

Anna Esposito  
Marcos Faundez-Zanuy  
Francesco Carlo Morabito  
Eros Pasero *Editors*



# Neural Approaches to Dynamics of Signal Exchanges



# **Smart Innovation, Systems and Technologies**

**Volume 151**

## **Series Editors**

Robert J. Howlett, Bournemouth University and KES International,  
Shoreham-by-sea, UK

Lakhmi C. Jain, Faculty of Engineering and Information Technology,  
Centre for Artificial Intelligence, University of Technology Sydney,  
Sydney, NSW, Australia

The Smart Innovation, Systems and Technologies book series encompasses the topics of knowledge, intelligence, innovation and sustainability. The aim of the series is to make available a platform for the publication of books on all aspects of single and multi-disciplinary research on these themes in order to make the latest results available in a readily-accessible form. Volumes on interdisciplinary research combining two or more of these areas is particularly sought.

The series covers systems and paradigms that employ knowledge and intelligence in a broad sense. Its scope is systems having embedded knowledge and intelligence, which may be applied to the solution of world problems in industry, the environment and the community. It also focusses on the knowledge-transfer methodologies and innovation strategies employed to make this happen effectively. The combination of intelligent systems tools and a broad range of applications introduces a need for a synergy of disciplines from science, technology, business and the humanities. The series will include conference proceedings, edited collections, monographs, handbooks, reference books, and other relevant types of book in areas of science and technology where smart systems and technologies can offer innovative solutions.

High quality content is an essential feature for all book proposals accepted for the series. It is expected that editors of all accepted volumes will ensure that contributions are subjected to an appropriate level of reviewing process and adhere to KES quality principles.

**\*\* Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, SCOPUS, Google Scholar and Springerlink \*\***

More information about this series at <http://www.springer.com/series/8767>

Anna Esposito · Marcos Faundez-Zanuy ·  
Francesco Carlo Morabito · Eros Pasero  
Editors

# Neural Approaches to Dynamics of Signal Exchanges



Springer

*Editors*

Anna Esposito

Department of Psychology

University of Campania Luigi Vanvitelli

Caserta, Italy

International Institute for Advanced

Scientific Studies (IIASS)

Italy

Francesco Carlo Morabito

Department of Civil, Environment, Energy  
and Materials Engineering

Mediterranea University of Reggio Calabria

Reggio Calabria, Italy

Marcos Faundez-Zanuy

Tecnocampus

Mataró, Spain

Eros Pasero

Dipartimento di Elettronica e

Telecomunicazioni

Politecnico di Torino

Turin, Italy

ISSN 2190-3018

ISSN 2190-3026 (electronic)

Smart Innovation, Systems and Technologies

ISBN 978-981-13-8949-8

ISBN 978-981-13-8950-4 (eBook)

<https://doi.org/10.1007/978-981-13-8950-4>

© Springer Nature Singapore Pte Ltd. 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.

The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721,  
Singapore

# **Technical Committee**

The chapters submitted to this book have been carefully reviewed by the following technical committee:

Alonso-Martinez Carlos, Universitat Pompeu Fabra  
Amorese Terry, Università degli Studi della Campania “Luigi Vanvitelli” and IIASS  
Angiulli Giovanni, Università Mediterranea di Reggio Calabria  
Bevilacqua Vitoantonio, Politecnico di Bari  
Camastra Francesco, Università Napoli Parthenope  
Campolo Maurizio, Università degli Studi Mediterranea Reggio Calabria  
Cauteruccio Francesco, Università degli Studi della Calabria  
Ciaramella Angelo, Università Napoli Parthenope  
Ciravegna Gabriele, Politecnico di Torino  
Cirrincione Giansalvo, Univeristé de Picardie  
Comajuncosas Andreu, Universitat Pompeu Fabra  
Cordasco Gennaro, Università degli Studi della Campania “Luigi Vanvitelli” and IIASS  
Cuciniello Marialucia, Università degli Studi della Campania “Luigi Vanvitelli” and IIASS  
Dattola Serena, Università degli Studi Mediterranea di Reggio Calabria  
Damasevicius Robertas, Kaunas University of Technology  
De Carlo Domenico, Università Mediterranea di Reggio Calabria  
Dell’Orco Silvia, School of Integrated Gestalt Psychotherapy—SIPGI  
Esposito Anna, Università degli Studi della Campania “Luigi Vanvitelli” and IIASS  
Esposito Antonietta Maria, Osservatorio Vesuviano sezione di Napoli  
Esposito Francesco, Università di Napoli Parthenope  
Esposito Marilena, International Institute for Advanced Scientific Studies (IIASS)  
Faundez-Zanuy Marcos, Universitat Pompeu Fabra  
Changhee Han, The University of Tokyo  
Ferone Alessio, University of Naples Parthenope  
Gabrieli Giulio, University of Trento

Gnisci Augusto, Università degli Studi della Campania “Luigi Vanvitelli”  
Gómez-Vilda Pedro, Universidad Politécnica de Madrid  
Gori Marco, University of Siena  
Koutsombogera Maria, Trinity College Dublin  
La Foresta Fabio, Università degli Studi Mediterranea Reggio Calabria  
Lo Giudice Paolo, University “Mediterranea” of Reggio Calabria  
Maldonato Mauro N., Università di Napoli “Federico II”  
Mammone Nadia, IRCCS Centro Neurolesi Bonino-Pulejo, Messina  
Martinez Olalla Rafael, Universidad Politécnica de Madrid  
Maskeliūnas Rytis, Kaunas University of Technology  
Matarazzo Olimpia, Università degli Studi della Campania “Luigi Vanvitelli”  
Mekyska Jiri, Brno University  
Morabito Francesco Carlo, Università Mediterranea di Reggio Calabria  
Nardone Davide, Università di Napoli “Parthenope”  
Parpinel Francesca, Ca’ Foscari University of Venice  
Panella Massimo, DIET Department, University of Rome “La Sapienza”  
Roure Josep, Universitat Pompeu Fabra  
Randazzo Vincenzo, Politecnico di Torino  
Reverdy Justine, Trinity College Dublin  
Rundo Leonardo, Università degli Studi di Milano-Bicocca  
Salvi Giampiero, KTH, Sweden  
Sappey-Marinier Dominique, Université de Lyon  
Scardapane Simone, Università di Roma “La Sapienza”  
Scarpiniti Michele, Università di Roma “La Sapienza”  
Senese Vincenzo Paolo, Università degli Studi della Campania “Luigi Vanvitelli”  
Sesa-Nogueras Enric, Universitat Pompeu Fabra  
Sgrò Annalisa, Università Mediterranea di Reggio Calabria  
Staiano Antonino, Università Napoli Parthenope  
Stamile Claudio, Université de Lyon  
Statue-Villar Antonio, Universitat Pompeu Fabra  
Suchacka Grażyna, Opole University  
Terracina Giorgio, Università della Calabria  
Troncone Alda, Università degli Studi della Campania “Luigi Vanvitelli” and  
IIASS  
Xavier Font-Aragones, Universitat Pompeu Fabra  
Uncini Aurelio, Università di Roma “La Sapienza”  
Ursino Domenico, Università Mediterranea di Reggio Calabria  
Vasquez Juan Camilo, University of Antioquia  
Vesperini Fabio, Università Politecnica delle Marche  
Vitabile Salvatore, Università degli Studi di Palermo  
Vogel Carl, Trinity College Dublin

## **Sponsoring Institutions**

International Institute for Advanced Scientific Studies (IIASS) of Vietri S/M (Italy)  
Department of Psychology, Università degli Studi della Campania “Luigi Vanvitelli” (Italy)

Provincia di Salerno (Italy)

Comune di Vietri sul Mare, Salerno (Italy)

International Neural Network Society (INNS)

Università Mediterranea di Reggio Calabria (Italy)

# Preface

The book aims to assemble research from different fields and create a common public data framework for a large variety of applications that may range from medical diagnosis to entertainment devices (speech, facial expressions, gaze and gesture), holding the promises to contribute to the development of intelligent interactive dialog systems that simplify everyday-life man-machine interaction by taking into account individual, socio-cultural differences, and contextual instances, intended here as the dynamics of values that contextual variables assume at a given instant. Interdisciplinary aspects are taken into account and research is proposed from different fields: mathematics, computer vision, speech analysis and synthesis, machine learning, signal processing, telecommunication, human-computer interaction, psychology, anthropology, sociology, neural networks, machine learning, and advanced sensing.

The topics of this book vary from the processing of audio-visual signals to the detection of user perceived states, dedicating a section to the last scientific discoveries in processing verbal (lexicon, syntax, and pragmatics), auditory (voice, intonation, vocal expressions) and visual signals (gestures, body language, facial expressions), as well as to algorithms for detecting communication disorders, remote health status monitoring, sentiment and affect analysis, social behaviors and engagements.

The remaining sections are dedicated to neural and machine learning algorithms for the implementation of advanced telecommunication systems, communication with people with special needs, emotion modulation by computer contents, advanced sensors for tracking changes in real-life and automatic systems, as well as the development of advanced human-computer interfaces. This socio-emotional content is vital for building trusting, productive relationships that go beyond purely factual and task oriented communication and the proposed technological solutions will enhance and improve the efficiency of European industry and the quality of service provided to citizens. Therefore, the proposed book has a wide view and does not focus on solving a particular problem, rather describe the results of a research that has positive effects in different fields and for different applications.

The contributions in the book cover different scientific areas according to the thematic classification reported below, even though these areas are closely connected in the themes they afford and provide fundamental insights for the cross-fertilization of different disciplines:

- Machine learning and Artificial Neural Networks: Algorithms and models,
- Social and biometric data for applications in human-computer interfaces

It must be said that human data analysis is central to many endeavors both in basic research and across application domains, and that the contributes proposed in this book aim to enable human centered informatics.

The chapters composing this book were first discussed at the international workshop on neural networks (WIRN 2018) held in Vietri Sul Mare from the 13th to the 15th of June 2018, in the regular and special sessions. In particular it is worth to mention the special session on: *Dynamics of Signal Exchanges* organized by Anna Esposito, Antonietta M. Esposito, Gennaro Cordasco, Mauro N. Maldonato, Francesco Carlo Morabito, Vincenzo Paolo Senese, Carl Vogel; and the special session on *Neural Networks and Pattern Recognition in Medicine* organized by Giansalvo Cirrincione and Vitoantonio Bevilacqua.

The scientists contributing to this book are specialists in their respective disciplines. We are indebted to them for making (through their chapters) the book a meaningful effort. The coordination and production of this book has been brilliantly conducted by the Springer project coordinator for books production Mr. **Maniarasan Gandhi**, the Springer executive editor Dr. **Thomas Ditzinger**, and the editor assistant Mr. **Holger Schaepe**. They are the recipient of our deepest appreciation. This initiative has been skillfully supported by the Editors in chief of the Springer series Smart Innovation, Systems and Technologies, Professors **Jain Lakhmi C.**, and **Howlett Robert James**, to whom goes out deepest gratitude.

Caserta, Italy  
Mataró, Spain  
Reggio Calabria, Italy  
Turin, Italy

Anna Esposito  
Marcos Faundez-Zanuy  
Francesco Carlo Morabito  
Eros Pasero

# Contents

## Part I Introduction

<b>1 Some Note on Artificial Intelligence . . . . .</b>	<b>3</b>
Anna Esposito, Marcos Faundez-Zanuy, Francesco Carlo Morabito and Eros Pasero	
1.1 Introduction . . . . .	3
1.2 Content of This Book . . . . .	6
1.3 Conclusions . . . . .	7
References . . . . .	8

## Part II Neural Networks and Related Applications

<b>2 Music Genre Classification Using Stacked Auto-Encoders . . . . .</b>	<b>11</b>
Michele Scarpiniti, Simone Scardapane, Danilo Comminiello and Aurelio Uncini	
2.1 Introduction . . . . .	11
2.2 The Proposed Architecture . . . . .	12
2.2.1 The Stacked Auto-Encoder . . . . .	14
2.2.2 Classification . . . . .	14
2.3 Feature Extraction . . . . .	15
2.4 Experimental Results . . . . .	16
2.5 Conclusions . . . . .	18
References . . . . .	18
<b>3 Linear Artificial Forces for Human Dynamics in Complex Contexts . . . . .</b>	<b>21</b>
Pasquale Coscia, Lamberto Ballan, Francesco A. N. Palmieri, Alexandre Alahi and Silvio Savarese	
3.1 Introduction . . . . .	22
3.2 Related Work . . . . .	23
3.3 Dynamic Model . . . . .	23
3.4 Artificial Forces . . . . .	24

3.5 Experiments . . . . .	27
3.6 Conclusion . . . . .	32
References . . . . .	32
<b>4 Convolutional Recurrent Neural Networks and Acoustic Data Augmentation for Snore Detection . . . . .</b>	<b>35</b>
Fabio Vesperini, Luca Romeo, Emanuele Principi, Andrea Monteriù and Stefano Squartini	
4.1 Introduction . . . . .	35
4.1.1 Related Works . . . . .	36
4.2 Proposed Approach . . . . .	37
4.2.1 Features Extraction . . . . .	37
4.3 Experiments . . . . .	39
4.3.1 Dataset . . . . .	39
4.3.2 Data Augmentation Techniques . . . . .	41
4.3.3 Performance Metrics . . . . .	42
4.3.4 Experimental Setup . . . . .	42
4.4 Results . . . . .	43
4.5 Conclusion . . . . .	44
References . . . . .	45
<b>5 Italian Text Categorization with Lemmatization and Support Vector Machines . . . . .</b>	<b>47</b>
Francesco Camastra and Gennaro Razi	
5.1 Introduction . . . . .	47
5.2 Italian Text Categorizer . . . . .	48
5.2.1 Tokenization Module . . . . .	48
5.2.2 Stopping Module . . . . .	49
5.2.3 Lemmatization Module . . . . .	49
5.2.4 Bag-of-Words Representation of Text . . . . .	50
5.2.5 Feature Ranker . . . . .	50
5.2.6 Classification . . . . .	51
5.3 Experimental Results . . . . .	52
5.4 Conclusions . . . . .	54
References . . . . .	54
<b>6 SOM-Based Analysis of Volcanic Rocks: An Application to Somma–Vesuvius and Campi Flegrei Volcanoes (Italy) . . . . .</b>	<b>55</b>
Antonietta M. Esposito, Andrea De Bernardo, Salvatore Ferrara, Flora Giudicepietro and Lucia Pappalardo	
6.1 Introduction . . . . .	56
6.2 The Selected Dataset and Parametrization . . . . .	56
6.3 The Self-organizing Map Method . . . . .	57
6.4 Conclusions . . . . .	59
References . . . . .	59

<b>7 Toward an Automatic Classification of SEM Images of Nanomaterials via a Deep Learning Approach . . . . .</b>	<b>61</b>
Cosimo Ieracitano, Fabiola Pantó, Nadia Mammone, Annunziata Paviglianiti, Patrizia Frontera and Francesco Carlo Morabito	
7.1 Introduction . . . . .	62
7.2 Methodology . . . . .	63
7.2.1 Electrospinning Process . . . . .	64
7.2.2 Convolutional Neural Network . . . . .	66
7.3 Experimental Results . . . . .	67
7.4 Conclusions . . . . .	70
References . . . . .	71
<b>8 An Effective Fuzzy Recommender System for Fund-raising Management . . . . .</b>	<b>73</b>
Luca Barzanti, Silvio Giove and Alessandro Pezzi	
8.1 Introduction . . . . .	74
8.2 Contacts' Characterization . . . . .	75
8.2.1 The Personal Characterization for Donors and Contacts . . . . .	75
8.2.2 Nonparametric Estimation of Volume and Frequency . . . . .	77
8.3 Computational Results . . . . .	79
8.4 Conclusion . . . . .	81
References . . . . .	81
<b>9 Reconstruction, Optimization and Quality Check of Microsoft HoloLens-Acquired 3D Point Clouds . . . . .</b>	<b>83</b>
Gianpaolo Francesco Trotta, Sergio Mazzola, Giuseppe Gelardi, Antonio Brunetti, Nicola Marino and Vitoantonio Bevilacqua	
9.1 Introduction . . . . .	83
9.2 Materials . . . . .	84
9.2.1 Mesh Reconstruction . . . . .	85
9.2.2 Quality Indexes . . . . .	87
9.3 Methods . . . . .	88
9.3.1 Performances Optimization . . . . .	88
9.3.2 Quality Check . . . . .	89
9.4 Results and Conclusion . . . . .	91
References . . . . .	92
<b>10 The “Probabilistic Rand Index”: A Look from Some Different Perspectives . . . . .</b>	<b>95</b>
Stefano Rovetta, Francesco Masulli and Alberto Cabri	
10.1 Introduction . . . . .	95
10.2 Co-Association Statistics . . . . .	96
10.3 The Probabilistic Rand Index . . . . .	98

10.4	Analysis of the Probabilistic Rand Index . . . . .	99
10.4.1	A Probabilistic Perspective . . . . .	99
10.4.2	An Information-Theoretic Perspective . . . . .	100
10.4.3	A Diversity-Theoretic Perspective . . . . .	101
10.5	Experimental Simulations . . . . .	102
10.6	Conclusion . . . . .	105
	References . . . . .	105
<b>11</b>	<b>Dimension Reduction Techniques in a Brain–Computer Interface Application</b> . . . . .	107
	Federico Cozza, Paola Galdi, Angela Serra, Gabriele Pasqua, Luigi Pavone and Roberto Tagliaferri	
11.1	Introduction . . . . .	108
11.2	Materials and Methods . . . . .	109
11.2.1	Population Used in the Study . . . . .	109
11.2.2	Experimental Design . . . . .	110
11.2.3	Signal Processing . . . . .	110
11.2.4	Feature Vector . . . . .	112
11.2.5	Dimension Reduction . . . . .	113
11.3	Results and Discussion . . . . .	114
11.4	Conclusions . . . . .	115
	References . . . . .	117
<b>12</b>	<b>Blind Source Separation Using Dictionary Learning in Wireless Sensor Network Scenario</b> . . . . .	119
	Angelo Ciaramella, Davide Nardone and Antonino Staiano	
12.1	Introduction . . . . .	119
12.2	Proposed Methodology . . . . .	120
12.2.1	Blind Source Separation . . . . .	121
12.2.2	Dictionary Learning . . . . .	121
12.2.3	Mixture Decomposition into Sparse Signals . . . . .	122
12.2.4	Estimate of the Mixing Matrix . . . . .	123
12.2.5	Separating Sources Using Sparse Coding . . . . .	123
12.2.6	Source Reconstruction from Sparseness . . . . .	123
12.3	Experimental Results . . . . .	124
12.3.1	Principles and Operating Simulation in a WSN . . . . .	124
12.3.2	Datasets and Evaluation Metrics . . . . .	124
12.3.3	Separation Results with Fixed Dictionary . . . . .	125
12.3.4	Comparing Different Strategies for Learning Dictionary . . . . .	125
12.3.5	Effects of the block strategy on the system performance . . . . .	126
12.4	Example of Separation of Four Female Speeches . . . . .	128
12.5	Conclusions . . . . .	129
	References . . . . .	130

<b>13 A Comparison of Apache Spark Supervised Machine Learning Algorithms for DNA Splicing Site Prediction . . . . .</b>	133
Valerio Morfino, Salvatore Rampone and Emanuel Weitschek	
13.1 Introduction . . . . .	133
13.1.1 Brief Biological Background and the DNA Splicing Site Prediction Problem . . . . .	134
13.2 Methods and Description of the Experiments . . . . .	135
13.2.1 Apache Spark . . . . .	135
13.2.2 Dataset Description . . . . .	136
13.2.3 Dataset Encoding . . . . .	137
13.2.4 Execution Environment . . . . .	137
13.3 Experimental Results . . . . .	138
13.3.1 Comparison with the Performance of the U-BRAIN Algorithm . . . . .	140
13.4 Conclusions and Future Works . . . . .	141
References . . . . .	142
<b>14 Recurrent ANNs for Failure Predictions on Large Datasets of Italian SMEs . . . . .</b>	145
Leonardo Nadali, Marco Corazza, Francesca Parpinel and Claudio Pizzi	
14.1 Introduction . . . . .	145
14.2 RNNs Versus MLPs . . . . .	146
14.3 The Dataset . . . . .	148
14.4 The Application . . . . .	149
14.4.1 ANN Architectures . . . . .	149
14.4.2 Performances . . . . .	150
14.5 Results . . . . .	150
14.6 Conclusions . . . . .	154
References . . . . .	155
<b>15 Inverse Classification for Military Decision Support Systems . . . . .</b>	157
Pietro Russo and Massimo Panella	
15.1 Introduction . . . . .	157
15.2 Proposed Methodology . . . . .	158
15.3 Validation and Results . . . . .	162
15.4 Conclusion . . . . .	165
References . . . . .	166
<b>16 Simultaneous Learning of Fuzzy Sets . . . . .</b>	167
Luca Cermenati, Dario Malchiodi and Anna Maria Zanaboni	
16.1 Introduction . . . . .	167
16.2 Inferring the Membership Function to a Fuzzy Set . . . . .	168
16.3 Simultaneously Inferring Several Membership Functions . . . . .	170

16.4	Experiments . . . . .	171
16.5	Conclusions . . . . .	174
	References . . . . .	174
<b>17</b>	<b>Trees in the Real Field . . . . .</b>	<b>177</b>
	Alessandro Betti and Marco Gori	
17.1	Introduction . . . . .	177
17.2	Uniform Real-Valued Tree Representations . . . . .	178
17.3	Non-commutative Left-Right Matrices . . . . .	186
17.4	Conclusions . . . . .	187
	References . . . . .	187
<b>18</b>	<b>Graded Possibilistic Meta Clustering . . . . .</b>	<b>189</b>
	Alessio Ferone and Antonio Maratea	
18.1	Introduction . . . . .	189
18.2	Meta Clustering . . . . .	190
18.2.1	Baseline Clusterings . . . . .	190
18.2.2	Clusterings Similarity . . . . .	190
18.2.3	Clustering Partitions by Relational Clustering . . . . .	191
18.2.4	Fuzzy $c$ -medoids . . . . .	192
18.2.5	Rough $c$ -Medoids . . . . .	193
18.2.6	Graded Possibilistic $c$ -medoids . . . . .	194
18.3	Experiments . . . . .	194
18.3.1	Data . . . . .	195
18.3.2	Performance Measures . . . . .	195
18.3.3	Results and Discussion . . . . .	196
18.4	Conclusions . . . . .	198
	References . . . . .	199
<b>19</b>	<b>Probing a Deep Neural Network . . . . .</b>	<b>201</b>
	Francesco A. N. Palmieri, Mario Baldi, Amedeo Buonanno, Giovanni Di Gennaro and Francesco Ospedale	
19.1	Introduction . . . . .	202
19.2	Multi-Layer Convolutional Networks . . . . .	203
19.3	Backward Reconstructions . . . . .	205
19.3.1	Filter Reconstruction . . . . .	208
19.4	Best Input Selection . . . . .	209
19.5	Activation Statistics . . . . .	209
19.6	Conclusions and Future Directions . . . . .	210
	References . . . . .	211
<b>20</b>	<b>Neural Epistemology in Dynamical System Learning . . . . .</b>	<b>213</b>
	Pietro Barbiero, Giansalvo Cirrincione, Maurizio Cirrincione, Elio Piccolo and Francesco Vaccarino	
20.1	The Need for an Epistemology . . . . .	213
20.2	Mathematical Models of Dynamical Systems . . . . .	214

20.3	The Poincaré Recurrence Theorem . . . . .	215
20.4	What Can Go Wrong . . . . .	216
20.5	How Machine Learning Sees Dynamical Systems . . . . .	217
20.6	A Song Experiment: The Frère Jacques Song . . . . .	217
20.7	Conclusion . . . . .	220
	References . . . . .	221
<b>21</b>	<b>Assessing Discriminating Capability of Geometrical Descriptors for 3D Face Recognition by Using the GH-EXIN Neural Network . . . . .</b>	<b>223</b>
	Gabriele Ciravegna, Giansalvo Cirrincione, Federica Marcolin, Pietro Barbiero, Nicole Dagnes and Elio Piccolo	
21.1	Introduction . . . . .	223
21.2	Database Creation and Goals . . . . .	224
21.3	Data Analysis . . . . .	224
21.3.1	Biclustering . . . . .	226
21.3.2	The GH-EXIN Neural Network . . . . .	227
21.4	Analysis of the Database . . . . .	230
21.5	Conclusion . . . . .	232
	References . . . . .	232
<b>22</b>	<b>Growing Curvilinear Component Analysis (GCCA) for Stator Fault Detection in Induction Machines . . . . .</b>	<b>235</b>
	Giansalvo Cirrincione, Vincenzo Randazzo, Rahul R. Kumar, Maurizio Cirrincione and Eros Pasero	
22.1	Introduction . . . . .	235
22.2	The Growing CCA (GCCA) . . . . .	237
22.3	Stator-Winding Fault Experiment . . . . .	238
22.4	Conclusions . . . . .	243
	References . . . . .	243
<b>Part III Neural Networks and Pattern Recognition in Medicine</b>		
<b>23</b>	<b>A Neural Based Comparative Analysis for Feature Extraction from ECG Signals . . . . .</b>	<b>247</b>
	Giansalvo Cirrincione, Vincenzo Randazzo and Eros Pasero	
23.1	Introduction . . . . .	247
23.2	The Proposed Approach . . . . .	249
23.3	Feature Analysis and Comparison . . . . .	249
23.3.1	ECG Raw Data . . . . .	250
23.3.2	Temporal Features . . . . .	251
23.3.3	Eigenvector Features . . . . .	253
23.3.4	Discussion . . . . .	254
	References . . . . .	255

<b>24 A Multi-modal Tool Suite for Parkinson's Disease Evaluation and Grading . . . . .</b>	257
Giacomo Donato Cascarano, Antonio Brunetti, Domenico Buongiorno, Gianpaolo Francesco Trotta, Claudio Loconsole, Ilaria Bortone and Vitoantonio Bevilacqua	
24.1 Introduction . . . . .	257
24.2 Materials and Methods . . . . .	259
24.2.1 Participants . . . . .	259
24.2.2 Experimental Set-up . . . . .	259
24.2.3 Feature Extraction . . . . .	260
24.2.4 Classification . . . . .	262
24.3 Results and Discussion . . . . .	263
24.4 Conclusion . . . . .	265
References . . . . .	265
<b>25 CNN-Based Prostate Zonal Segmentation on T2-Weighted MR Images: A Cross-Dataset Study . . . . .</b>	269
Leonardo Rundo, Changhee Han, Jin Zhang, Ryuichiro Hataya, Yudai Nagano, Carmelo Militello, Claudio Ferretti, Marco S. Nobile, Andrea Tangherloni, Maria Carla Gilardi, Salvatore Vitabile, Hideki Nakayama and Giancarlo Mauri	
25.1 Introduction . . . . .	270
25.2 Background . . . . .	271
25.3 Materials and Methods . . . . .	271
25.3.1 MRI Datasets . . . . .	271
25.3.2 CNN-Based Prostate Zonal Segmentation . . . . .	273
25.3.3 Influence of Pre-training . . . . .	275
25.4 Results . . . . .	276
25.5 Discussion and Conclusions . . . . .	276
References . . . . .	279
<b>26 Understanding Cancer Phenomenon at Gene Expression Level by using a Shallow Neural Network Chain . . . . .</b>	281
Pietro Barbiero, Andrea Bertotti, Gabriele Ciravagna, Giansalvo Cirrincione, Elio Piccolo and Alberto Tonda	
26.1 Biological Introduction . . . . .	282
26.2 Data Set . . . . .	282
26.3 Objective . . . . .	282
26.4 Shallow Neural Networks . . . . .	283
26.4.1 Mathematical Model of the Network Architecture . . . . .	283
26.4.2 Objective Function . . . . .	284
26.4.3 Parameter Optimization . . . . .	284
26.5 Experiments and Discussion . . . . .	286
26.6 Conclusion . . . . .	288
References . . . . .	289

<b>27 Infinite Brain MR Images: PGGAN-Based Data Augmentation for Tumor Detection . . . . .</b>	291
Changhee Han, Leonardo Rundo, Ryosuke Araki, Yujiro Furukawa, Giancarlo Mauri, Hideki Nakayama and Hideaki Hayashi	
27.1 Introduction . . . . .	292
27.2 Generative Adversarial Networks . . . . .	294
27.3 Materials and Methods . . . . .	294
27.3.1 BRATS 2016 Training Dataset . . . . .	294
27.3.2 PGGAN-Based Image Generation . . . . .	294
27.3.3 Tumor Detection Using ResNet-50 . . . . .	296
27.3.4 Clinical Validation Using the Visual Turing Test . . . . .	297
27.3.5 Visualization Using t-SNE . . . . .	298
27.4 Results . . . . .	298
27.4.1 MR Images Generated by PGGANs . . . . .	298
27.4.2 Tumor Detection Results . . . . .	299
27.4.3 Visual Turing Test Results . . . . .	300
27.4.4 t-SNE Result . . . . .	300
27.5 Conclusion . . . . .	301
References . . . . .	302
<b>28 DNA Microarray Classification: Evolutionary Optimization of Neural Network Hyper-parameters . . . . .</b>	305
Pietro Barbiero, Andrea Bertotti, Gabriele Ciravegna, Giansalvo Cirrincione and Elio Piccolo	
28.1 Introduction . . . . .	305
28.2 Proposed Approach . . . . .	306
28.3 Experiments and Discussion . . . . .	309
28.4 Conclusions . . . . .	310
References . . . . .	311
<b>29 Evaluation of a Support Vector Machine Based Method for Crohn’s Disease Classification . . . . .</b>	313
S. Franchini, M. C. Terranova, G. Lo Re, S. Salerno, M. Midiri and Salvatore Vitabile	
29.1 Introduction . . . . .	314
29.1.1 Related Works . . . . .	314
29.1.2 Our Contribution . . . . .	315
29.2 Materials . . . . .	316
29.3 Methods . . . . .	317
29.3.1 Classification Methods Comparison . . . . .	317
29.3.2 Feature Extraction . . . . .	318
29.3.3 Feature Reduction Techniques . . . . .	318
29.3.4 Support Vector Machines . . . . .	318
29.3.5 K-Fold Cross-Validation . . . . .	320

29.4	Results and Discussions . . . . .	320
29.4.1	Performance Evaluation . . . . .	321
29.4.2	Feature Space Reduction Techniques . . . . .	323
29.4.3	Radiologist Driven Reduction Techniques . . . . .	324
29.5	Conclusions . . . . .	325
	References . . . . .	326
<b>Part IV Dynamics of Signal Exchanges</b>		
30	<b>Seniors' Appreciation of Humanoid Robots . . . . .</b>	331
	Anna Esposito, Marialucia Cuciniello, Terry Amorese, Antonietta M. Esposito, Alda Troncone, Mauro N. Maldonato, Carl Vogel, Nikolaos Bourbakis and Gennaro Cordasco	
30.1	Introduction . . . . .	332
30.2	Materials and Methods . . . . .	333
30.2.1	Stimuli . . . . .	333
30.2.2	Participants . . . . .	333
30.2.3	Tools and Procedures . . . . .	335
30.3	Analysis and Results . . . . .	335
30.4	Discussion and Conclusion . . . . .	341
	References . . . . .	344
31	<b>The Influence of Personality Traits on the Measure of Restorativeness in an Urban Park: A Multisensory Immersive Virtual Reality Study . . . . .</b>	347
	Vincenzo Paolo Senese, Aniello Pascale, Luigi Maffei, Federico Cioffi, Ida Sergi, Augusto Gnisci and Massimiliano Masullo	
31.1	Introduction . . . . .	348
31.2	Method . . . . .	349
31.2.1	Sample . . . . .	349
31.2.2	Procedure . . . . .	350
31.2.3	Measures . . . . .	350
31.2.4	Virtual Scenarios . . . . .	351
31.3	Data Analysis . . . . .	352
31.4	Results . . . . .	352
31.5	Discussion . . . . .	354
31.6	Conclusions . . . . .	355
	References . . . . .	355
32	<b>Linguistic Repetition in Three-Party Conversations . . . . .</b>	359
	Justine Reverdy, Maria Koutsombogera and Carl Vogel	
32.1	Introduction . . . . .	359
32.2	Data set . . . . .	361
32.2.1	Conversational Roles . . . . .	362
32.2.2	Annotation Process . . . . .	362

32.3	Method .....	364
32.4	Results and Discussion .....	365
32.4.1	Overview of Repetition Types and Dialogue Sections .....	365
32.4.2	Above Chance Repetitions and Facilitators' Feedback .....	367
32.5	Conclusion .....	369
	References .....	369
<b>33</b>	<b>An Experiment on How Naïve People Evaluate Interruptions as Effective, Unpleasant and Influential .....</b>	<b>371</b>
	Ida Sergi, Augusto Gnisci, Vincenzo Paolo Senese and Angelo Di Gennaro	
33.1	Introduction .....	371
33.1.1	Hypotheses .....	373
33.2	Study .....	374
33.2.1	Method .....	374
33.3	Results .....	375
33.3.1	The Effect of Interruptions in Terms of Effectiveness, Pleasantness and Influence .....	375
33.3.2	Correlations of Perception of Interruption with Effectiveness, Unpleasantness and Influence .....	376
33.4	Discussion and Conclusions .....	377
	References .....	381
<b>34</b>	<b>Analyzing Likert Scale Inter-annotator Disagreement .....</b>	<b>383</b>
	Carl Vogel, Maria Koutsombogera and Rachel Costello	
34.1	Introduction .....	383
34.2	Data Set .....	385
34.3	Method .....	387
34.4	Results .....	389
34.5	Discussion .....	390
34.6	Conclusion .....	392
	References .....	393
<b>35</b>	<b>PySiology: A Python Package for Physiological Feature Extraction .....</b>	<b>395</b>
	Giulio Gabrieli, Atiqah Azhari and Gianluca Esposito	
35.1	Introduction .....	395
35.2	Development Workflow .....	396
35.3	Modules, Features, and Installation .....	397
35.3.1	Package Structure .....	397
35.3.2	Feature Estimation .....	397
35.3.3	Installation .....	398

<b>35.4 Advanced Example: Predicting Valence of an Image from Physiological Features . . . . .</b>	398
35.4.1 Data Collection . . . . .	399
35.4.2 Preprocessing . . . . .	399
35.4.3 Feature Extraction . . . . .	399
35.4.4 Classification . . . . .	400
35.4.5 Results and Discussion . . . . .	400
<b>35.5 Future Development . . . . .</b>	401
<b>35.6 Conclusion . . . . .</b>	401
<b>References . . . . .</b>	401
<b>36 Effect of Sensor Density on eLORETA Source Localization Accuracy . . . . .</b>	403
Serena Dattola, Fabio La Foresta, Lilla Bonanno, Simona De Salvo, Nadia Mammone, Silvia Marino and Francesco Carlo Morabito	
36.1 Introduction . . . . .	403
36.2 Materials and Methods . . . . .	405
36.2.1 The EEG Inverse Problem . . . . .	405
36.2.2 LORETA . . . . .	406
36.2.3 Acquisition System . . . . .	407
36.3 Results . . . . .	409
36.3.1 Data Description . . . . .	409
36.3.2 Analysis of Result . . . . .	409
36.4 Discussion . . . . .	411
36.5 Conclusions . . . . .	413
<b>References . . . . .</b>	413
<b>37 To the Roots of the Sense of Self: Proposals for a Study on the Emergence of Body Awareness in Early Infancy Through a Deep Learning Method . . . . .</b>	415
Alfonso Davide Di Sarno, Raffaele Sperandeo, Giuseppina Di Leva, Irene Fabbricino, Enrico Moretto, Silvia Dell'Orco and Mauro N. Maldonato	
37.1 Introduction . . . . .	416
37.2 The Observational Methods in Psychology . . . . .	417
37.3 The Observation of the Dyad . . . . .	418
37.4 Theory of Reference . . . . .	419
37.4.1 Movements: The Child First Language . . . . .	419
37.4.2 The Six Fundamental Movements Theory . . . . .	420
37.4.3 Effects of the Mismatches in the Relationship . . . . .	421
37.5 The Deep Learning Model Applied to Human Behaviour . . . . .	422
37.5.1 Long Short-Term Memory Recurrent Neural Network . . . . .	423

37.6	Study Goals . . . . .	423
37.7	Method . . . . .	424
37.8	Conclusion and Expected Results . . . . .	424
	References . . . . .	425
<b>38</b>	<b>Performance of Articulation Kinetic Distributions Vs MFCCs in Parkinson's Detection from Vowel Utterances . . . . .</b>	<b>431</b>
	Andrés Gómez-Rodellar, Agustín Álvarez-Marquina, Jiri Mekyska, Daniel Palacios-Alonso, Djamila Meghraoui and Pedro Gómez-Vilda	
38.1	Introduction . . . . .	432
38.2	Kinematic Model of Speech Articulation . . . . .	432
38.3	Materials and Methods . . . . .	434
38.3.1	Validation of Articulatory Gesture to Acoustic Feature Mapping . . . . .	434
38.3.2	AKV-Based Parkinson Disease Detection . . . . .	438
38.4	Results and Discussion . . . . .	439
38.5	Conclusions . . . . .	440
	References . . . . .	441
<b>39</b>	<b>From “Mind and Body” to “Mind in Body”: A Research Approach for a Description of Personality as a Functional Unit of Thoughts, Behaviours and Affective States . . . . .</b>	<b>443</b>
	Daniela Iennaco, Raffaele Sperandeo, Lucia Luciana Mosca, Martina Messina, Enrico Moretto, Valeria Cioffi, Silvia Dell'Orco and Mauro N. Maldonato	
39.1	Introduction . . . . .	444
39.2	Theoretical Assumptions . . . . .	444
39.3	Aims of the Study . . . . .	446
39.4	Methodology . . . . .	447
39.5	Statistical Analysis . . . . .	448
39.6	Tools . . . . .	448
39.6.1	Sixteen Personality Factor Questionnaire-Fifth Edition (16PF-5) . . . . .	448
39.6.2	Electrophysiological Recording (EECG, sEMG, HRV, GSR, TEMP) . . . . .	449
39.7	Expected Results and Conclusions . . . . .	450
	References . . . . .	450
<b>40</b>	<b>Online Handwriting and Signature Normalization and Fusion in a Biometric Security Application . . . . .</b>	<b>453</b>
	Carlos Alonso-Martinez and Marcos Faundez-Zanuy	
40.1	Introduction . . . . .	453
40.2	Experimental Results . . . . .	454
40.2.1	Biometric Database . . . . .	454
40.2.2	Online Signature Biometric Recognition . . . . .	454

40.2.3	Online Handwritten Capital Letters Biometric Recognition . . . . .	456
40.2.4	Combined Approach for Biometric Recognition . . . . .	457
40.2.5	Score Normalization . . . . .	457
40.3	Conclusions . . . . .	461
	References . . . . .	463
<b>41</b>	<b>Data Insights and Classification in Multi-sensor Database for Cervical Injury . . . . .</b>	<b>465</b>
	Xavi Font, Carles Paul and Eloi Rodriguez	
41.1	Introduction . . . . .	465
41.2	Experiment Setup and the Data Set . . . . .	466
41.2.1	Inertial Data . . . . .	468
41.2.2	Thermographic Data . . . . .	468
41.2.3	EEG Data . . . . .	468
41.2.4	Data Summary . . . . .	468
41.3	Data Insights . . . . .	469
41.4	Classification Through Penalized Regression . . . . .	469
41.5	Results . . . . .	472
41.5.1	Sensor Data . . . . .	472
41.5.2	Survey Data . . . . .	472
41.6	Conclusions . . . . .	473
	References . . . . .	474
<b>42</b>	<b>Estimating the Asymmetry of Brain Network Organization in Stroke Patients from High-Density EEG Signals . . . . .</b>	<b>475</b>
	Nadia Mamnone, Simona De Salvo, Silvia Marino, Lilla Bonanno, Cosimo Ieracitano, Serena Dattola, Fabio La Foresta and Francesco Carlo Morabito	
42.1	Introduction . . . . .	475
42.2	Estimating the Brain Network Organization . . . . .	477
42.2.1	Patients' Recruitment . . . . .	477
42.2.2	HD-EEG Recording and Preprocessing . . . . .	477
42.2.3	EEG-based Complex Network Analysis . . . . .	478
42.2.4	Permutation Disalignment Index . . . . .	479
42.3	Results . . . . .	480
42.4	Conclusions . . . . .	482
	References . . . . .	482
<b>43</b>	<b>Preliminary Study on Biometric Recognition Based on Drawing Tasks . . . . .</b>	<b>485</b>
	Josep Lopez-Xarbau, Marcos Faundez-Zanuy and Manuel Garnacho-Castaño	
43.1	Introduction . . . . .	485
43.2	Database . . . . .	486
43.3	Recognition Algorithms . . . . .	487

43.4 Experimental Results . . . . .	488
43.5 Conclusions . . . . .	493
References . . . . .	493
<b>44 Exploring the Relationship Between Attention and Awareness. Neurophenomenology of the Centroencephalic Space of Functional Integration . . . . .</b>	<b>495</b>
Mauro N. Maldonato, Raffaele Sperandeo, Anna Esposito, Ciro Punzo and Silvia Dell'Orco	
44.1 Modal Levels of Attention . . . . .	496
44.2 Procedures in Continuous Evolution . . . . .	497
44.3 Functional Asymmetries Between Attention and Awareness . . . . .	498
44.4 Future Research Lines . . . . .	500
References . . . . .	500
<b>45 Decision-Making Styles in an Evolutionary Perspective . . . . .</b>	<b>503</b>
Silvia Dell'Orco, Raffaele Sperandeo, Ciro Punzo, Mario Bottone, Anna Esposito, Antonietta M. Esposito, Vincenzo Bochicchio and Mauro N. Maldonato	
45.1 Introduction . . . . .	504
45.2 Recognition-Primed Decision Model . . . . .	505
45.3 Cognition and Individual Differences in Decision-Making Styles . . . . .	506
45.4 Conclusions and Future Perspectives of Research . . . . .	509
References . . . . .	510
<b>46 The Unaware Brain: The Role of the Interconnected Modal Matrices in the Centrencephalic Space of Functional Integration . . . . .</b>	<b>513</b>
Mauro N. Maldonato, Paolo Valerio, Raffaele Sperandeo, Antonietta M. Esposito, Roberto Vitelli, Cristiano Scandurra, Benedetta Muzii and Silvia Dell'Orco	
46.1 Unconscious Cognitive Processing . . . . .	514
46.2 The Thresholds of Subliminal Perception . . . . .	515
46.3 Unconscious Affective Logic . . . . .	516
46.4 Could Desire Perhaps Be an Illusion? . . . . .	517
46.5 Conclusions and Future Explorations . . . . .	518
References . . . . .	519
<b>Author Index . . . . .</b>	<b>523</b>

## About the Editors

**Anna Esposito** received her “Laurea Degree” summa cum laude in Information Technology and Computer Science from Università di Salerno in 1989. She received her PhD Degree in Applied Mathematics and Computer Science from Università di Napoli “Federico II” in 1995. She is currently working as an Associate Professor in Computer Science at the Department of Psychology, Università della Campania. She is associated with WSU as Research Affiliate. Previously, she worked as Assistant Professor at the Department of Physics at the Università di Salerno (Italy). She also held a position of Research Professor (2000–2002) at the Department of Computer Science and Engineering at Wright State University(WSU), Dayton, OH, USA. She has authored 190+ peer reviewed publications in international journals, books, and conference proceedings. She edited/co-edited 28+ books and conference proceedings in collaboration with Italian, EU and overseas colleagues. She has also guest edited several journal special issues.

She has been the proposer and chair of COST 2102. Currently, she is the Italian Management Committee Member of COST Action CA15218 and the Italian Management Committee Substitute Member of COST Action IS1406. Previously also she has been Italian Management Committee member of many COST ACTIONS. Since 2006, she is a Member of the European Network for the Advancement of Artificial Cognitive Systems, Interaction and Robotics.

**Marcos Faundez-Zanuy** received his B.Sc. degree in Telecommunication in 1993 and the Ph.D. degree in 1998 from the Polytechnic University of Catalunya. He is now Full Professor at ESUP Tecnocampus Mataro and heads the Signal Processing Group. He also held the position of Dean of Escola Superior Politecnica Tecnocampus from September, 2009 to April, 2018. His area of research interests are in the fields of biometrics applied to security and health. He was the initiator and Chairman of the European COST action 277 “Nonlinear speech processing”, and the secretary of COST action 2102 “Cross-Modal Analysis of Verbal and

Non-Verbal Communication". He has authored more than 50 papers indexed in international journals, more than 100 conference papers, and around 10 books. He is also responsible of 10 national and European research projects.

**Francesco Carlo Morabito** got the "Laurea" Degree (Summa cum laude) from the University of Napoli (Italy) in 1985. He now serves as Vice-Rector for Internationalisation at University of Reggio Calabria, Italy. Previously, he was Professor of Electrical Engineering at same University. He has also worked at Radar Department, Selenia SpA, Rome, as Researcher. He also taught at the EPFL, Lausanne, Switzerland, University of Naples, University of Messina, and University of Cosenza. He was a Visiting Researcher at Max-Planck Institute fuer Plasmaphysiks Muenchen, Germany from 1994–2001. He has authored 280+ papers in international journals/conferences and 10 book chapters. He is also editor/co-editor of 6 international books. He is Foreign Member of Royal Academy of Doctors, Barcelona, Spain (from 2004). He received Gold Medal "Henry Coanda", Rumanian Academy for Researches in Neural Networks and Fuzzy Systems, Iasi, Rumania in 2003. He was President of the Italian Society of Neural Networks (SIREN) from 2008–2014. He also held positions in INNS: Member of the Board of Governors (2000–2002); Secretary (2003); Member of the Board of Governors (2004–2006; 2007–2009; 2010–2012); Chair of the Nomination Committee (2010–2012). He is Associate Editor of Neural Networks (Elsevier); Editorial Board Member of many reputed journals.

**Eros Pasero** is Professor of Electronics at the Politecnico of Turin since 1991. Previously, he was Professor at the University of Roma, Visiting Professor at ICSI, UC Berkeley, CA, Professor of digital electronics and electronic systems at Tongji University, Shanghai, China in 2001 and "Electronic Devices" in 2015 and Professor of digital electronics and electronic systems at TTPU (Turin Tashkent Politechic University), Tashkent, Uzbekistan. His area of research interests are in Artificial Neural Networks and Electronic Sensors.

He is now the President of SIREN, the Italian Society for Neural Networks; he was General Chairman of IJCNN2000 in Como, General Chairman of SIRWEC2006 in Turin and General Chairman of WIRN 2015 in Vietri. He has received several awards and holds 5 international patents. He was supervisor of tens of international PhD's and hundreds of Master students. He has authored more than 100 international publications. He is involved in several funded research projects.

# **Part I**

## **Introduction**

# Chapter 1

## Some Note on Artificial Intelligence



Anna Esposito, Marcos Faundez-Zanuy, Francesco Carlo Morabito  
and Eros Pasero

**Abstract** This introductory chapter discusses some basics of artificial intelligence and in particular some theoretical issues concerning neural networks that are still open problems and need further investigations. This is because, independently of the large use of neural networks in several fields, using neural networks and similar artificial intelligence techniques to solve problems of non-polynomial complexity is by itself a creative and intelligent problem, not rigidly tied to procedural methods and fully explainable on theoretical bases.

### 1.1 Introduction

The term artificial intelligence was coined by McCarthy et al. [6] in the attempt to collect funds to organize a series of seminars. In the presentation of the project McCarthy et al. wrote: “*We propose.. a ....study of Artificial Intelligence on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it*” [6]. Artificial intelligence, therefore, as it was initially defined, aimed to identify machine learning algorithms and procedures that could complement and even replace the

---

A. Esposito (✉)

Dipartimento di Psicologia and IIASS, Università degli Studi della Campania Luigi Vanvitelli, Caserta, Italy

e-mail: [iiass.annaesp@tin.it](mailto:iiass.annaesp@tin.it)

M. Faundez-Zanuy

Pompeu Fabra University, Barcelona, Spain

e-mail: [faundez@tecnocampus.cat](mailto:faundez@tecnocampus.cat)

F. C. Morabito

Università degli Studi “Mediterranea” di Reggio Calabria, Reggio Calabria, Italy

e-mail: [morabito@unirc.it](mailto:morabito@unirc.it)

E. Pasero

Politecnico di Torino - Dip. Elettronica e Telecomunicazioni, Turin, Italy

e-mail: [eros.pasero@polito.it](mailto:eros.pasero@polito.it)

man in the execution of tasks requiring the intervention of “intelligence.” However, establishing that certain tasks require some “intelligence” has proved to be a difficult issue, since a comprehensive understanding of the concept of “intelligence” has not yet been formulated and may remain an “ill-posed problem” for the time to come. There are, however, skills or abilities which, although do not fully provide an all-inclusive rendering of the concept of intelligence, yet can be considered as important components of it. We may agree with Rhine [7] that, some essential features of intelligence are as follows:

- react flexibly to various situations;
- take advantage of fortuitous circumstances;
- recognize the relative importance of different elements in a situation;
- find similarities in situations despite of their differences;
- synthesize new concepts connecting old ones in new ways;
- produce new ideas
- ...more.

Nevertheless, when such features are going to be incorporated into a machine algorithm or an automatic procedure one comes across an apparent paradox. Machines, and in particular computers, adopt intrinsically rigid behaviors. However fast they may be, they represent the very essence of unawareness, and therefore, how can they plan intelligent behaviors? The basic belief in artificial intelligence is that there is no contradiction in the idea of programming a computer to be intelligent; i.e., it is, in principle, perfectly conceivable, to teach a machine through precise instructions to be flexible and creative.

The fundamental characteristic that distinguishes artificial intelligence from other disciplines—such as psychology, philosophy, linguistics, logic—which investigate aspects and forms of intelligent behaviors, lies in the methodological approach, best defined as “computational approach.” The computational approach requires that hypotheses and theories constituting artificial intelligence (henceforth AI) must be expressed in the form of efficient computational procedures; i.e., the description of such procedures must be so explicit and articulated that they can be translated into a computer program. The two basic concepts to which AI refers are *learning* and *adaptation* and, as previously stated, AI tries to formalize these concepts in rigorous mathematical models. A category of learning that has been easily formalized in a rigorous mathematical model is “*learning from examples*” [8]. Once procedures to automatically learn from examples have been formally defined, AI proposes to identify a system that can implement such learning behaviors. Among the several systems proposed by the literature, neural networks have proved to be the most successful and particularly efficient in learning by examples. Neural networks are composed of many nonlinear computational elements that operate in parallel. Instead of executing a set of elementary instructions, neural networks simultaneously explore different hypotheses by taking advantage of the implicit parallelism present in their architecture which consists of many simple elements connected through weighted connections. The computational elements of a neural network typically perform nonlinear functions, the simplest of which is a weighted sum of the filtered input through

a nonlinear function called “*activation function*”. A neural network is completely specified by:

- (a) the architecture or network topology;
- (b) the activation functions of its nodes in the different layers;
- (c) the type of learning algorithm the network uses and the way it learns (supervised or unsupervised).

There are several theoretical aspects that must be accounted for the correct functioning of a neural network the most important are as follows:

- (a) Avoid the overfitting problem which is tied to the dimensioning of the network architecture and complexity of the task assigned to the net. Overfitting gives rise to an undersized or oversized network with respect to the number of needed nodes for the task at the hand and affect the network performances. To overcome overfitting, it is necessary not only to understand how many nodes and how many layers a network must have in order to solve the assigned task, but also ensure that the activation functions of the nodes in the various layers possess certain properties. For example, while the functions of the first layer may be strictly monotonous, those of the second layer should be chosen depending on the complexity of the task. Overfitting produces effects similar to those observed when polynomials with a degree higher than the number of available points are used to interpolate a function. The network recognizes exactly the examples seen. Nevertheless, for those not seen it responds with a behavior far from the current data distribution. This involves the loss of the network’s ability to generalize, which represents one of the most important properties making neural networks appealing to solve problems of non-polynomial computational complexity. How to size the network architecture is a problem still handled through trial and error processes, advocating the need of further theoretical investigations.
- (b) Avoid the overtraining problem. When the set of training data is presented to the network for a large number of epochs, the error committed by the network on such data (training error) decreases. However, the decrease in the training error does not imply that in turn the generalization error decreases. The network tends to specialize on the training set, losing the generalization objective. Heuristics have been suggested to avoid the overtraining problem, consisting in introducing a further set of data, the validation set, which act as a testing set for deciding when to stop the training. However, when fewer data are available this becomes a problem. Is there any other theoretical approach helping to avoid overtraining? More investigations are needed.
- (c) The input/output data encoding. Very often, the input to a neural network is not in a suitable form and it is necessary to work on it with transformations. Consider, for example, a voice or a video signal. Even quantized and sampled, it is practically unconceivable to consider giving in input to the network these signals in their coarse digital representation. Similarly, the output of a neural network does not always come in the form of a direct response to the problem

under consideration and may require an interpretation. Consequently, pre-processing and post-processing the network inputs and outputs play a fundamental role for the network's accuracy.

With deep learning algorithms [1, 3–5], it seems that the preprocessing problem has been solved. However, the advantage of using deep learning is that networks can handle large amounts of input data. More huge are these data, more work for annotating them is required, and less benefits are gained because the time required to annotate large amount of data, as for example satellites or geophysical data, reduces the advantage of eliminating the preprocessing phase. In addition, deep learning algorithms do not satisfy the European Union guidelines on developing ethical AI systems: “*AI systems should be accountable, explainable, and unbiased*” since it remains still unclear how data are processed along the deep layering of the corresponding network architecture [9]. More theoretical investigations are required in the future.

- (d) The convergence of a learning algorithm is one of the most important aspect to ensure that the network can achieve good performances in solving a given problem. There are no theoretical bases that can always guarantee the convergence, and heuristics are exploited depending on both the available computational resources and values used to initialize the network's parameters. For example, the value of the learning rate must not be very small otherwise, learning (and therefore convergence) is slow. However, the learning rate cannot even take very large values otherwise, the network does not converge to the required solution. Theoretical procedures to infer the optimal value of the learning rate do not exist. Nevertheless, it has experimentally observed that the optimal initial learning rate value depends on network's dimensions, network's architecture, used learning algorithm, size of the training data, number of epochs for which the net is trained, and error committed to each training cycle. Another parameter that plays an important role in ensuring the network convergence is the way weights are initialized on the connections between one node and another. Also in this case, it is necessary to take into account the architecture of the network, the learning algorithm, distinguish between recurring and non-recurring network models, and more. Studies providing accurate theoretical bases for determining the correct learning rate value or the correct way to initialize weights should be taken into consideration in the future.

## 1.2 Content of This Book

The themes of this book tackled aspects of dynamics of signal exchanges either processed by natural or artificial neural networks. The volume is organized in sections, one dealing with very general neural network applications and the remaining with more specific issues associated to social exchanges and pattern recognition procedures in medicine. The contributions collected in this book were initially discussed

at the International Workshop on Neural Networks (WIRN 2018) held in Vietri sul Mare, Italy, from June 13 to 15, 2018. The workshop, being at its 29th edition, is nowadays a historical and traditional scientific event gathering together researcher on artificial intelligence and human–machine interaction from Europe and overseas.

**Section I** is introductory and contains this paper [2].

**Section II** is dedicated to neural networks and their related applications in several research fields. It includes 21 original contributes.

**Section III** describes the use of neural networks and pattern recognition techniques in medicine. This section includes 7 chapters discussing on advanced artificial intelligent methods for supporting healthcare services.

**Section IV** discusses themes multidisciplinary in nature and closely connected in their final aims to identifying features from realistic dynamics of signal exchanges. Such dynamics characterize formal and informal social signals, communication modes, hearing and vision processes, and brain functionalities. The section includes 17 chapters.

### 1.3 Conclusions

Taking all of the above discussion in mind, we can conclude that using neural networks and similar artificial intelligence techniques to solve problems of non-polynomial computational complexity is by itself a creative and intelligent problem, not rigidly tied to procedural methods. We expect that contributors of this book who use intelligence (natural or artificial) to solve daily problems would consider in the future investigating more on the highlighted theoretical issues.

#### Acknowledgements



The research leading to these results has received funding from the European Union Horizon 2020 research and innovation programme under grant agreement N. 769872 (EMPATHIC) and N. 823907 (MENHIR) and from the project SIROBOTICS that received funding from Ministero dell’Istruzione, dell’Università, e della Ricerca (MIUR), PNR 2015–2020, Decreto Direttoriale 1735 del 13 luglio 2017.

## References

1. Bengio, Y., Thibodeau-Laufer, E., Alain, G., Yosinski, J.: Deep generative stochastic networks trainable by backprop. In: Proceeding 31st International Conference on Machine Learning, pp. 226–234 (2014)
2. Esposito, A., Faundez-Zanuy, M., Morabito, F.C., Pasero E.: Some note on artificial intelligence. This volume (2019)
3. Graves, A., Mohamed, A.R., Hinton, G.: Speech recognition with deep recurrent neural networks. In: Proceeding International Conference on Acoustics, Speech and Signal Processing pp. 6645–6649 (2013)
4. Hinton, G., et al.: Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Process. Mag. **29**, 82–97 (2012)
5. Krizhevsky, A., Sutskever, I., Hinton, G.: ImageNet classification with deep convolutional neural networks. In: Proceeding of Advances in Neural Information Processing Systems, vol. 25, pp. 1090–1098 (2012)
6. McCarthy, J., Minsky, M.L., Rochester, N., Shannon, C.E.: A proposal for the Dartmouth summer research project on artificial intelligence. AI Mag. **27**(4), 12–14 (2006)
7. Rhines, W. (1985) AI: Out of the lab and into business. J. Bus. Strat. Summer **6**(1), 50–57 (1985)
8. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature **323**, 533–536 (1986)
9. Vincent, J.: <https://www.theverge.com/2019/4/8/18300149/eu-artificial-intelligence-ai-ethical-guidelines-recommendations> (2019)

**Part II**

**Neural Networks and Related Applications**

# Chapter 2

## Music Genre Classification Using Stacked Auto-Encoders



Michele Scarpiniti, Simone Scardapane, Danilo Comminiello  
and Aurelio Uncini

**Abstract** In this paper, we propose an architecture based on a stacked auto-encoder (SAE) for the classification of music genre. Each level in the stacked architecture works by stacking some hidden representations resulting from the previous level and related to different frames of the input signal. In this way, the proposed architecture shows a more robust classification compared to a standard SAE. The input to the first level of the SAE is fed by a set of 57 peculiar features extracted from the music signals. Some experimental results show the effectiveness of the proposed approach with respect to other state-of-the-art methods. In particular, the proposed architecture is compared to the support vector machine (SVM), multi-layer perceptron (MLP) and logistic regression (LR).

### 2.1 Introduction

A very challenging problem that has attracted researchers from the last decades is the automatic classification of the genre of music tracks [15]. The importance in classification of the music genre is to allow suppliers, simple users, and Web sites collecting music tracks, to organize the file collections in a simple and searchable form. Hence, music genres are the main top-level descriptors among a plethora of possible ones. The main difficulty in this problem is that music genres are categories that are strongly dependent of an interplay of cultures, artists and market forces: The

---

M. Scarpiniti (✉) · S. Scardapane · D. Comminiello · A. Uncini

Department of Information Engineering, Electronics and Telecommunications (DIET),  
“Sapienza” University of Rome, Rome, Italy

e-mail: [michele.scarpiniti@uniroma1.it](mailto:michele.scarpiniti@uniroma1.it)

S. Scardapane

e-mail: [simone.scardapane@uniroma1.it](mailto:simone.scardapane@uniroma1.it)

D. Comminiello

e-mail: [danilo.comminiello@uniroma1.it](mailto:danilo.comminiello@uniroma1.it)

A. Uncini

e-mail: [aurelio.uncini@uniroma1.it](mailto:aurelio.uncini@uniroma1.it)

differences between genres remain blurry for definition and make the considered problem of nontrivial solution [10, 15, 18]. The problem of music genre classification can be extended to the related automatic segmentation and classification of broadcast audio news [3, 19].

So far, the approaches used to solve such a problem are based on different methods [5]. Specifically, there exist some unsupervised approaches, where music files are clustered based on a suitable cost function in order to dynamically build a peculiar taxonomy depending on the clustering outcome [16]. The main drawback relies on the lack of the class label and so the need of experts to construct a good taxonomy. On the other hand, most common approaches are based on supervised methods [15]. Primarily, attention has been focused on support vector machine (SVM) classifiers [9], kNN classifiers [11], extreme learning machines (ELMs) [14] or other artificial neural networks approaches [5]. Also, ensemble of different methods has been proposed [17]. Finally, some semi-supervised approaches have been proposed [13] in order to outperforms a number of standard supervised learning techniques.

As a general approach, all the considered classifiers work on a set of suitable features extracted from the musical signal [10, 12, 18]. Both temporal and spectral features have been here considered. However, although several works exist on the problem, no unique results have been found [7]. These results are very sensitive to the particular features used and the specific dataset used.

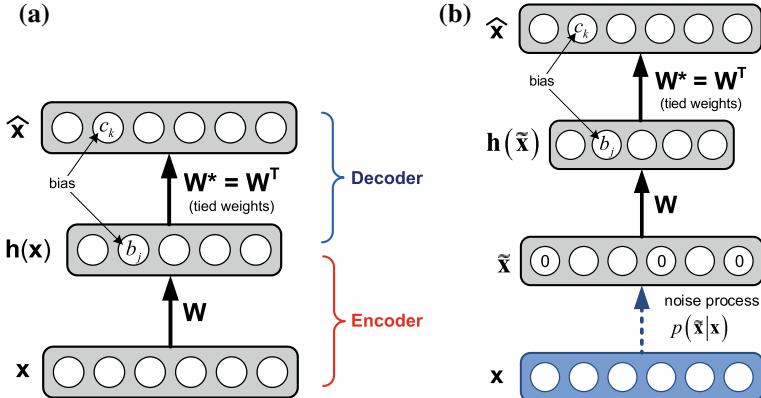
Recently, several approaches coming from the deep learning community have been applied to the musical genre classification [4]. From these approaches, those based on the auto-encoder (AE) architectures seem to be very promising [1, 6, 20]. Specifically, an architecture obtained by stacking different AEs, called stacked AE (SAE) is proposed to construct a powerful high-level representation of the input features extracted from the music data [21].

In this paper, we propose a SAE architecture able to solve the challenging problem of music genre classification. Differently from the classical definition of SAEs, the proposed architecture involves input layers of the internal levels of the SAE constructed by concatenating two or more hidden representations of previous levels in the stacked architecture.

The rest of the paper is organized as follows. In Sect. 2.2, we briefly introduce the AE architecture, while Sect. 2.2.1 describes the idea of SAE, and Sect. 2.2.2 introduces the chosen classification scheme. The description of the features used in the work is provided in Sect. 2.3. Finally, we validate our approach in Sect. 2.4 and we conclude with some final remarks in Sect. 2.5.

## 2.2 The Proposed Architecture

An *auto-encoder* (AE) is a feedforward neural network trained to *reproduce* its input at the output layer [1, 2, 8]. The aim of an auto-encoder is to learn a compressed and distributed representation  $\mathbf{h}(\mathbf{x})$  (*encoding*) for a set of data  $\mathbf{x}$  (typically for the purpose of dimensionality reduction, like in the case of PCA), using a weight



**Fig. 2.1** Auto-encoder (AE) (a) and denoising auto-encoder (DAE) (b) architectures

matrix  $\mathbf{W}$ . Then, an estimate  $\hat{\mathbf{x}}$  of data is reconstructed using tied weights (*decoding*), as depicted in Fig. 2.1a. The AE is called undercomplete if the hidden layer is smaller than the input layer. In this case, the hidden vector  $\mathbf{h}$ , also called *embedding*, is a good compressed representation of the input  $\mathbf{x}$ . Hence, hidden units will be good features for the training distribution.

As shown in Fig. 2.1a, the encoding is performed as:

$$\mathbf{h}(\mathbf{x}) = \sigma(\mathbf{b} + \mathbf{W}\mathbf{x}), \quad (2.1)$$

while the decoding is obtained as follows:

$$\hat{\mathbf{x}} = \sigma(\mathbf{c} + \mathbf{W}^T \mathbf{h}(\mathbf{x})), \quad (2.2)$$

where  $\mathbf{W}$  is the weight matrix,  $\mathbf{b}$  and  $\mathbf{c}$  are the bias vectors of the hidden and output layers, and  $\sigma(\cdot)$  is the element-wise sigmoid activation function.

As a powerful variant, the *denoising auto-encoder* (DAE) is a stochastic version of the AE where a stochastically corrupted  $\tilde{\mathbf{x}}$  version of the input  $\mathbf{x}$  is used to feed the AE (usually using a Gaussian additive noise), but the uncorrupted input  $\mathbf{x}$  is still used as target for the reconstruction (see Fig. 2.1b). Intuitively, a DAE tries to: (i) encode the input (preserve the information about the input) and (ii) undo the effect of a corruption process stochastically applied to the input of the AE. An interesting property of the DAE is that it corresponds to a generative model and it naturally lends itself to data with missing values.

The *loss function* to be minimized in order to train the AE/DAE is usually the sum of squared differences:

$$\mathcal{L}(\hat{\mathbf{x}}) = \frac{1}{2} \sum_k |\hat{\mathbf{x}}_k - \mathbf{x}_k|^2, \quad (2.3)$$

where  $\mathbf{x}_k$  is the  $k$ th batch of the input vector. In our experiments, we optimize (2.3) which is obtained by a stochastic gradient descent with a momentum term  $\alpha$  and a regularization factor  $\delta$  [6].

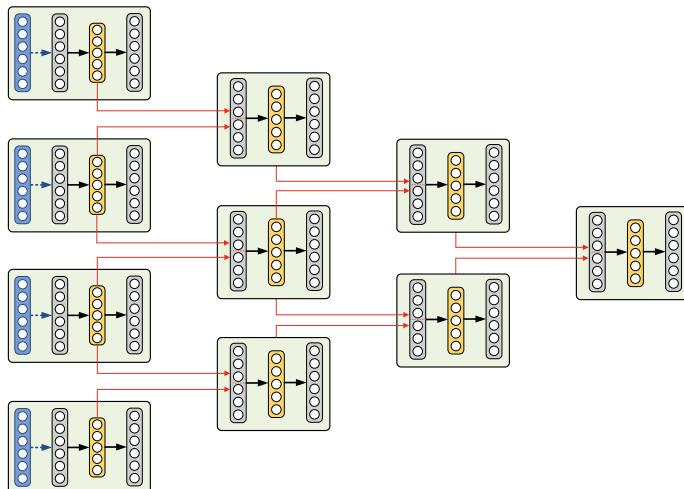
### 2.2.1 The Stacked Auto-Encoder

DAE can be stacked to form a deeper network by feeding the latent representation  $\mathbf{h}$  of the single DAEs found on the layer below as input to the current layer. This kind of deep architecture is known as *stacked AE* (SAE) [20, 21].

In order to improve the performance of the stacked network, in the proposed work the input to each internal layer is constructed by stacking the hidden representations of two or more DAEs of the previous layer, as shown in Fig. 2.2, for the case of four layers and concatenating only two hidden representations. Each layer of the SAE is here also called level.

### 2.2.2 Classification

The latent representation of each layer of the SAE architecture is used to feed the input of a classifier based on the support vector machines (SVMs). In this work, we consider a classifier for each of the SAE layer in order to track the effect of the



**Fig. 2.2** A four-level stacked auto-encoder (SAE) architecture

number of levels. All results are shown in terms of classification accuracy. An SVM with a radial basis kernel with parameter  $\gamma$  and soft-margin  $C$  is used.

The parameters  $\gamma$  and  $C$  of the SVM classifiers are optimized by a grid search algorithm in the sets:  $\gamma \in \{0.1, 0.2, \dots, 3.9, 4\}$  and  $C \in \{10^{-3}, 10^{-2}, \dots, 10^6, 10^7\}$ , respectively. It has been found that the optimal parameters are  $\gamma = 1$  and  $C = 100$  for all the layers.

## 2.3 Feature Extraction

The proposed stacked architecture of Sect. 2.2.1 is evaluated on an input vector  $\mathbf{x}$  composed by collecting several features from the audio signal [10, 18, 20]. Such features should be sufficiently discriminatory in order to obtain a successful classification. A total of  $M = 57$  features, grouped into eight types, are considered [12]. Specifically, in this paper we have chosen the following features:

- *Zero-Crossing Rate* (ZCR): one coefficient that evaluates the zero-crossing rate of an audio time series.
- *Root Mean Square Energy* (RMSE): one coefficient that evaluates the RMS energy for each frame of the audio samples.
- *Spectral Centroid* (SC): one coefficient that evaluates the mean (centroid) of each frame of the magnitude spectrogram, normalized and treated as a distribution over frequency bins.
- *Spectral Bandwidth* (SBW): one coefficient that evaluates the  $p$ th-order spectral bandwidth, defined as:

$$\text{SBW}_p = \sqrt[p]{\sum_k |S[k]| (f[k] - c)^p}, \quad (2.4)$$

where  $|S[k]|$  is the spectrogram magnitude,  $f[k]$  the  $k$ th frequency bin and  $c$  the spectral centroid.

- *Spectral Roll-off* (SRO): one coefficient that evaluates the roll-off frequency.
- *Chroma STFT* (CSTFT): twelve coefficients that evaluate the chromagram from the power spectrogram.
- *Mel-Frequency Cepstral Coefficients* (MFCC): twenty coefficients that evaluate the Mel-frequency cepstrum.
- *Delta MFCC* (DMFCC): twenty coefficients that evaluate the delta features, a local estimate of the derivative of the input data along frames.

All the considered features are evaluated by using the LibROSA Python library.<sup>1</sup> Data obtained from these  $M$  features for  $P$  total frames are represented as a  $P \times M$  matrix. In addition, since the different features present a huge variation between them, all values have been normalized into the interval  $[-0.8, 0.8]$ .

---

<sup>1</sup>The library can be downloaded from: <https://librosa.github.io/librosa/feature.html>.

## 2.4 Experimental Results

In the proposed experimental results, we use a four-level stacked architecture: Each of the inputs to the internal levels is obtained by concatenating two hidden representations of the previous level. In this set of experiments, we consider five musical genres: blues, jazz, rap, reggae, and rock. A total of  $N_f = 25$  sound files per class have been manually selected from a personal database. All music tracks (in WAV format), sampled at  $F_s = 44.1$  kHz, have been transformed into monophonic tracks, by simply taking the mean between channels. The considered dataset is quite well balanced, since each class presents a duration of about one hour and half. Table 2.1 summarizes the main characteristics of the used dataset. 80% of the dataset is used for training while the remaining 20% for testing. A fivefold cross-validation is used.

The input signals have been segmented into frames of duration  $T = 500$  ms. Subsequently, the  $M = 57$  features described in Sect. 2.3 have been computed from  $P = 200$  frames randomly selected from each file in order to form a batch feeding the proposed architecture. A total of  $N_e = 1600$  epochs are run. The number  $N_h$  of hidden units in the internal layer of each AE is a suitable fraction  $\eta < 1$  of the input ones.

The SAE is optimized by setting the learning rate to  $\mu = 10^{-4}$ , the momentum factor to  $\alpha = 0.5$ , and the regularization term to  $\delta = 0.005$ , for all the four levels. The learning rate  $\mu$  has been obtained by a grid search algorithm in the range  $\mu \in [10^{-6}, 5 \times 10^{-6}, 10^{-5}, 5 \times 10^{-5}, \dots, 10^{-1}]$ .

Results of the proposed approach are evaluated in terms of the accuracy of classification. Accuracy is defined as the number of correctly classified songs divided by the total number of songs. Results in terms of accuracy for the proposed four-level SAE are shown in Table 2.2 for two different values of the frame size  $T$ . Specifically, we used a frame size of  $T = 0.5$  s and  $T = 1$  s, respectively. Table 2.2 indicates also the number of hidden units  $N_h$  of the representation of each level and the standard deviation of the reached accuracy. As we can see from Table 2.2, in both the cases the accuracy is quite high. Generally, a better accuracy is reached at the higher level, while a lower standard deviation is obtained.

In an additional simulation, we changed the number of concatenated hidden representations. Specifically, the second and third levels use as input the concatenation of three hidden vectors from the previous levels, while the four-level use, as usual, two hidden vectors. Results in terms of accuracy for the proposed four-level SAE are

**Table 2.1** Description of the used dataset: class, number of files ( $N_f$ ), and time duration

Class	$N_f$	Duration
BLUES	25	1:14:24
JAZZ	25	1:35:51
RAP	25	1:23:06
REGGAE	25	1:33:28
ROCK	25	1:58:10

**Table 2.2** Results in terms of accuracy for the proposed four-level SAE and a frame size of  $T = 0.5$  and  $T = 1$  second

Level	$T = 0.5$ s		$T = 1$ s	
	$N_h$	Accuracy (%)	$N_h$	Accuracy (%)
1	45	$92.0 \pm 2.83$	45	$92.8 \pm 3.35$
2	67	$93.6 \pm 3.58$	67	$94.4 \pm 2.19$
3	100	$94.4 \pm 2.19$	100	$96.0 \pm 2.83$
4	140	$96.0 \pm 1.79$	140	$96.8 \pm 1.79$

**Table 2.3** Results in terms of accuracy for the proposed four-level SAE and a frame size of  $T = 0.5$  second

Level	$N_h$	Accuracy [%]
1	45	$92.8 \pm 1.79$
2	94	$94.4 \pm 2.19$
3	112	$95.2 \pm 3.34$
4	134	$96.0 \pm 2.83$

shown in Table 2.3. The results presented in Table 2.3 show a behavior very similar to the previous one in the case of a frame size of  $T = 1$  s, with a slightly worse accuracy in the last two levels but still acceptable.

Finally, the proposed architecture has been compared with other state-of-the-art classifiers. Specifically, the results of SAE have been compared with a SVM, a logistic regression (LR), and a multi-layer perceptron (MLP) with a single hidden layer made of 100 units. All these state-of-the-art classifiers use as input all the  $M = 57$  features. Parameters of the LR and those of MLP have been selected by a grid search algorithm over suitable intervals. Specifically, for LR we have optimized the inverse of regularization strength  $C_l$ , in range  $C_l \in \{10^{-1}, 1, 10, \dots, 10^7, 10^8\}$ , while for MLP we have optimized the learning rate  $\mu_m$  in the range  $\mu_m \in \{10^{-6}, 10^{-5}, \dots, 10^{-1}\}$ . Results of comparisons in terms of accuracy, averaged across each fold, are shown in Table 2.4. This table shows that the proposed approach outperforms the other state-of-the-art considered methods. In addition, SVM and MLP perform very similar while the LR presents the poorest results.

**Table 2.4** Results in terms of accuracy for the compared state-of-the-art classifiers

Classifier	Accuracy (%)
SAE	$96.0 \pm 1.79$
SVM	$93.2 \pm 2.83$
MLP	$92.4 \pm 2.19$
LR	$89.1 \pm 3.58$

## 2.5 Conclusions

In this paper, we have proposed a stacked auto-encoder (SAE) network to solve the challenging problem of music genre classification. In particular, we focus our attention on the classification of five different genres based on 57 peculiar features (both temporal and spectral). However, the approach is simply extendible to a greater number of genres and a different number and types of features. Some experimental results, implemented by using a four-level SAE and evaluated in terms of accuracy, have shown the effectiveness of the proposed approach.

## References

1. Bengio, Y.: Learning deep architectures for AI. *Found. Trends Mach. Learn.* **2**(1), 1–127 (2009)
2. Bourlard, H., Kamp, Y.: Auto-association by multilayer perceptrons and singular value decomposition. *Biol. Cybern.* **59**, 291–294 (1988)
3. Castán, D., Ortega, A.A.M., Lleida, E.: Audio segmentation-by-classification approach based on factor analysis in broadcast news domain. *EURASIP J. Audio, Speech, Music. Process.* **2014**(34), 1–13 (2014)
4. Choi, K., Fazekas, G., Cho, K., Sandler, M.: A tutorial on deep learning for music information retrieval [arXiv:1709.04396](https://arxiv.org/abs/1709.04396) (2018)
5. Fu, Z., Lu, G., Ting, K.M., Zhang, D.: A survey of audio-based music classification and annotation. *IEEE Trans. Multimed.* **13**(2), 303–319 (2011)
6. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. The MIT Press (2016)
7. Goulart, A.J.H., Guido, R.C., Maciel, C.D.: Exploring different approaches for music genre classification. *Egypt. Inform. J.* **13**(2), 59–63 (2012)
8. Hinton, G.E., Zemel, R.S.: Autoencoders, minimum description length, and Helmholtz free energy. In: Proceeding of NIPS 1993 (1994)
9. Mandel, M., Ellis, D.: Song-level features and support vector machines for music classification. In: Proceeding of 6th International Symposium on Music Information Retrieval. London, UK (2005)
10. Mierswa, I., Morik, K.: Automatic feature extraction for classifying audio data. *Mach. Learn.* **58**(2–3), 127–149 (2005)
11. Pampalk, E., Flexer, A., Widmer, G.: Improvements of audio based music similarity and genre classification? In: Proceeding of 6th International Symposium on Music Information Retrieval. London, UK (2005)
12. Patsis, Y., Verhelst, W.: A speech/music/silence/garbage/ classifier for searching and indexing broadcast news material. In: Proceeding of 19th International Workshop on Database and Expert Systems Application (DEXA '08). Turin, Italy (2008)
13. Poria, S., Gelbukh, A., Hussain, A., Bandyopadhyay, S., Howard, N.: Music genre classification: A semi-supervised approach. In: Proceeding of the Mexican Conference on Pattern Recognition (MCPR 2013), pp. 254–263 (2013)
14. Scardapane, S., Comminello, D., Scarpiniti, M., Uncini, A.: Music classification using extreme learning machines. In: 8th International Symposium on Image and Signal Processing and Analysis (ISPA2013), pp. 377–381. Trieste, Italy (2013)
15. Scaringella, N., Zoia, G., Mlynec, D.: Automatic genre classification of music content: a survey. *IEEE Signal Process. Mag.* **23**(2), 133–141 (2006)
16. Shao, X., Xu, C., Kankanhalli, M.: Unsupervised classification of musical genre using hidden Markov model. In: IEEE International Conference of Multimedia Explore (ICME 2004). Taiwan (2004)

17. Silla, C.N., Kaestner, C.A., Koerich, A.L.: Automatic music genre classification using ensemble of classifiers. In: IEEE International Conference on Systems, Man and Cybernetics, pp. 1687–1692 (2007)
18. Tzanetakis, G., Cook, P.: Musical genre classification of audio signals. *IEEE Trans. Speech Audio Process.* **10**(5), 293–302 (2002)
19. Vavrek, J., Vozáriková, E., Pleva, M., Juhár, J.: Broadcast news audio classification using SVM binary trees. In: Proceeding of the 35th International Conference on Telecommunications and Signal Processing (TSP 2012) (2012)
20. Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P.A.: Extracting and composing robust features with denoising autoencoders. In: Proceedings of the Twenty-fifth International Conference on Machine Learning (ICML'08), pp. 1096–1103 (2008)
21. Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A.: Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* **11**, 3371–3408 (2010)

# Chapter 3

## Linear Artificial Forces for Human Dynamics in Complex Contexts



Pasquale Coscia, Lamberto Ballan, Francesco A. N. Palmieri, Alexandre Alahi and Silvio Savarese

**Abstract** In complex contexts, people need to adapt their behavior to interact with the surrounding environment to reach the intended destination or avert collisions. Motion dynamics should therefore include both social and kinematic rules. The proposed analysis aims at defining a linear dynamic model to predict future positions of different types of agents, namely pedestrians and cyclists, observing a limited number of frames. The dynamics are defined in terms of artificial potentials fields (APFs) obtained by static (e.g., walls, doors or benches) and dynamic (e.g., other agents) elements to produce attractive and repulsive forces that influence the motion. A linear combination of such forces affects the resulting behavior. We exploit the context using a semantic scene segmentation to derive static forces while the interactions between agents are defined in terms of their reciprocal physical distances. We conduct experiments both on synthetic and on subsets of publicly available datasets to corroborate the proposed model.

---

P. Coscia (✉) · F. A. N. Palmieri

Dipartimento di Ingegneria, Università della Campania, via Roma, 29, Aversa, Italy

e-mail: [pasquale.coscia@unicampania.it](mailto:pasquale.coscia@unicampania.it)

F. A. N. Palmieri

e-mail: [francesco.palmieri@unicampania.it](mailto:francesco.palmieri@unicampania.it)

L. Ballan

Dipartimento di Matematica, Università di Padova, via Trieste, 63, Padova, Italy

e-mail: [lamberto.ballan@unipd.it](mailto:lamberto.ballan@unipd.it)

A. Alahi

École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

e-mail: [alexandre.alahi@epfl.ch](mailto:alexandre.alahi@epfl.ch)

S. Savarese

Department of Computer Science, Stanford University, Stanford, CA 94305, USA

e-mail: [silvio@stanford.edu](mailto:silvio@stanford.edu)

### 3.1 Introduction

The prediction of human trajectories both at short and long term, along with their intentions, is still a challenging task even considering the most advanced technologies and methods. The majority of problems arises from the understanding of non-verbal cues which are very complicated to model. Both physiological (related to motion) and psychological (related to personal attributes) aspects also play key roles in such a task. Mimicking such aspects could, therefore, support us in a wide range of real-world applications, from public safety, to recognize imminent collisions, to anomaly detection, to unveil potentially hazardous situations, and in robotics, to enhance social robots capabilities when they interact with humans.

Images provided by RGB cameras, which are largely used for security reasons or space optimization, typically capture complex human behaviors and the prediction of what will happen in the next seconds could be exploited to improve the quality of life, or to provide a rapid support in case of danger. Many factors must be considered, such as gained experience, obstacle avoidance, attractive elements such as vending machines and exits. In fact, humans are mostly guided by intentions, social rules, and common sense which take them to prefer certain paths. Dynamics are also influenced by the environment's perception. Complex situations may not—or may not completely—rely on classical identification system techniques.

We rely on attractive and repulsive forces obtained by artificial potential fields which influence human behaviors. We consider three main elements that define the underlying structure of our model: *agent dynamics*, which refer to the kinematic model in terms of position and velocity; *space perception*, in terms of obstacles and attractive elements and past gained experience, and an *interaction factor* which represents the reaction to the environmental stimuli or changes. Our aim is to predict future positions of two types of targets, namely pedestrians and cyclists, observing several frames.

Our preliminary results corroborate the effectiveness of the proposed procedure to predict future human positions compared to a constant velocity model. It is worth noting that no assumptions regarding the type of dynamics are made. Depending on the available data, the model could capture different types of motions, from disordered acting typical of panic situations to more ordinary situations. In this work, we refer to urban scenarios where people walk to reach the entrance of a building or their intended destinations and cyclists crossing road intersections.

The article is outlined as follows. Section 3.2 recalls previous work on human trajectory forecasting. Sections 3.3 and 3.4 describe the proposed approach and detail each factor. Section 3.5 describes the research results. Section 3.6 summarizes the method and discusses future work.

### 3.2 Related Work

Human trajectory forecasting is an active research area whether we consider individuals, small groups of people, or crowded spaces such as shopping malls, parking lots, or airports, where thousands of people walk simultaneously in any direction.

The first attempt to model human dynamics relies on artificial potential fields widely used in robotics [1], simulation [2] and computer vision [3] during the last decades. It assumes the existence of a field of forces which fills the environment attracting the agent toward the destination avoiding collisions with obstacles through imaginary repulsive surfaces. Such idea was used in Ref. [4] for collision avoidance of robotic manipulators in real-time applications, and it has been investigated and enhanced since then to overcome well-known issues, such as local minima [5], and more recently, oscillations between multiple obstacles and goals non-reachable with obstacles nearby [6].

Such approaches have been inspired by the well-known "Social Force Model" [7] which describes movements guided by external forces. Such model has also been combined with different techniques for several tasks. For example, in Ref. [8], abnormal behaviors in crowded scenarios are detected using both a generative model and force flows.

Several energetic methods have also been proposed to solve the trajectory forecasting task. In Ref. [9], an energy minimization approach is exploited both for behavioral prediction and tracking. In Ref. [10], the behavior of a large number of people is analyzed in real situations such as at entrances of concert halls or immigration desks. People are treated as particles and integrated with psychological and sociological aspects related to behaviors and actions.

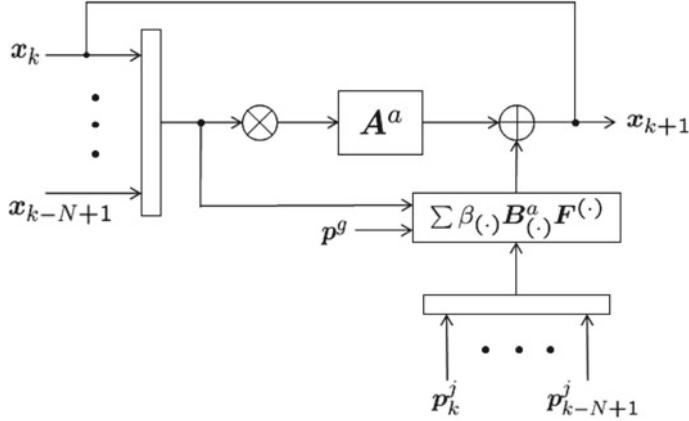
Several works have been inspired by recent deep learning techniques. Alahi et al. [11] connect long–short-term memories (LSTMs) of each agent by a pooling layer based on spatial distances to learn movements and predict future positions. Bartoli et al. [12] propose a similar approach enhancing the prediction with context information. In Ref. [13], generative adversarial networks (GANs) are used to describe inherently multimodal paths with a more sophisticated pooling mechanism which is also based on relative positions.

### 3.3 Dynamic Model

To define human dynamics, we rely on the following linear dynamic model:

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k) + g(\mathbf{i}_k) + \mathbf{w}_k, \quad (3.1)$$

The agent is assumed to have no shape and to be represented by the 4D vector  $\mathbf{x}(t) = [\mathbf{p}(t), \dot{\mathbf{p}}(t)]^T = [x(t), y(t), \dot{x}(t), \dot{y}(t)]^T$ , where dot represents the first-order derivative with respect to time  $t$ , and  $\mathbf{i}_k$  are the inputs at time  $t_k$ . The function  $f(\cdot)$



**Fig. 3.1** Graphical representation of the proposed model. It consists of a closed-loop system where the estimated state  $x_{k+1}$  is used as new input. Matrices  $A^a$ ,  $B_{(\cdot)}^a$  and  $F^{(\cdot)}$  along with the scalars  $\beta_{(\cdot)}$  are computed during the training phase. The value  $N$  represents the number of past observations on which the estimated state depends

is the state-transition function while the function  $g(\cdot)$  is the sum of four factors. More specifically, in our framework,  $i_k$  can be associated to the agent's space perception related to both static and dynamic elements. The process  $w$  is a 4D Gaussian distributed random sequence that accounts for model uncertainty. Assuming linear dynamics with respect to the previous state, the system equations take form:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{F}_k^{tot}(\mathbf{p}_k, \mathbf{p}_g, \mathbf{p}_k^j) + \mathbf{w}_k; \quad \mathbf{x}_0 = \mathbf{x}(0), \quad (3.2)$$

where  $\mathbf{A} \in \mathbb{R}^{4 \times 4}$  is the state-transition matrix,  $\mathbf{x}_0$  is the estimated agent state, i.e., initial position and velocity,  $\mathbf{F}_k^{tot} \in \mathbb{R}^{2 \times 1}$  is related to the space perception at time  $t_k$  while  $\mathbf{p}_g$  and  $\mathbf{p}_k^j$  represent the destination and other agents' positions, respectively. The matrix  $\mathbf{B} \in \mathbb{R}^{4 \times 2}$  is used to affect position and/or velocity components. The model is depicted in Fig. 3.1 and it will be described in detail, along with each factor of the term  $\mathbf{F}_k^{tot}$  in Eq. (3.2) in the following section.

### 3.4 Artificial Forces

**Goal.** The destination of an agent can represent an attractive source toward which the agent feels the desire to move. The closer he/she is, the greater is the need to reach the final position. To model the interest toward such point, given the current position  $\mathbf{p}_k$  at time  $t_k$ , we define an attractive potential based on euclidean distance as:

$$U_k^g(\mathbf{p}_k, \mathbf{p}_g) = \frac{1}{2} \beta_g \rho^m(\mathbf{p}_k, \mathbf{p}_g), \quad \rho(\mathbf{p}_k, \mathbf{p}_g) = \|\mathbf{p}_g - \mathbf{p}_k\|_2. \quad (3.3)$$

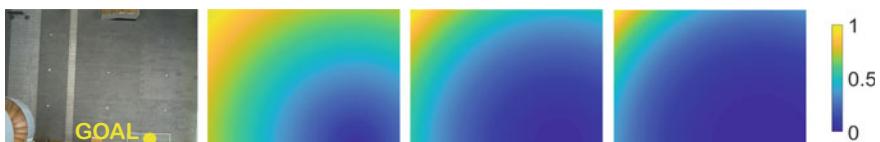
The forces are computed then as:

$$\mathbf{F}_k^g(\mathbf{p}_k, \mathbf{p}_g) = -\nabla_{\mathbf{p}_k} U_k^g(\mathbf{p}_k, \mathbf{p}_g) = \frac{m}{2} \beta_g \|\mathbf{p}_g - \mathbf{p}_k\|_2^{m-2} (\mathbf{p}_g - \mathbf{p}_k), \quad (3.4)$$

where  $\nabla_{\mathbf{p}_k}$  is the spatial gradient,  $\mathbf{p}_g = [x_g, y_g]$  represents the 2D destination's coordinates and  $\mathbf{F}_k^g(\mathbf{p}_k, \mathbf{p}_g) : \mathbb{R}^4 \rightarrow \mathbb{R}^2$  are the force components along with the  $x$  and  $y$  directions, respectively. We consider an attractive potential parabolic in shape, i.e.,  $m = 2$ . Figure 3.2 show a graphical representation of such potential field considering different values of  $m$ .

**Static elements.** Walls, flowerbeds, doors, or trees are certainly elements that produce a repulsive effect on human dynamics. Conversely, sidewalks and benches for pedestrians and street for cyclists should act as attractive elements. To derive such potential fields, we manually annotate such elements. In fact, since we consider fixed cameras, there is no need to use automatic segmentation tools which could make errors compared to a human operator. Firstly, we apply a Gaussian low-pass filter to the segmented images in order to blur the elements' contours. The size and the standard deviation of the Gaussian kernel control the fields' influence and have been determined experimentally. These potential fields are then used to extract the forces in both  $x$  and  $y$  directions considering their spatial derivatives, i.e.,  $\mathbf{F}_k^{s_{rep}}(\mathbf{p}_k) = \nabla_{\mathbf{p}_k} \mathbf{O}(\mathbf{p}_k)$  and  $\mathbf{F}_k^{s_{att}}(\mathbf{p}_k) = \nabla_{\mathbf{p}_k} \mathbf{A}_t(\mathbf{p}_k)$ , where  $\mathbf{O}$  and  $\mathbf{A}_t$  represent the maps of the repulsive and attractive static elements. Both functions are defined as follows:  $\mathbf{F}_k^{s_{rep}}(\mathbf{p}_k) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  and  $\mathbf{F}_k^{s_{att}}(\mathbf{p}_k) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ . A linear combination of attractive and repulsive forces due to static objects is finally considered, i.e.,  $\mathbf{F}_k^s(\mathbf{p}_k) = \mathbf{F}_k^{s_{rep}}(\mathbf{p}_k) + \mathbf{F}_k^{s_{att}}(\mathbf{p}_k)$ .

**Dynamic elements.** The agent dynamics are also influenced by other agents in their field of view. They may change their current direction because of such dynamic elements which modify their behavior, for example, according to social etiquettes. We assume that agents do not interact with each other, so they undergo repulsive forces in case of imminent collisions. Similarly to static electrically charged particles, we assume that the total repulsive force acting on a agent due to  $N_d$  agents in his/her neighborhood obeys the superposition principle and is defined in terms of Coulomb's law as follows:



**Fig. 3.2** Graphical representation of the normalized potential field due to the destination's attraction for a scene of the used dataset along with the final position (in yellow). The parameter  $m$  in Eq. 3.3 is set to 1, 2 and 3, respectively (from left to right)

$$\mathbf{F}_k^d(\mathbf{p}_k, \mathbf{p}_k^j) = \sum_{j=1}^N r_{jk}^{-2} [\cos \Omega_k^j \sin \Omega_k^j]^T, r_{jk} = \|\mathbf{p}_k - \mathbf{p}_k^j\|_2, \Omega_k^j = \angle(\mathbf{p}_k - \mathbf{p}_k^j), \quad (3.5)$$

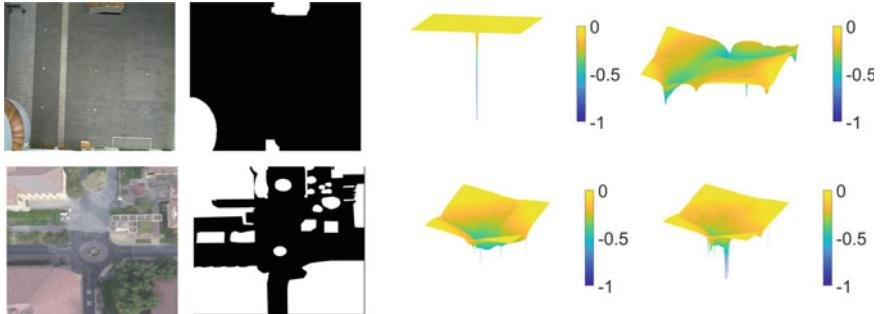
where  $\mathbf{p}_k^j$  is the position of the  $j$ th agent. Such dynamic repulsive force is evaluated only if the distance between two agents is less than a fixed value, i.e.,  $r_{jk} \leq r_{max}$ , and the predicted agent is pointing toward the  $j$ th agent, i.e.,  $(\Omega_k^j + \pi) \in [\angle \dot{\mathbf{p}}_k - \frac{\pi}{3}, \angle \dot{\mathbf{p}}_k + \frac{\pi}{3}]$ . In this way, we define a field of view in order to avoid repulsions due to elements that are behind and cannot be seen. Such function is defined as  $\mathbf{F}_k^d(\mathbf{p}_k, \mathbf{p}_k^j) : \mathbb{R}^{2+2N_d} \rightarrow \mathbb{R}^2$ , where  $N_d$  represents the number of dynamic elements in the agent's neighborhood.

**Past observations.** The knowledge of previous motion may let the dynamic model to better predict new unobserved behaviors. Similarly to the above element, to take into account past observations, we assume that each trajectories' point define a *gravitational* field, like generated by a fictitious homogeneous mass, which acts in any given point of the image plane. Such observation field is obtained, in a point  $\mathbf{p}_k$ , as sum of all the past observed contributes, say  $\mathbf{p}_l$ ,  $l = 1, \dots, L$ , as follows:

$$\mathbf{U}_k^o(\mathbf{p}_k) = - \sum_{i=1}^L \frac{1}{\|\mathbf{r}_i\|_2}, \quad \mathbf{F}_k^o(\mathbf{p}_k) = -\nabla_{\mathbf{p}_k} \mathbf{U}_k^o(\mathbf{p}_k) = - \sum_{i=1}^L \frac{\hat{\mathbf{r}}_i}{\|\mathbf{r}_i\|_2^2}, \quad (3.6)$$

where  $\nabla_{\mathbf{p}_k}$  is the spatial gradient and  $\mathbf{r}_i = \mathbf{p}_k - \mathbf{p}_i$ . A graphical representation of the potential fields generated by both one observation and a group of observations is shown in Fig. 3.3. The described function is defined as  $\mathbf{F}_k^o(\mathbf{p}_k) : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ .

Finally, the total force exerted on the agent, at time  $t_k$ , is a linear combination of the above components, i.e.,  $\mathbf{F}_k^{tot}(\mathbf{p}_k, \mathbf{p}_g, \mathbf{p}_k^j) = \beta_g \mathbf{F}_k^g(\mathbf{p}_k, \mathbf{p}_g) + \beta_r \mathbf{F}_k^{rep}(\mathbf{p}_k) + \beta_a \mathbf{F}_k^{att}(\mathbf{p}_k) + \beta_d \mathbf{F}_k^d(\mathbf{p}_k, \mathbf{p}_k^j) + \beta_o \mathbf{F}_k^o(\mathbf{p}_k)$ , where the scalar potential gains  $\beta_{(.)} \in \mathbb{R}$



**Fig. 3.3** First row shows the monitored area drawn from the EIFPD dataset, the semantic scene segmentation due to obstacles, the potential field's shape due to one observation in the center of the image and the potential field's shape due to observations in the training set. The second row shows such elements for a scene of the SDD dataset, and the potential fields for cyclist and pedestrian target classes, respectively

to define the amount of each component. Hence, the discrete time-invariant state-space model has the following form:

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{F}_k^{tot}(\mathbf{p}_k, \mathbf{p}_g, \mathbf{p}_k^j) + \mathbf{w}_k \\ &= \mathbf{A}\mathbf{x}_k + \beta_g \mathbf{B}_g \mathbf{F}_k^g(\mathbf{p}_k, \mathbf{p}_g) + \beta_r \mathbf{B}_r \mathbf{F}_k^{srep}(\mathbf{p}_k) \\ &\quad + \beta_a \mathbf{B}_a \mathbf{F}_k^{sat}(\mathbf{p}_k) + \beta_d \mathbf{B}_d \mathbf{F}_k^d(\mathbf{p}_k, \mathbf{p}_k^j) + \beta_o \mathbf{B}_o \mathbf{F}_k^o(\mathbf{p}_k) + \mathbf{w}_k\end{aligned}\quad (3.7)$$

where the matrices  $\mathbf{B}_{(\cdot)}$  define the effects of the several artificial fields on the agent's state. Since humans tend to avoid quickly direction changes and to manifest, in a certain measure, a memory of their past actions, could be useful to incorporate previous states in our model. To take into account  $N$  past observations for the prediction at time  $k + 1$ , we modify the model as follows:

$$\mathbf{x}_{k+1} = \mathbf{A}^a [\mathbf{x}_k \ \mathbf{x}_{k-1} \ \dots \ \mathbf{x}_{k-N+1}]^T + \mathbf{B}^a \mathbf{F}_k^{tot}(\mathbf{p}_{k,\dots,k-N+1}, \mathbf{p}_g, \mathbf{p}_{k,\dots,k-N+1}^j) + \mathbf{w}_k, \quad (3.8)$$

where  $\mathbf{A}^a \in \mathbb{R}^{4 \times (4 \times N)}$  denotes the *augmented* state matrix and  $\mathbf{B}^a \in \mathbb{R}^{4 \times (2 \times N)}$ . The parameter  $N$  is set to  $1, 2, \dots, n_p$  to influence the current prediction by  $1, 2, \dots, n_p$  past observations, respectively. In order to compute the parameters of the model, i.e., the matrices  $\mathbf{A}^a$  and  $\mathbf{B}^a_{(\cdot)}$ , we use the least square approach. We arrange the available data as the columns of matrices, i.e.,  $\mathbf{X}_N = \mathbf{A}\mathbf{X} + \beta_g \mathbf{B}_g \mathbf{F}_g + \beta_r \mathbf{B}_r \mathbf{F}^{srep} + \beta_a \mathbf{B}_a \mathbf{F}^{sat} + \beta_d \mathbf{B}_d \mathbf{F}^d + \beta_o \mathbf{B}_o \mathbf{F}^o = \mathbf{H}\mathbf{Y}$  and then minimize the energy error  $J(\mathbf{H}) = \|\mathbf{X}_N - \mathbf{H}\mathbf{Y}\|_2^2$  using the Moore–Penrose pseudoinverse. We assume a Gaussian distributed noise with mean and covariance estimated as the sample mean and covariance of the residual matrix  $\mathbf{R}$ , i.e.,  $\mathbf{w} \sim \mathcal{N}(\mathbf{x}; \hat{\mu}_R, \hat{\Sigma}_R)$ .

### 3.5 Experiments

In the following, we describe the used protocol to test the model, along with the datasets, the metric, and the experimental results, both qualitative and quantitative.

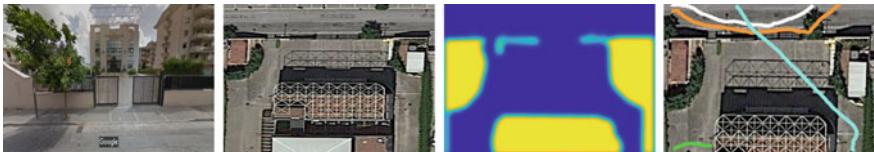
**Evaluation Protocol.** Firstly, each scene is manually annotated by segmenting obstacles and attractive elements using the software *Ratsnake* [14]. Due to noisy trajectories, we perform a fivefold cross-validation for each scenario considering as error the average over the five folds. The prediction is terminated when the agent leaves the scene or hits an obstacle. We generate 500 trajectories for each trajectory of the testing set. In particular, the model is trained to observe the first part of the trajectory, considering eight frames ( $N = 8$ ) to predict the next 12 frames, as in [11, 15]. We study the impact on the prediction by adding in turn each factor in Eq. (3.2). Firstly, we consider the free evolution of the system, i.e., the estimated state in absence of inputs, and then the forced response, which takes into account external forces. Specifically, we conduct experiments considering the following models:

$$\begin{aligned}
\text{I } \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{w}_k; \\
\text{II } \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \beta_r \mathbf{B}_r \mathbf{F}_k^{s_{rep}}(\mathbf{p}_k) + \mathbf{w}_k; \\
\text{III } \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \beta_r \mathbf{B}_r \mathbf{F}_k^{s_{rep}}(\mathbf{p}_k) + \beta_o \mathbf{B}_o \mathbf{F}_k^o(\mathbf{p}_k) + \mathbf{w}_k; \\
\text{IV } \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \beta_r \mathbf{B}_r \mathbf{F}_k^{s_{rep}}(\mathbf{p}_k) + \beta_o \mathbf{B}_o \mathbf{F}_k^o(\mathbf{p}_k) + \beta_d \mathbf{B}_d \mathbf{F}_k^d(\mathbf{p}_k, \mathbf{p}_k^j) + \mathbf{w}_k; \\
\text{V } \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \beta_r \mathbf{B}_r \mathbf{F}_k^{s_{rep}}(\mathbf{p}_k) + \beta_o \mathbf{B}_o \mathbf{F}_k^o(\mathbf{p}_k) + \beta_d \mathbf{B}_d \mathbf{F}_k^d(\mathbf{p}_k, \mathbf{p}_k^j) + \beta_g \mathbf{B}_g \mathbf{F}_k^g(\mathbf{p}_k, \mathbf{p}_g) + \mathbf{w}_k.
\end{aligned}$$

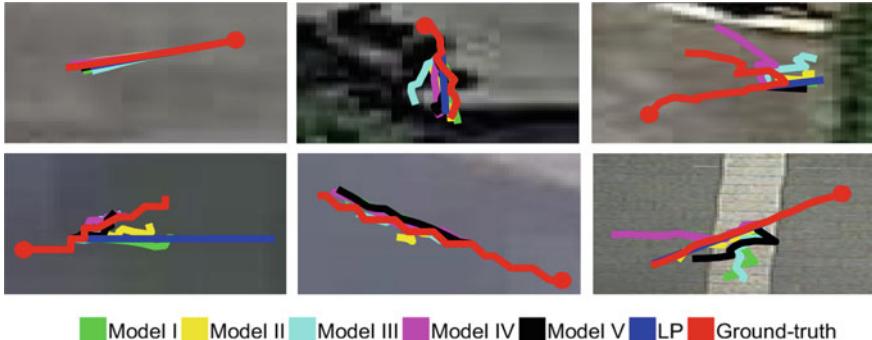
**Datasets.** We conduct experiments on a simulated dataset (SD), which represents the entrance of our department depicted in Fig. 3.4 along with subsets of two publicly available datasets: the Stanford Drone Dataset (SDD) [16] and the Edinburgh Informatics Forum Pedestrian Database (EIFPD) [17]. The former comprises videos of six areas collected by a drone upon a university campus and contains annotated tracked of various types of agents. The latter covers about six months of observations of people walking in the main free space of a university building. Both datasets are captured from a bird’s eye view, so there is no need of projective transformations. From the SDD dataset, we select the *DeathCircle0* scene considering as target classes both pedestrians and cyclists while from the EIFPD dataset, we select four scenes (*Aug 24*, *Jan 15*, *Jul 13*, *Jun 14*) averaging the results.

**Metrics.** To asses the model’s performances, we consider three different metrics. More specifically, given the observed and predicted trajectory, their physical distance is evaluated considering the average modified Hausdorff distance (MHD) [18]. Moreover, the negative log-likelihood (NLL) is evaluated building a transition probability matrix using the predicted trajectories to measure the observed trajectory’s probability to be sampled by such distribution. As stated, we consider square patches of  $15 \times 15$  pixels. Finally, we also report the average final position displacement (FPD) error obtained evaluating the Euclidean distance between the predicted final positions and the ground-truth ones.

**Baselines.** We consider as baseline a linear prediction (LP), i.e., a constant velocity model, whose constant velocity parameter  $\mathbf{v}_{CV}$  is sampled from a Gaussian distribution defined as  $\mathcal{N}(\mathbf{v}; \hat{\mu}_{CV}, \hat{\mathbf{R}}_{CV})$  where  $\hat{\mu}_{CV}$  and  $\hat{\mathbf{R}}_{CV}$  represent the sample mean and covariance of the velocity of the first  $N$  observations. Compared to Model I, the baseline does not take into account any coupling effects between  $x$  and  $y$  components.



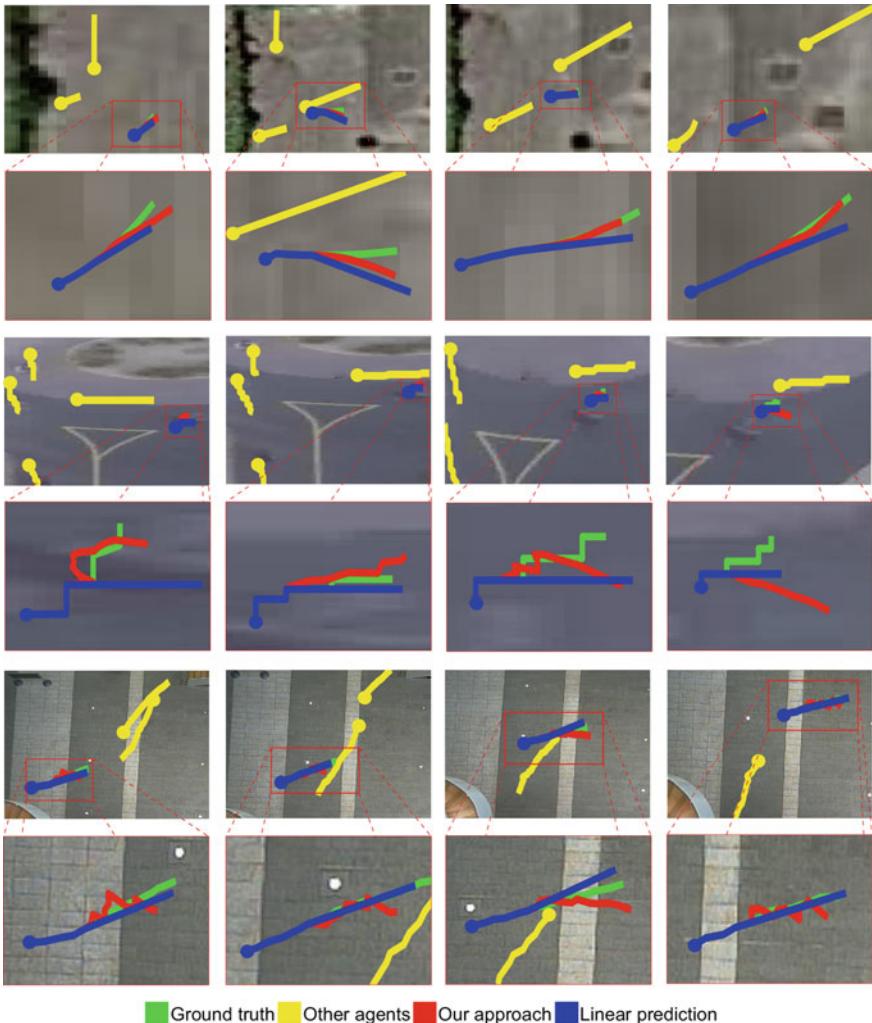
**Fig. 3.4** Figures show front view, top-down view, obstacles map, and some trajectories of the simulated scenario (from left to right). The trajectories are generated using a linear model including repulsions due to obstacles and other agents and the destination’s attraction. We set  $\mathbf{w} \sim \mathcal{N}(\mathbf{x}; \mathbf{0}, \sigma^2 \mathbf{I}_4)$  using three levels of noise:  $\sigma = 0$ ,  $\sigma = 0.5$ , and  $\sigma = 2$ , respectively



**Fig. 3.5** Qualitative results of predicted trajectories for the selected scenarios. The red circles represent the starting positions. The first row shows the simulated scenario using three noise levels, i.e.,  $\sigma = 0$ ,  $\sigma = 0.5$  and  $\sigma = 2$ , respectively. The second row shows the SDD dataset for a pedestrian and a cyclist trajectory (from left to right) while the last box shows an example drawn from the EIFPD dataset

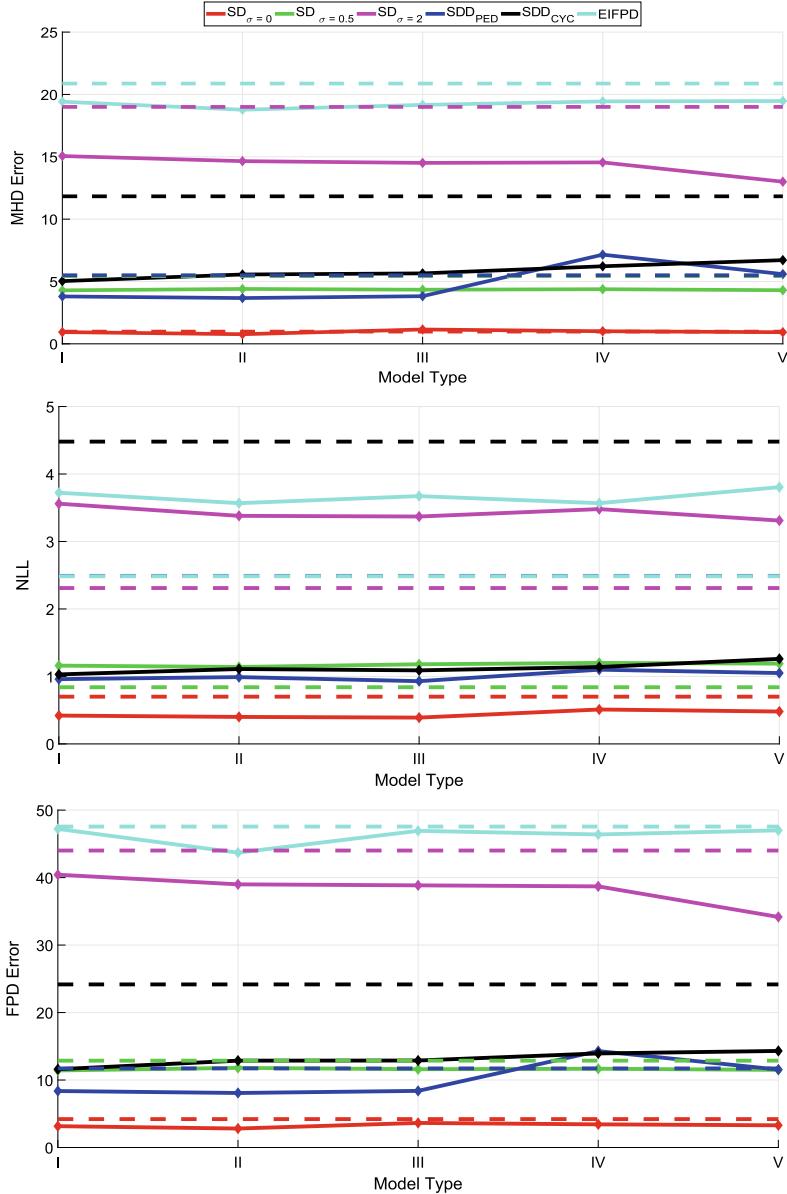
**Qualitative experiments.** We report in Fig. 3.5 the predicted trajectories for the selected scenarios. Specifically, the models appear able to predict short-term horizons with good accuracy when trajectories do not have complex behaviors due to turning points or remarkable noise. In this case, the LP baseline performs better than the several models. We also observe that, for the SDD dataset, the cyclist target appears better predicted than the pedestrian one due to more regular behaviors. Finally, as shown in the last box, the high level of noise does not allow the model to even predict quite linear behaviors. Figure 3.6 shows some examples of predicted trajectories including the interaction forces. For the simulated scenario (first row), the predicted trajectories are very similar to the ground truths even though, when the agents are too close, the repulsive force strongly deflects the target from his/her real path. Such effect is also evident for a cyclist target although the predicted dynamics appear better than the ones of the LP baseline. For a pedestrian target of the EIFPD dataset, our approach produces somewhat evident zigzag courses due to the high level of noise of annotated trajectories compared to the smoother trajectories of the LP baseline.

**Quantitative experiments.** Figure 3.7 shows the results for the selected metrics. MHD errors show that our models predict trajectories that are closer to the ground truths than the ones predicted by a constant velocity model, even considering their final positions. Furthermore, the NLL metric shows that, for noisy trajectories, our models have more irregular behaviors compared to the baseline. In some cases, the adding of artificial factors to the basic model has a limited impact on the prediction. The main reasons could be ascribed to the choice of several parameters of our framework, such as the influence distance among the agents ( $r_{max}$ ) along with the goal's potential field which seems not to appropriately influence the motion due to the short-walked distance. Nevertheless, the model outperforms the LP baseline,



**Fig. 3.6** Qualitative results of randomly selected predicted trajectories including the interaction forces. Results of model *IV* are considered. Each row shows successive frames (from left to right). The circles denote the starting positions. The first row represents the simulated scenario ( $\sigma = 0$ ); the third row shows a cyclist target of the SDD dataset; finally, the fifth row an example drawn from the EIFPD dataset

especially for the target cyclist in the SDD dataset. The non-regular behaviors of annotated pedestrians in the EIFPD dataset appear more challenging to predict since our results are comparable, in this case, to the selected baseline.



**Fig. 3.7** Quantitative results for the selected metrics: MHD, NLL, and FPD, respectively. We report the corresponding results for the selected scenes (solid lines) along with the results of the LP baseline (dot lines). The x-axis reports the type of model while the y-axis the corresponding metric value. The results show approximately constant behaviors for the five models even though they outperform the baseline

### 3.6 Conclusion

We have focused on human trajectory forecasting task in urban scenarios including several types of contexts and agents. The proposed comprehensive framework is based both on human–scene and human–human factors to foresee future positions. Dynamics are guided by artificial fields, extracted from a preliminary scene analysis, and dynamic factors, such as destination and other agents, which allows us to model significant elements that guide human behaviors. Our preliminary results show that model can be used to predict future positions in urban contexts despite its simplicity.

Our future work will be toward a more detailed analysis of interaction forces between two or more agents along with considering hidden states to model complex situations that may occur in urban context and that cannot be considered due to the inherent limitations of a linear model.

## References

1. Ge, S.S., Cui, Y.J.: Dynamic motion planning for mobile robots using potential field method. *Auton. Robot.* **13**(3), 207–222 (2002)
2. Bounini, F., Gingras, D., Pollart, H., Gruyer, D.: Modified artificial potential field method for online path planning applications. In: 2017 IEEE Intelligent Vehicles Symposium (IV), pp. 180–185 (June 2017)
3. Xie, D., Todorovic, S., Zhu, S.C.: Inferring dark matter and dark energy from videos. In: 2013 IEEE International Conference on Computer Vision, pp. 2224–2231 (Dec 2013)
4. Khatib, O.: Real-time obstacle avoidance for manipulators and mobile robots. In: Proceedings. 1985 IEEE International Conference on Robotics and Automation vol. 2, pp. 500–505 (Mar 1985)
5. Koren, Y., Borenstein, J.: Potential field methods and their inherent limitations for mobile robot navigation. In: Proceedings of 1991 IEEE International Conference on Robotics and Automation, vol. 2, pp. 1398–1404 (Apr 1991)
6. Ge, S.S., Cui, Y.J.: New potential functions for mobile robot path planning. *IEEE Trans. Robot. Autom.* **16**(5), 615–620 (2000)
7. Helbing, Dirk, Molnár, Péter: Social force model for pedestrian dynamics. *Phys. Rev. E* **51**, 4282–4286 (1995)
8. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 935–942 (June 2009)
9. Yamaguchi, K., Berg, A.C., Ortiz, L.E., Berg, T.L.: Who are you with and where are you going? *CVPR* **2011**, 1345–1352 (2011)
10. Sieben, A., Schumann, J., Seyfried A.: Collective phenomena in crowds - where pedestrian dynamics need social psychology. *PLOS ONE*, **12**(6):1–19, 06 2017
11. Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., Savarese, S.: Social lstm: human trajectory prediction in crowded spaces. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 961–971 (June 2016)
12. Bartoli, F., Lisanti, G., Ballan, L., Del Bimbo, A.: Context-aware trajectory prediction. ArXiv e-prints (May 2017)
13. Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S., Alahi, A.: Social GAN: socially acceptable trajectories with generative adversarial networks. ArXiv e-prints (March 2018)

14. Iakovidis, D.K., Smailis, C.V.: Efficient semantically-aware annotation of images. In: 2011 IEEE International Conference on Imaging Systems and Techniques, pp. 146–149 (May 2011)
15. Pellegrini, S., Ess, A., Schindler, K., van Gool, L.: You'll never walk alone: Modeling social behavior for multi-target tracking. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 261–268 (Sept 2009)
16. Robicquet, A., Sadeghian, A., Alahi, A., Savarese, S.: Learning social etiquette: human trajectory understanding in crowded scenes, pp. 549–565. Springer International Publishing, Cham (2016)
17. Majecka, B.: Statistical models of pedestrian behaviour in the Forum. Master's thesis, School of Informatics, University of Edinburgh (2009)
18. Dubuisson, M.P., Jain, A.K.: A modified hausdorff distance for object matching. In: Proceedings of 12th International Conference on Pattern Recognition, vol. 1, pp. 566–568 (Oct 1994)

# Chapter 4

## Convolutional Recurrent Neural Networks and Acoustic Data Augmentation for Snore Detection



Fabio Vesperini, Luca Romeo, Emanuele Principi,  
Andrea Monteriù and Stefano Squartini

**Abstract** In this paper, we propose an algorithm for snoring sounds detection based on convolutional recurrent neural networks (CRNN). The log Mel energy spectrum of the audio signal is extracted from overnight recordings and is used as input to the CRNN with the aim to detect the precise onset and offset time of the sound events. The dataset used in the experiments is highly imbalanced toward the non-snore class. A data augmentation technique is introduced, that consists in generating new snore examples by simulating the target acoustic scenario. The application of CRNN with the acoustic data augmentation constitutes the main contribution of the work in the snore detection scenario. The performance of the algorithm has been assessed on the A3-Snore corpus, a dataset which consists of more than seven hours of recordings of two snorers and consistent environmental noise. Experimental results, expressed in terms of Average Precision (AP), show that the combination of CRNN and data augmentation in the raw data domain is effective, obtaining an AP up to 94.92%, giving superior results within the related literature.

### 4.1 Introduction

People spend almost a third of their life sleeping, thus the sleep quality is very important to people's health. Most of us have experienced trouble sleeping just sometimes, while for some people sleep problems occur regularly; in former cases, a sleep disorder is diagnosed. Among these, one of the most common sleep disorders [7] is the chronic snoring. Snoring is a noise produced when vibration occurs at several levels of the upper airway. It may be associated with various degrees of upper airway resistance. It is a highly prevalent disorder among the community: it has been

---

F. Vesperini · L. Romeo · E. Principi (✉) · A. Monteriù · S. Squartini  
Department of Information Engineering,  
Università Politecnica delle Marche Via Brecce Bianche, 60131 Ancona, Italy  
e-mail: [e.principi@univpm.it](mailto:e.principi@univpm.it)

F. Vesperini  
e-mail: [f.vesperini@pm.univpm.it](mailto:f.vesperini@pm.univpm.it)

confirmed that approximately 30% of the overall population suffers from chronic snoring almost every night [24].

Sound snoring can have a negative impact on normal physical, mental, social, and emotional functioning of the person suffering from it and their bed partner [3]. It is also a typical symptom for obstructive sleep apnea (OSA), a chronic sleep disease sharing a frequent occurrence [19]. OSA is characterized by posterior pharyngeal occlusion for at least ten seconds with apnea/hypopnea and attendant arterial oxyhemoglobin desaturation. If left untreated, this life-threatening sleep disorder may induce high blood pressure, coronary heart disease, pulmonary heart failure, and even nocturnal death [2]. In addition, OSA is indicated between the main causes of significant morbidity among children [13].

#### **4.1.1 Related Works**

Recent studies have found that acoustic signal carries important information about the snore source and obstruction site in the upper airway of OSA patients [14]. This significant discovery has motivated several researchers to develop acoustic-based approaches that could provide inexpensive, convenient, and non-invasive monitoring and screening apparatuses to be combined with traditional diagnostic tools. The study described in [5] consists of the identification and segmentation process by using energy and zero crossing rate (ZCR), which were used to determine the boundaries of sound segments. Episodes have been efficiently represented into two-dimensional spectral features by using principal component analysis and classified as snores or non-snores with robust linear regression (RLR). The system was tested by using the manual annotations of an ear-nose-throat (ENT) specialist as a reference. The accuracy for simple snorers was found to be 97.3% when the system was trained using only simple snorers' data. It drops to 90.2% when the training data contain both simple snorers' and OSA patients' data. In the case of snore episode detection with OSA patients, the accuracy is 86.8%. In [15] an automatic detection, segmentation and classification algorithm of snore-related signals (SRS) from overnight audio recordings is proposed by combining acoustic signal processing with machine learning techniques. The specialists have found from OSA patients four classes of SRS, connected to the sound origin mechanism. It is typically defined as the velumopharyngeal-tongue-epiglottis (VOTE) scheme and identifies the tissues in upper airway that are involved in the noise generation. For classification a k-nearest neighbor (k-NN) model is adopted for its good performance on pattern recognition.

In the occasion of the Interspeech 2017 ComParE challenge [17] and subsequent investigations, different approaches based on deep neural networks (DNNs) have been presented [1, 10, 21] with the aim to classify isolated snore sound events among the four classes based of the VOTE scheme.

In this work, we propose the application of convolutional recurrent neural networks (CRNNs) to detect snoring episodes from overnight recordings acquired in real-life conditions. The method is a two-step process: the acoustic spectral features

extraction and the CRNN combined with gated recurrent units (GRU) processing. The algorithm is evaluated by using the average precision (AP) score.

Differently from other deep learning approaches [1, 10, 21], this choice offers a viable and natural solution for jointly learning the spatio-temporal dependencies of audio sequence for discovering snoring event. Additionally, we deal with the high unbalanced setting which exists in the task of snore detection. The original snore/background ratio in the aforementioned signals has been increased by adding isolated snore events from the Munich-Passau Snore Sound Corpus dataset [17]. The reliability of the proposed approach is investigated using the A3-snore dataset, leading to significant improvement in term of average precision (AP) with respect to convolutional neural network (up to 9.48%) and other data augmentation techniques.

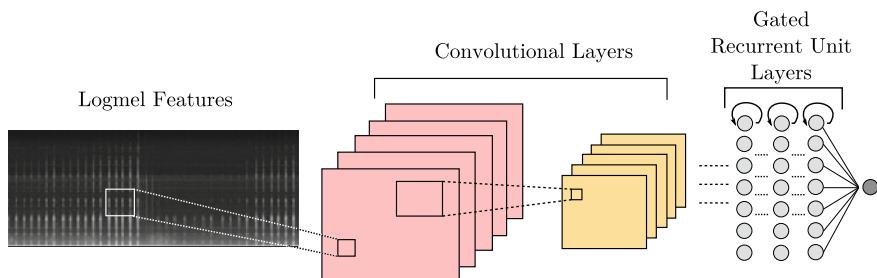
The outline of this paper is the following. Section 4.2 describes the acoustic features extraction procedure and the employed DNN architecture. Section 4.3 are reports the dataset details, the evaluated data augmentation techniques, and the experimental setup. Section 4.4 reports the results of the best performing models and Sect. 4.5 concludes this contribution and presents future development ideas.

## 4.2 Proposed Approach

The two steps of the proposed approach are detailed in this section, starting from the spectral features extraction and ending with the convolutional recurrent neural network (Fig. 4.1).

### 4.2.1 Features Extraction

The feature extraction stage operates on stereo audio signals sampled at 44.1 kHz. Following the results obtained in recent works related to sound event detection [23], we use the log Mel energy coefficients (Logmel) as an efficient representation of the



**Fig. 4.1** The proposed approach scheme

audio signal. The stereo signal is firstly down-mixed to mono by averaging the two channels. The resulting audio signal is split into 30ms frames and a frame step of 10ms, then the Logmel coefficients are obtained by filtering the power spectrogram of the frame by means of a mel filter bank, then applying a logarithmic transformation to each sub-band energy in order to match the human perception of loudness. We used a filter bank with 40 mel scaled channels, obtaining 40 coefficients/frame.

#### 4.2.1.1 Convolutional Recurrent Neural Networks

CRNNs used in this work are composed of four types of layers: convolutional layers, pooling layers, recurrent layers, and detection layer.

The mathematical operation of convolution has been introduced in artificial neural network layers since 1998 [12] for image processing. Recently, CNNs have been often used also in audio tasks, where they exploit one input dimension to keep track of the temporal evolution of a signal [22]. In this particular case, the CNNs perform better with filter size small with respect to the input matrix, and this means the temporal context observed in these layers is typically less than two hundred milliseconds. Each convolutional layer is followed by batch normalization per feature map [11], a leaky rectified linear unit activation function (LeakyReLU) and a dropout layer [18] with rate equal to 0.3. A frequency domain max-pooling layer is then applied to the resulting feature map, in order to enhance the relevant information from frequency bands without lose the temporal resolution of the Logmels, as proposed in [4]. The extracted features over the CNN feature maps are stacked along the frequency axis. Max-pooling operation combined with shared weight in convolutional layers provide robustness to frequency shifts in the input features and this is crucial to overcome the problem of intra-class acoustic variability for snore events.

In the recurrent block, the stacked features resulting from the last pooling layer are fed to layers composed of GRUs [9], where tanh and hard sigmoid activation functions are used for update and reset gates, respectively. GRU layers control the information flow through a gated unit structure, which depends on the layer input, on the activation at the previous frame and on the reset gate. Fast response to the changes in the input and the previous activation information is fundamental for high performance in the proposed algorithm, where the task is to detect a small chunk of consecutive time frames where the target event is present. In addition, a previous work [20] demonstrates improvements provided by recurrent architectures in the sound event detection in real-life audio.

The detection layer is a feed-forward layer of composed of a single neuron with sigmoid activation function, corresponding to the probability the event onset. The layer is time distributed, this means that while computing the output of the classification layer, the same weight and bias values are used over the recurrent layer outputs for each frame.

In a comparative aim, we implemented also a CNN architecture very similar to the CRNN, the only difference being that the recurrent layers of the CRNN are re-

placed with time distributed feed-forward layers with ReLU activations. In following section, we will refer it as CNN.

The neural networks training was accomplished by the AdaDelta stochastic gradient based optimization algorithm [25] for a maximum of 500 epochs on the binary cross entropy loss function. The optimizer hyperparameters were set according to [25] (i.e., initial learning rate  $lr = 1.0$ ,  $\rho = 0.95$ , and  $\epsilon = 10^{-6}$ ). An early stopping strategy, monitoring the validation AP score, was employed in order to reduce the computational burden and avoid overfitting.

## 4.3 Experiments

This section details the composition of the dataset used in this work, the data augmentation techniques, the performance metrics, and the experimental setup.

### 4.3.1 Dataset

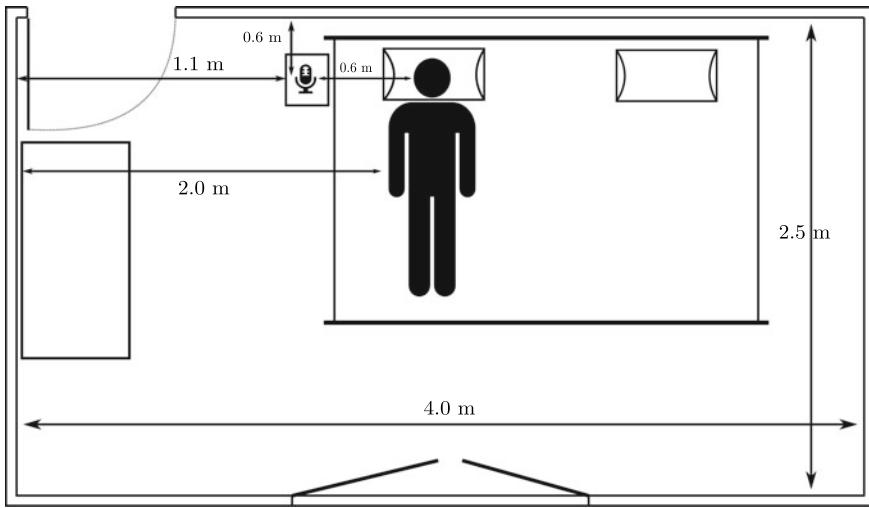
The snore detection algorithm has been evaluated on the A3-Snore dataset. A brief description of the acquisition setup and dataset splitting is provided in the following.

#### 4.3.1.1 Acquisition Setup

In order to capture the overnight audio recordings, a ZOOM-H1 Handy Recorder has been used. It is equipped with two unidirectional microphones set at a 90 degree angle relative to one another. The signals are stored in WAV files with a sampling rate of 44.1 kHz and bit depth equal to 16. The input gain is automatically set by the recorder to prevent overload and distortion, while the high-pass filter was enabled in order to eliminate pops, wind noise, blowing, and other kinds of low-frequency rumble.

#### 4.3.1.2 Acquisition Environment

The acquisition environment consists of a simple bedroom, with two access points (door and window). The recorder is placed near the patient, at same height of the bed and in line with the subject's mouth. During the recordings, the patient is the only one that can occupy the bedroom, in order to avoid contaminations on recorded audio signals. The room dimensions are reported in Fig. 4.2. Background sounds include traffic noise, breathing and speech signals, and house and animal noises. We acquired some samples measurements of the event-to-background (EBR) ratios considering



**Fig. 4.2** Plant of the recording room

background noise, snoring events, and noise events such as “car passing by” or “dog barking.” The EBR resulted equal to 6.5 dB and 1.1 dB, respectively, for noise to background EBR and snore to background EBR.

#### 4.3.1.3 Dataset Splitting

The original recordings have been manually labeled, annotating the snore events onset and offset with a resolution of 1 second. The audio sequences have been divided into chunks of 10 min, and only those with the highest number of snore events have been used in the experiments. The dataset is organized into subjects, which can be, respectively, used as *training* or *validation* sets in a twofold cross-validation strategy (i.e., Leave One Subject Out procedure). The number of events per class in the database is strongly unbalanced as reported in Table 4.1. Thus, the snore detection task is challenging, due to the high number of noises on the A3-SNORE dataset.

**Table 4.1** Difference of recording times for each class, divided by snorers

A3-SNORE dataset

#	Gender	Age	Snoring (SN)	Total duration (Tot)	Ratio (SN/Tot) (%)
Snorer 1	M	48	33m-27s	3h-12m-0s	14.5
Snorer 2	M	55	21m-21s	3h-50m-0s	11.1
Total			54m-48s	7h-02m-0s	12.8

### 4.3.2 Data Augmentation Techniques

In this application, what we are really interested is to detect the minority class (e.g., snoring events) rather than the majority class (e.g., background). Thus, we need to adequately train the models in order to obtain a fairly high prediction for the minority class. In order to counteract the dataset unbalancing existing in the task of snoring detection, different techniques of data augmentation have been evaluated. The literature suggests that it is possible to augment training data in data space or in feature space. In this work, both data augmentation approaches have been evaluated, by using the *Synthetic Minority Over-sampling Technique* (SMOTE) [6] in the feature space, and by generating simulated data with an increased number of snore events. The original snore/background ratio in the acquired signals has been increased with these transformations to approximately 30% [24], maintaining anyway a natural unbalance which is properly of this task. In the following subsections, a brief description of each method is provided.

#### 4.3.2.1 Majority Class Under-Sampling

it is not a properly data augmentation technique but it is a fast and easy way to balance the data. It consists in randomly selecting a subset of data from the training sets in order to modify the ratio of the sample occurrences in two classes.

#### 4.3.2.2 SMOTE

It is an over-sampling approach in which the minority class is over-sampled by creating new synthetic examples. The minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the  $k$  minority class nearest neighbors ( $k$ -NN). Depending upon the amount of required over-sampling, neighbors from the  $k$ -NNs are randomly chosen. In particular, synthetic samples are generated in the following way: the difference between the feature vector (sample) under consideration and its nearest neighbor is multiplied by a random number between 0 and 1, and this is added to the feature vector under consideration. In details, for a sample  $x_i$ :

$$x_j^{\text{SMOTE}} = x_i + (\tilde{x}_{i,k} - x_i) \cdot r(j) \quad (4.1)$$

where  $r(j) \in [0, 1]$ . This causes the selection of a random point along the line segment between two specific features. This approach effectively forces the decision region of the minority class to become more general.

### 4.3.2.3 Proposed Approach—Generating Simulated Data

The simulated training sets have been created starting from the folds described in Sect. 4.3.1. The impulse responses between the snore source and the microphones have been recreated by using the library Pyroomacoustics [16]. Isolated snore sounds have been taken from the Munich-Passau Snore Sound Corpus (MPSSC) dataset [17]. It is composed of 843 snore events which have been extracted and manually screened by medical experts from drug-induced sleep endoscopy (DISE) examinations of 224 subjects. The augmented training set has been created by convolving the isolated snore sound events of the MPSSC corpus with the synthetic impulse responses. Then, the obtained signals have been mixed with the original recordings without overlap with the already present events. The artificial added event dynamic was normalized to the maximum value observed in the original signals. The resulting total time of snore signals is 55 min for Snorer 1, and 56 min and 5 s for Snorer 2.

### 4.3.3 Performance Metrics

The performance of the algorithms has been evaluated in term of AP score, a metric that summarizes the Precision and Recall curve. AP score is calculated as follows:

$$\text{AP} = \sum_n (R_n - R_{n-1})P_n, \quad (4.2)$$

where  $R_n$  and  $P_n$  are the Recall and Precision for threshold  $n$ , respectively. Precision and Recall for a generic threshold are calculated as follows:

$$R_n = \frac{TP_n}{TP_n + FN_n}, \quad P_n = \frac{TP_n}{TP_n + FP_n}, \quad (4.3)$$

where  $TP_n$  is the number of snore frames correctly detected,  $FN_n$  is the number of snore frames erroneously detected as background, and  $FP_n$  is the number of background frames erroneously detected as snoring.

### 4.3.4 Experimental Setup

To assess the performance of the models, we explored different hyper-parameter configurations and, for each of these, we repeated the whole experiments training the models both with the original data and with data processed with techniques described in Sect. 4.3.2. Table 4.2 shows the hyper-parameter configurations analyzed in our experiments. They regard kernels size, kernel number, and GRUs for a total of 120 experiments. In the case of CNN, the number of units and layer refers to a multilayer perceptron (MLP) architecture. The experiments were conducted in a twofold cross-

**Table 4.2** Explored network layout parameters

Convolutional layers number	3
Kernel number	4, 8, 16, 32, 64
Kernel size	$5 \times 5$ , $3 \times 3$ , $2 \times 2$
Pooling size	$5 \times 1$ , $4 \times 1$ , $2 \times 1$
Recurrent layers number	2, 3
Dense layers number	2, 3
Number of units	4, 8, 16, 32, 64

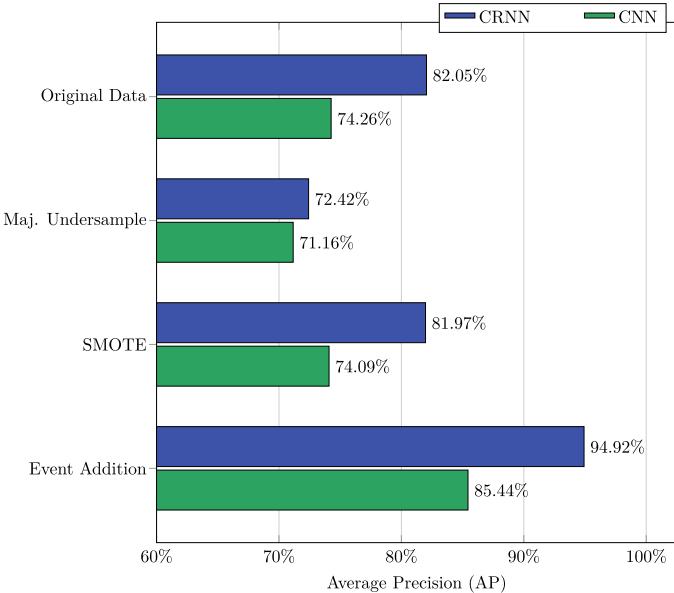
validation strategy corresponding on a leave one subject out procedure, thus in fold 1, we used Snorer 1 as training set and Snorer 2 as validation set and in fold 2 vice-versa. The models were selected on the performance based on the AP score averaged on the two folds. The algorithm has been implemented in the Python language using Keras [8] as deep learning library. All the experiments were performed on a computer equipped with a 6-core Intel i7, 32 GB of RAM and two Nvidia Titan X graphic cards.

## 4.4 Results

The performance of the CRNN and CNN architectures using different data augmentation techniques are reported in Fig. 4.3. In blue are depicted results with CRNN, in green the results of the CNNs. The CRNNs show to be effective for snore event detection yet with the original data, although the dataset imbalance. The best performing model is composed of three CNN layers with, respectively [64,64,64], filters of size  $3 \times 3$  and two GRU layers of 64 units. This configuration obtains an AP up to 82.05%, with a difference of +7.79% with respect to the CNN.

The majority class under-sampling and the SMOTE techniques obtain worst performance with respect to original recordings. For majority class under-sampling, this can be motived by the necessity of DNN models of a large amount of data to be trained properly; thus, a reduction of data samples cannot benefit to their detection ability. Regarding the SMOTE, the performance reduction is less dramatic ( $-0.16\%$  and  $-0.08\%$ , respectively, for CNNs and CRNNs) but its employment remains vain. In this case, the motivation can be found on the complexity to generate new samples of audio signals in the feature space which can really improve a DNN performance.

The addition of isolated snore samples convolved with the simulated room impulse response has a tangible beneficial effect on the examined models. In fact, with this technique we obtain an AP improvement equal to 11.18% and 12.87%, respectively, for the CNN and the CRNN. The latter obtains an AP equal to 94.92% with an architecture composed of 3 CNN layers with, respectively [64,32,32], filters of size  $5 \times 5$  and two GRU layers of 32 units. This model is composed of 91,553 free-parameters and occupies approximately 1.2 MB, providing to the algorithm a feasible complexity in an application scenario.



**Fig. 4.3** Results with different data augmentation techniques for the best models of the evaluated architectures

## 4.5 Conclusion

In this paper, a deep learning algorithm based on a CRNN architecture fed with Log-mel spectral features extracted from the audio signal has been proposed for snore detection. The A3-Snore dataset has been acquired in real-world conditions, containing overnight recordings of two male subjects and it has been used to assess the performance of the models. The original snore/background ratio has been increased by adding isolated snore events from the Munich-Passau Snore Sound Corpus dataset [17]. The reliability of the proposed approach has been investigated with respect to baseline CNN and different data augmentation techniques such as over-sampling (i.e., SMOTE) and downsampling. Results show that the presented snore detection methodology is able to better generalize across different users. In particular, the CRNN is able to extract salient information from the spectral features in order to discriminate snore events, while the implemented data augmentation provide additional samples of the minority class (i.e., snore events). These samples contain supplementary information that can be exploited by the CRNN for learning and discriminate snore events. Future works will be addressed to employ this methodology in a weakly supervised setting. Specifically, in the real-life applications, the precise annotation of existing events from overnight recordings can be onerous and can result in sparse labeling. Machine learning models trained in a weakly supervised fashion can help to counteract this problem without losing the state-of-the-art performance.

## References

1. Amiriparian, S., Gerczuk, M., Ottl, S., Cummins, N., Freitag, M., Pugachevskiy, S., Baird, A., Schuller, B.: Snore sound classification using image-based deep spectrum features. In: Proceeding of Interspeech. Stockholm, Sweden (Aug 20–24 2017)
2. Banno, K., Kryger, M.H.: Sleep apnea: clinical investigations in humans. *Sleep Med.* **8**, 400–426 (2007)
3. Blumen, M.B., Salva, M.A.Q., Vaugier, I., Leroux, K., d'Ortho, M.P., Barbot, F., Chabolle, F., Lofaso, F.: Is snoring intensity responsible for the sleep partner's poor quality of sleep? *Sleep Breath.* **16**(3), 903–907 (2012)
4. Cakir, E., Virtanen, T.: Convolutional recurrent neural networks for rare sound event detection. In: Proceeding of DCASE. pp. 27–31 (2017)
5. Cavusoglu, M., Kamasak, M., Erogul, O., Ciloglu, T., Serinagaoglu, Y., Akcam, T.: An efficient method for snore/nonsnore classification of sleep sounds. *Physiol. Meas.* **28**(8), 841 (2007)
6. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002)
7. Chokroverty, S.: Sleep Disorders Medicine E-Book: Basic Science, Technical Considerations, and Clinical Aspects. Elsevier Health Sciences (2009)
8. Chollet, F., et al.: Keras (2015)
9. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint [arXiv:1412.3555](https://arxiv.org/abs/1412.3555) (2014)
10. Freitag, M., Amiriparian, S., Cummins, N., Gerczuk, M., Schuller, B.: An ‘end-to-evolution’hybrid approach for snore sound classification. In: Proceeding of Interspeech. Stockholm, Sweden (Aug 20–24 2017)
11. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proceeding of ICMC, pp. 448–456 (2015)
12. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998)
13. Lumeng, J.C., Chervin, R.D.: Epidemiology of pediatric obstructive sleep apnea. *Proc. Am. Thorac. Soc.* **5**(2), 242–252 (2008)
14. Pevernagie, D., Aarts, R.M., De Meyer, M.: The acoustics of snoring. *Sleep Med. Rev.* **14**(2), 131–144 (2010)
15. Qian, K., Xu, Z., Xu, H., Wu, Y., Zhao, Z.: Automatic detection, segmentation and classification of snore related signals from overnight audio recording. *IET Signal Proc.* **9**(1), 21–29 (2015)
16. Scheibler, R., Bezzam, E., Dokmanić, I.: Pyroomacoustics: a Python package for audio room simulations and array processing algorithms. In: Proceeding of ICASSP. Calgary, Canada (Apr 15–20 2018)
17. Schuller, B., Steidl, S., Batliner, A., Bergelson, E., Krajewski, J., Janott, C., Amatuni, A., Casillas, M., Seidl, A., Soderstrom, M., et al.: The interspeech 2017 computational paralinguistics challenge: addressee, cold & snoring. In: Computational Paralinguistics Challenge (ComParE), Interspeech 2017. pp. 3442–3446 (2017)
18. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**(1), 1929–1958 (2014)
19. Strollo, P.J.J., Rogers, R.M.: Obstructive sleep apnea. *N. Engl. J. Med.* **334**(2), 99–104 (1996)
20. Valenti, M., Tonelli, D., Vesperini, F., Principi, E., Squartini, S.: A neural network approach for sound event detection in real life audio. In: 2017 25th European Signal Processing Conference (EUSIPCO), pp. 2754–2758. IEEE (2017)
21. Vesperini, F., Galli, A., Gabrielli, L., Principi, E., Squartini, S.: Snore sounds excitation localization by using scattering transform and deep neural networks. In: Proceeding of the Int. Joint Conf. on Neural Networks (IJCNN). Rio de Janeiro, Brazil (Jul 8–13 2018), to appear
22. Vesperini, F., Vecchiotti, P., Principi, E., Squartini, S., Piazza, F.: Localizing speakers in multiple rooms by using deep neural networks. *Comput. Speech Lang.* **49**, 83–106 (2018)

23. Virtanen, T., Mesaros, A., Heittola, T., Diment, A., Vincent, E., Benetos, E., Elizalde, B.M.: Proceedings of the Detection and Classification of Acoustic Scenes and Events Workshop (DCASE2017). Tampere University of Technology, Laboratory of Signal Processing (2017)
24. Young, T., Finn, L., Kim, H., et al.: Nasal obstruction as a risk factor for sleep-disordered breathing. *J. Allergy Clin. Immunol.* **99**(2), S757–S762 (1997)
25. Zeiler, M.D.: AdaDelta: an adaptive learning rate method. arXiv preprint [arXiv:1212.5701](https://arxiv.org/abs/1212.5701) (2012)

# Chapter 5

## Italian Text Categorization with Lemmatization and Support Vector Machines



Francesco Camastra and Gennaro Razi

**Abstract** The paper describes an Italian language text categorizer by *Lemmatization* and support vector machines. The categorizer is composed of six modules. The first module performs the tokenization, removing the punctuation signs; the second and third ones carry out stopping and lemmatization, respectively; the fourth module implements the bag-of-words approach; the fifth one performs feature dimensionality reduction eliminating poor discriminant features; finally, the last module does the classification. The Italian text categorizer has been validated on a database composed of more than 1100 articles, extracted from online edition of three Italian language newspapers, belonging to eight different categories. The work is highly novel, since to the best our knowledge, there are no works in literature on Italian text categorization.

### 5.1 Introduction

The task of text categorization is to assign a document to one or more classes or categories from a predefined set. This task has several applications, including the automated indexing of scientific articles according to a predefined thesaurus of technical terms, selective dissemination of information to consumers, spam filtering [1, 2], news categorization, classification of email contents, identification of document topic [3], and web page classification [4].

Although the categorization of English texts was widely investigated in past, the categorization of other language texts has been poorly addressed until now. In particular, to the best of our knowledge, there are no works about the categorization of Italian texts in literature.

---

F. Camastra (✉) · G. Razi

Department of Science and Technology, University of Naples Parthenope,  
Centro Direzionale Isola C4, 80143 Naples, Italy

e-mail: [camastra@ieee.org](mailto:camastra@ieee.org)

G. Razi

e-mail: [infomail.razi@gmail.com](mailto:infomail.razi@gmail.com)

Having said that, this work aims to develop a categorizer of Italian language text, paying particular attention, as applicative scenario, the categorization of articles of Italian newspapers. The proposed categorizer has six modules performing tokenization, stopping, lemmatization, bag-of-words, dimensionality reduction, and classification, respectively. The main peculiarity of the proposed categorizer, with respect to categorizers of English texts, resides in replacing stemming process by suffix-stripping algorithms (e.g. Porter algorithm [5]) with lemmatization.

The paper is organized as follows. Section 5.2 describes the Italian language text categorizer; Sect. 5.3 presents some experimental results; finally, in Sect. 5.4, some conclusions are drawn.

## 5.2 Italian Text Categorizer

The Italian Text Categorizer is composed of six units, i.e. the *Tokenization module*, the *Stopping Module*, the *Lemmatization Module*, the *Bag-of-Words Representer*, the *Feature Ranker* and the *Classifier*.

### 5.2.1 Tokenization Module

This module has the task of performing *tokenization*, i.e. removing from the original text the punctuation signs (e.g. commas, full marks, question marks, semicolons). Moreover, the uppercase letters are transformed into lowercase. An example of tokenization is shown in Fig. 5.1.

Lo schema centrodestra M5S salta. Ma solo al termine di una giornata segnata da colpi di scena, aperture, finte trattative e rapido ritorno al punto di partenza. Di Maio non va oltre il sostegno esterno a Berlusconi. Il leader di Fli rifiuta. Matteo Salvini furibondo per il tira e molla e per la rigidità del Cavaliere. Nella Lega già in nottata parte il pressing per mollare gli alleati. Per ora Salvini resiste. La presidente del Senato incaricata, Elisabetta Alberti Casellati, si concede un secondo giro di consultazioni.



lo schema centrodestra m5s salta ma solo al termine di una giornata segnata da colpi di scena aperture finte trattative e rapido ritorno al punto di partenza dimaio non va oltre il sostegno esterno a berlusconi il leader di fli rifiuta matteo salvini furibondo per il tira e molla e per la rigidita del cavaliere nella lega gia in nottata parte il pressing per mollare gli alleati per ora salvini resiste la presidente del senato incaricata elisabetta alberti casellati si concede un secondo giro di consultazioni

**Fig. 5.1** Tokenization. On the left, there is the original text; on the right, the text after tokenization. The original text was extracted by Repubblica at page 2, on April 20 2018

### 5.2.2 Stopping Module

This stage has the task of individuating and then removing the so-called *stop words*. Stop words are the most frequent words in a text. Since they are the most frequent words in any text, they have poor discriminative power and they do not allow discriminating a text from another one. For this reason, stop words are removed in any text. Stop words can vary from language to language. They are usually articles, prepositions, conjunctions, auxiliary and modal verbs. In Italian language, the number of stop words is more than 400. A list of Italian stop words can be found in [6]. An example of stopping on an Italian text is shown in Fig. 5.2.

### 5.2.3 Lemmatization Module

This stage in English language text categorizers usually carries out stemming process, namely reduces each word to its linguistic root, i.e. the *stem*. Whereas for the English language reliable stemming algorithms (stemmers), e.g. Porter's algorithm [5], exist, in Italian language, it is not possible to construct good stemmers due to the high presence of irregular forms. For this reason, stemming process has been replaced with *lemmatization*, which consists of reducing the word to its *lemma*. To this purpose, we recall that in linguistics the lemma of a word is the word that is conventionally chosen to represent all flexed forms of a given term in a dictionary. For instance, the lemma of the English verb "has" is "to have", whereas, the lemma of the Italian verb "vado" is "andare". Lemmatization, in this work, has been performed using *morphit* [7] (Fig. 5.3).

lo schema centrodestra m5s salta  
ma solo al termine di una giornata  
segnata da colpi di scena aperture  
finte trattative e rapido ritorno al  
punto di partenza dimaio non va  
oltre il sostegno esterno a  
berlusconi il leader di fi rifiuta  
matteo salvini furibondo per il tira  
e molla e per la rigidita del  
cavaliere nella lega gia in nottata  
parte il pressing per mollare gli  
alleati per ora salvini resiste la  
presidente del senato incaricata  
elisabetta alberti casellati si  
concede un secondo giro di  
consultazioni

schema centrodestra m5s salta  
solo termine di giornata segnata  
colpi scena aperture finte  
trattative rapido ritorno punto  
partenza dimaio non va oltre  
sostegno esterno berlusconi  
leader fi rifiuta matteo salvini  
furibondo tira molla rigidita  
cavaliere lega nottata parte  
pressing mollare alleati ora salvini  
resiste presidente senato  
incaricata elisabetta alberti  
casellati concede secondo giro  
consultazioni

**Fig. 5.2** Stopping. On the left there is the text after tokenization, on the right the text after the removing of stop words

schema centrodestra m5s salta solo termine di giornata segnata colpi scena aperture finte trattative rapido ritorno punto partenza dimaio non va oltre sostegno esterno berlusconi leader fi rifiuta matteo salvini furibondo tira molla rigidita cavaliere lega nottata parte pressing mollare alleati ora salvini resiste presidente senato incaricata elisabetta alberti casellati concede secondo giro consultazioni		schema centrodestra m5s saltare solo termine di giornata segnare colpo scena apertura finto trattativa rapido ritorno punto partenza dimaio non andare oltre sostegno esterno berlusconi leader fi rifiutare matteo salvini furibondo tirare mollare rigidita cavaliere lega nottata partire pressing mollare alleato ora salvini resistere presidente senato incaricato elisabetta alberti casellati concedere secondo giro consultazione
---	--	---

**Fig. 5.3** Lemmatization On the left there is the text after stopping, on the right the text after lemmatization

#### 5.2.4 Bag-of-Words Representation of Text

In this stage, the text is converted in a format, suitable to be processed by a machine learning classification algorithm.

In order to do that, *Bag-of-Words (BOW)* approach [8] is applied to the text. If we denote by  $\mathcal{T} = (t_1, t_2, \dots, t_N)$  the set of words that appear in the text, the text is represented by a  $N$ -dimensional feature vector  $\mathbf{T} = (x_1, x_2, \dots, x_N)$ , with  $x_i = h(v(t_i))$ ,  $i = 1, \dots, N$ , where  $v(t_i)$  is the occurrence of the term  $t_i$  in the text and  $h(\cdot)$  is an appropriate function.

Among several proposed methods for computing  $h(\cdot)$  [9], we adopted *tf-idf* (i.e. *term frequency-inverse document frequency*) approach [9]. In tf-idf approach, the feature  $x_i$  associated to the term  $t_i$  in the text  $\mathbf{T} \in \mathcal{T}$  is given by:

$$x_i = v_{t_i T} \log \left( \frac{|\mathcal{T}|}{v_{t_i}} \right), \quad (5.1)$$

where  $v_{t_i T}$  is the number of occurrences of the term  $t_i$  in the text  $\mathbf{T}$  and  $v_{t_i}$  the number of occurrences of the term  $t_i$  in all texts of  $\mathcal{T}$ .

#### 5.2.5 Feature Ranker

After Bag-of-Words representation, a text is usually represented by several thousands of features. Hence, it is necessary to apply any feature selection method to reduce the features as low as possible. For this reason, in this stage, we compute a score for each term and all terms are sorted on the score basis. Finally, the top  $K$  terms are picked, where  $K$  is a value *a priori* fixed.

As feature score, we have adopted the *information gain* (or *mutual information*) [10]. The information gain  $\tau(x_i)$  assigned to each feature  $x_i$  is given by:

$$\tau(x_i) = \sum_{c \in \mathcal{C}} \sum_{t \in \{x_i, x'_i\}} P(t, c) \log \left[ \frac{P(t, c)}{P(t)P(c)} \right] \quad (5.2)$$

where  $x'_i$  denotes the absence of  $x_i$  and  $\mathcal{C}$  the set of the all possible categories a text can belong to.

### 5.2.6 Classification

The classification is performed by support vector machines that we remind. Support vector machine (*SVM*) [11, 12] is a binary classifier and the patterns with output +1 and -1 are called positive and negative, respectively. The underlying idea of SVM is the computation of the *optimal hyperplane algorithm*, i.e. the plane that yields the maximum separation margin between two classes. Let  $\mathcal{T} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_\ell, y_\ell)\}$  be the training set, the optimal hyperplane  $y(\mathbf{x}) = \mathbf{m} \cdot \mathbf{x}$  can be obtained solving the following constrained optimization problem

$$\min_{\mathbf{m}} \frac{1}{2} \|\mathbf{x}\|^2 \quad \text{subject to } y_i((\mathbf{m} \cdot \mathbf{x}_i + b)) \geq 1 \quad i = 1, \dots, \ell. \quad (5.3)$$

However, in real-world applications, mislabelled examples might exist yielding a partial overlapping of the two classes. To cope with this problem, it allows that some patterns can violate the constraints, introducing slack variables. They are strictly positive when the respective patterns violate the constraint, otherwise they are null. SVM allows controlling at the same time the margin, expressed by  $\|\mathbf{m}\|$ , and the number of the training errors, given by the non-null  $\xi_i$ , by the minimization of the constrained optimization problem:

$$\min_{\mathbf{m}} \frac{1}{2} \|\mathbf{x}\|^2 + C \sum_{i=1}^{\ell} \xi_i \quad \text{subject to } y_i((\mathbf{m} \cdot \mathbf{x}_i + b)) \geq 1 \quad i = 1, \dots, \ell. \quad (5.4)$$

The constant  $C$ , sometimes called *regularization constant*, manages the trade-off between the separation margin and the number of the misclassifications. Since both the function and constraints are convex, the problem (5.4) can be solved by means of the method of Lagrange multipliers  $\alpha_i$ , obtaining the following final form:

$$\max_{\alpha} \sum_{i=1}^{\ell} \alpha_i - \frac{1}{2} \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \quad \text{subject to} \quad (5.5)$$

$$0 \leq \alpha_i \leq C \quad i = 1, \dots, \ell \quad (5.6)$$

$$\sum_{i=1}^{\ell} \alpha_i y_i = 0. \quad (5.7)$$

The vectors, whose respective multipliers are non-null, are called *support vectors*, justifying the name of the classifier. Optimal hyperplane algorithm implements the following decision function:

$$f(\mathbf{x}) = \operatorname{sgn} \left( \sum_{i=1}^{\ell} \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{x}) + b \right). \quad (5.8)$$

In order to get SVM, it is adequate to map nonlinearly input data in a feature space  $\mathcal{F}$ , before computing the optimal hyperplane. This can be performed by using *kernel trick* [13], i.e. replacing the inner product with an appropriate Mercer kernel  $G(\cdot)$ , obtaining the final form of the SVM decision function:

$$f(\mathbf{x}) = \operatorname{sgn} \left( \sum_{i=1}^{\ell} \alpha_i y_i G(\mathbf{x}_i, \mathbf{x}) + b \right). \quad (5.9)$$

In the recognizer it is used, as Mercer Kernel, the *Gaussian Kernel* [13], defined as  $G(\mathbf{x}, \mathbf{y}) = \exp(-\gamma \|\mathbf{x} - \mathbf{y}\|^2)$ , where  $\gamma \in \mathbf{R}^+$ .

Since SVM is a binary classifier, several strategies [13, 14] have been proposed to allow SVM usage when the classification task has more than two classes. Among these strategies, in the classifier, it is used *One-Versus-All (OVA)* method since it is the one that requires the minimum number of SVMs to train.

OVA method trains a classifier for each of  $P$  classes versus all the other classes. The method consists in training  $P$  SVM  $f_j$  by labelling the training data that have  $y_i = k$  with 1 and the rest of training data with  $-1$  during the training of the  $k^{th}$  classifier. In testing, the decision function is given by:

$$F(\mathbf{x}) = \arg \max_k f_k(\mathbf{x}) \quad (5.10)$$

SVM experiments, in this work, have been performed using *SVMLight* [15] software package.

### 5.3 Experimental Results

For the experimental validation of Italian text categorizer, we have constructed a database, using RSS (Rich Site Summary) [16] technology, extracting automatically articles from online editions of three Italian language newspapers, namely *La Repubblica*, *Il Mattino* and *Il Giornale*. The category assigned to each article was the same ascribed by the online newspaper containing it. The articles of the database belong

to eight different categories, namely *Sport*, *Motors*, *Entertainment and Culture*, *Science*, *News Section*, *Business*, *Politics* and *Foreign*. The database, formed by 2224 articles, was divided randomly in training and test set, composed of 1124 and 1110, respectively. In the validation, several SVMs were trained by using diverse values of the parameter  $\gamma$  of gaussian kernel and regularization constant  $C$ . Moreover, the number of features (i.e. words) was also varied, due to information gain criterion. The number of classes used in the experiments was eight, namely the number of different hand poses in our database. SVM trainings were carried out by using *SVM-pak* [15] software library. The best SVMs were chosen by *k-fold cross-validation* [17].

Table 5.1 reports the classifier performances on the test set, measured in terms of accuracy, whereas Table 5.2 presents the confusion matrix for Italian text classifier. The best result in terms of accuracy was 81.62% using the top 50% features, ranked in terms of information gain. To the best of our knowledge, no Italian text categorizer has been developed in past. Hence, no comparison with analogous systems is possible.

**Table 5.1** Accuracy of SVMs classifier on the test set

Category	# Articles	# Correctly classified	Accuracy (%)
Sport (SP)	240	223	92.92
Motors (M)	124	105	84.68
Entertainment and culture (EC)	56	47	83.93
Science (S)	127	101	79.53
News section (NC)	159	123	77.36
Business (B)	132	101	76.51
Politics (P)	157	119	75.80
Foreign (F)	115	87	75.65
All test set	1110	906	81.62

**Table 5.2** Confusion matrix of Italian text classifier, without rejection, on the test set. The values are expressed in terms of percentage rates

	SP	M	E	SC	NC	B	P	F
SP	92.92	0.42	0	0.42	0.42	0	0	5.82
M	3.23	84.68	0	4.03	3.23	0	1.60	3.23
E	1.79	0	83.93	7.13	1.79	1.79	3.57	0
SC	8.66	3.94	0	79.53	0	3.94	0.79	3.14
NC	6.28	1.26	0.63	2.52	77.36	1.26	2.52	8.17
B	6.82	3.03	0	4.55	3.03	76.51	1.51	4.55
P	3.83	1.27	1.27	0	7.01	5.09	75.80	5.73
F	1.74	0	0	2.61	12.17	1.74	6.09	75.65

## 5.4 Conclusions

In this paper, an Italian text categorizer by lemmatization and support vector machines has been described. The system is composed of six modules, i.e. the *Tokenization module*, the *Stopping Module*, the *Lemmatization Module*, the *Bag-of-Words Representer*, the *Feature Ranker* and the *Classifier*. The Italian text categorizer has been validated on a database composed of 1110 articles of eight diverse categories, extracted from online edition of three Italian language newspapers, showing an average accuracy larger than 81%. To the best of our knowledge, the system described in the work is the first categorizer for Italian language texts.

In the next future, we plan to improve the categorizer investigating the following research lines, namely improving bag-of-words using N-gram representation, and enriching the lemmatization taking into account of word synonyms.

**Acknowledgements** Gennaro Razi developed part of the work, during his M. Sc. thesis in Computer Science at University of Naples Parthenope, with the supervision of Francesco Camastra. The research was funded by *Sostegno alla ricerca individuale per il triennio 2015–17* project of University of Naples Parthenope.

## References

1. Camastra, F., Ciaramella, F., Staiano, A.: Machine learning and soft computing for ict security: an overview of current trends. *J. Ambient. Intell. Intelligence Comput.* **4**(2), 235–247 (2013)
2. Guzella, T., Caminhas, W.: A review of machine learning approaches to spam filtering. *Expert. Syst. Appl.* **36**(7), 10206–10222 (2009)
3. Marturana, F., Tacconi, S.: A machine learning-based triage methodology for automated categorization of digital media. *Digit. Investig.* **10**, 193–204 (2013)
4. Prieto, V.M., Ivarez, M., Cacheda, F.: Saad, a content based web spam analyzer and detector. *J. Syst. Softw.* **86**, 2906–2918 (2013)
5. Porter, M.: An algorithm for suffix stripping. *Program* **14**(3), 130–137 (1980)
6. Razi, G.: Categorizzazione di articoli di quotidiani italiani mediante apprendimento automatico. M. Sc. Applied Computer Science Thesis (in Italian), University of Naples “Parthenope” (2017)
7. Zanchetta, E., Baron, M.: Morph-it! a free corpus-based morphological resource for the italian languagel. In: *Proceedings of Corpus Linguistic 2005* (2005)
8. Salton, G., Wong, A., Yang, C.: A vector-space model for automatic indexing. *Commun. ACM* **18**(11), 613–620 (1975)
9. Sebastiani, F.: Machine learning in automated text categorization. *ACM Comput. Surv.* **34**(1), 1–47 (2002)
10. Cover, T.M., Thomas, J.: Elements of Information Theory. Wiley (1991)
11. Cortes, C., Vapnik, V.: Support vector networks. *Mach. Learn.* **20**, 1–25 (1995)
12. Vapnik, V.: Statistical Learning Theory. Wiley, New York (1998)
13. Schölkopf, B., Smola, A.: Learning with Kernels. MIT Press, Cambridge (USA) (2002)
14. Shawe-Taylor, J., Cristianini, N.: Kernels Methods for Pattern Analysis. Cambridge University Press (2004)
15. Joachim, T.: Making large-scale svm learning practical. In: *Advances in Kernel Methods-Support Vector Learning*, MIT Press, pp. 169–184 (1999)
16. Wadham, R.: Rich site summary (rss). *Libr. Mosaics* **16**(1), 25–25 (2005)
17. Hastie, T., Tibshirani, R., Friedman, R.: *The Elements of Statistical Learning*, 2nd edn. Springer (2009)

## Chapter 6

# SOM-Based Analysis of Volcanic Rocks: An Application to Somma–Vesuvius and Campi Flegrei Volcanoes (Italy)



**Antonietta M. Esposito, Andrea De Bernardo, Salvatore Ferrara,  
Flora Giudicepietro and Lucia Pappalardo**

**Abstract** Algorithms based on artificial intelligence (AI) have had a strong development in recent years in different research fields of earth science such as seismology and volcanology. In particular, they have been applied to the study of the volcanic eruptive products of the recent activity of Mount Etna volcano. This work presents an application of the self-organizing map (SOM) neural networks to perform a clustering analysis on petrographic patterns of rocks of Somma–Vesuvius and Campi Flegrei volcanoes, in the Neapolitan area. The goal is to highlight possible affinity between the magmatic reservoirs of these two volcanic complexes. The SOM is known for its ability to cluster data by using intrinsic similarity measures without any previous information about their distribution. Moreover, it allows an easy understandable data visualization by using a two-dimensional map. The SOM has been tested on a geochemical dataset of 271 samples, consisting of 134 samples of Campi Flegrei eruptions (named CF), 24 samples of Somma–Vesuvius effusive eruptions (VF), 73 samples of Somma–Vesuvius explosive eruptions (VX), and finally 40 samples of “foreign” eruptions (ET), included to verify the neural net classification capability. After a pre-processing phase, applied to have a more appropriate data representation as input for the SOM, each sample has been encoded through a vector of 23 features, containing information about major bulk components, trace elements, and Sr isotopic ratio. The resulting SOM identifies three main clusters, and in particular, the foreign patterns (ET) are well separated from the other ones being mainly grouped in a single node. In conclusion, the obtained results suggest the ability of SOM neural network to

---

A. M. Esposito (✉) · A. De Bernardo · S. Ferrara · F. Giudicepietro · L. Pappalardo  
Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Napoli Osservatorio Vesuviano,  
Naples, Italy  
e-mail: [antonietta.esposito@ingv.it](mailto:antonietta.esposito@ingv.it)

F. Giudicepietro  
e-mail: [flora.giudicepietro@ingv.it](mailto:flora.giudicepietro@ingv.it)

L. Pappalardo  
e-mail: [lucia.pappalardo@ingv.it](mailto:lucia.pappalardo@ingv.it)

A. De Bernardo · S. Ferrara  
Università degli Studi di Napoli Federico II, Naples, Italy

associate volcanic rock suites on the basis of their geochemical imprint and can be consistent with the hypothesis that there might be a common magma source beneath the whole Neapolitan area.

## 6.1 Introduction

From a petrographic point of view, the volcanic rocks are classified according to their chemical composition. Our dataset includes geochemical analysis of alkaline volcanic rocks of explosive and effusive eruptions representative of the entire eruptive history of Somma–Vesuvius and Campi Flegrei active volcanoes. The chemical composition is expressed in terms of the *major elements* (as weight (wt.) % oxides, each >0.1%), which classify the suite of rocks and their degree of evolution (i.e., basalt, latite, trachyte, or rhyolite); some *trace elements* (expressed in ppm; present in concentrations <0.1%), which characterize the magma source (Earth's mantle) or the geodynamic environment in which magma is formed; and the *Sr isotopic ratios* ( $^{87}\text{Sr}/^{86}\text{Sr}$ ), indicative of magma chamber processes and timescale. Thus, the values of these petrological parameters provide information both on the mantle source, from which the magmas derive, and on the evolutionary processes and the residence times in the crustal magma chambers.

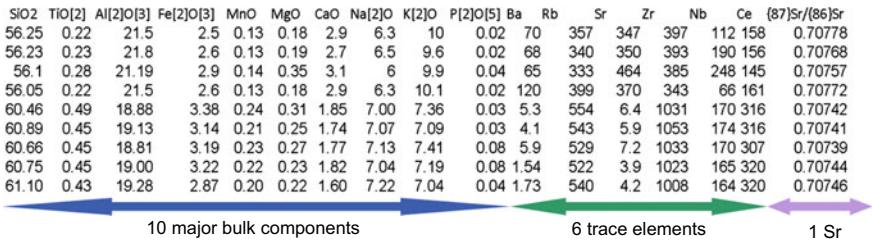
A widespread way to compare/distinguish suites of rocks is the use of the Harker variation diagrams in which the concentrations of an element or oxide are plotted against those of another element (generally  $\text{SiO}_2$ ). However, bivariate diagrams can have important limitations as similar trends may result from more than one geochemical process. It can be therefore useful to cluster several key parameters (e.g., major and trace elements, rare-earth elements, and isotopes) to better discriminate between different magma sources and/or processes.

The purpose of our analysis is to highlight a possible compositional affinity between magmas of Somma–Vesuvius and Campi Flegrei that could be indicative of a common deep magmatic source. In addition, to test the reliability of the proposed neural method, the geochemical analysis of volcanic rocks, with similar composition but originating from a different volcanic context (i.e., that of the Stromboli volcano, in the Aeolian Islands), has been added to the examined dataset.

In the following, the dataset and its parameterization are introduced, the SOM neural network, selected for this application, is presented, and finally, its results and our conclusions are illustrated.

## 6.2 The Selected Dataset and Parametrization

The dataset consists of a series of geochemical analyses of the products of the Neapolitan volcanic complex (Somma–Vesuvius and Campi Flegrei), sampled from deposits of representative well-studied eruptions (geochemical data come from [11,



**Fig. 6.1** An example of the chemical analysis vector of 17 elements: ten major elements (blue arrow), six trace elements (green arrow), and one Sr isotopic ratio (pink arrow)

[12] and reference therein). In total, there are 271 volcanic patterns, divided into the following four classes:

- 134 samples of Campi Flegrei eruptions (named **CF**);
- 24 samples of Somma–Vesuvius effusive eruptions (**VF**);
- 73 samples of Somma–Vesuvius explosive eruptions (**VX**);
- 40 samples of “foreign” eruptions (**ET**), belonging to the Stromboli volcano and included to test the SOM classification ability.

Initially, each chemical analysis sample is a vector of 17 features: ten *major elements*, six *trace elements*, and one *Sr* isotopic ratio. Figure 6.1 shows an example of this vector of 17 features.

In order to have an appropriate data representation as input variables for neural network, a pre-processing phase has been carried out on these vectors. First of all, to have a uniform distribution of values, we have normalized the trace element concentrations by dividing them for a common factor. Moreover, being the isotopic ratio of the strontium, *Sr*; a very discriminating variable, it was inserted seven times in the vector to emphasize its contribution in the network training.

At the end of this phase, each pattern input is encoded by using a vector of 23 features.

### 6.3 The Self-organizing Map Method

Techniques based on artificial intelligence (AI) are increasingly used to approach problems concerning the monitoring of natural hazards and geosciences [4–8, 13].

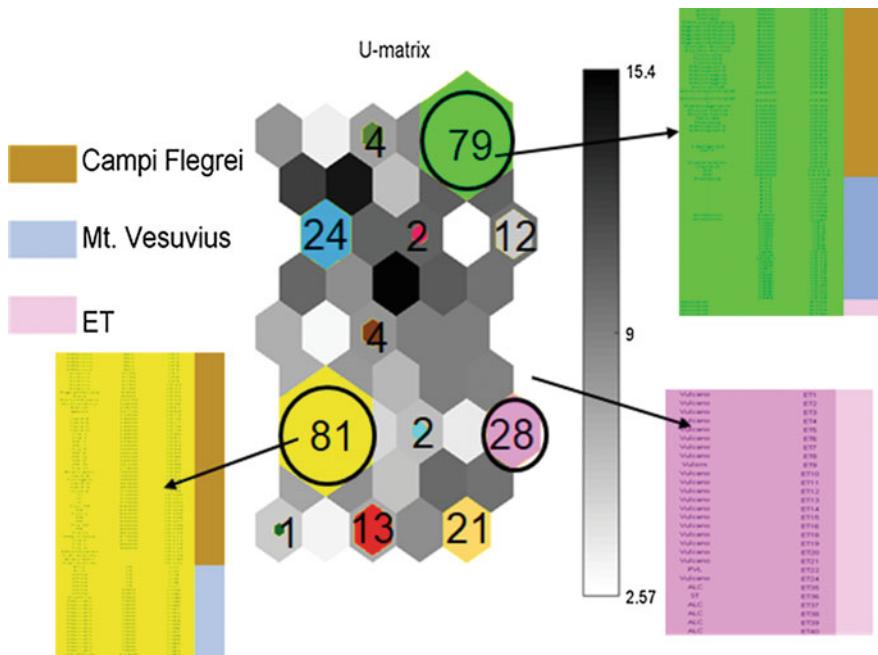
Recently, artificial neural networks have been used to study petrographic patterns of volcanic rocks [1, 2], in particular products erupted in the recent activity of Mt. Etna [3].

In this work, we adopted the self-organizing map (SOM) [10] to cluster our data. The SOMs are unsupervised neural networks mainly suitable for large and high-dimensional dataset analyses. They are able to identify clusters using intrinsic similarity measures in the data. Moreover, their bidimensional data representation is

particularly understandable for users. Finally, being unsupervised, no a priori information about the data is necessary to obtain the final clusters.

In our experiments, the SOM parameter values were chosen according to [9] and the SOM Toolbox for MATLAB (<http://www.cis.hut.fi/somtoolbox/>). We used a hexagonal map with  $5 \times 3 = 15$  nodes, obtaining the clustering illustrated in Fig. 6.2. All colored nodes are those that contain at least one data. The node size indicates how many feature vectors fall into each node, i.e., the data density, also specified by the number displayed within each node. The gray hexagons among the nodes represent instead the Euclidean distances visualized on the map according to a gray scale [9]. In this way, the dark gray hexagons correspond to large distances between the nodes and therefore between the clusters.

Observing Fig. 6.2, we note that three main clusters are identified on the SOM indicated by black circles: The first one of 81 elements (yellow) includes products from explosive eruptions of both Campi Flegrei and Somma–Vesuvius. The second one of 79 samples (green) contains mainly all the effusive products of Somma–Vesuvius eruptions and some poorly evolved products of Campi Flegrei. Finally, the third cluster of 28 components (pink) includes only explosive and effusive products from the group of “foreign” samples (ET).



**Fig. 6.2** SOM clustering obtained by using a hexagonal map with  $5 \times 3 = 15$  nodes. The node size indicates the data density, also displayed as a number within each node. The gray hexagons separating the nodes represent the Euclidean distances among them, where dark gray hexagons corresponding to high distance values

## 6.4 Conclusions

This work presents the preliminary results of a cluster analysis based on the self-organizing map (SOM) neural network to investigate possible relations between the products of different volcanic complexes in the Neapolitan area: Somma–Vesuvius and Campi Flegrei. The examined database includes chemical analyses (Fig. 6.1) that give information both on the magmatic mantle source and on the evolutionary processes and the residence times in the crustal magma chambers. It is composed of 271 volcanic patterns, divided into four classes, named CF, VF, VX, and ET, as previously described. The “foreign” eruption patterns (ET), related to the Stromboli volcano, have been considered to verify the robustness of the proposed classification method. A pre-processing phase has been performed in order to standardize the data used as input for the SOM. The final feature vector has a size of 23 elements.

The SOM clustering identifies three main clusters. In particular, the foreign patterns (ET) result well separated from the other ones, being mainly grouped in a single node (Fig. 6.2).

In conclusion, the obtained results suggest the ability of SOM neural network to associate volcanic rock suites based on their geochemical imprint and can be consistent with the hypothesis that there might be a common magma source beneath the whole Neapolitan area [11].

## References

1. Ali, M., Chawathé, A.: Using artificial intelligence to predict permeability from petrographic data. *Comput. Geosci.* **26**(8), 915–925 (2000)
2. Aminian, K., Ameri, S.: Application of artificial neural networks for reservoir characterization with limited data. *J. Petrol. Sci. Eng.* **49**(3–4), 212–222 (2005)
3. Corsaro, R.A., Falsaperla, S., Langer, H.: Geochemical pattern classification of recent volcanic products from Mt. Etna, Italy, based on Kohonen maps and fuzzy clustering. *Int. J. Earth Sci.* **102**(4), 1151–1164 (2013)
4. Dowla, F. U., Rogers, L. L.: Solving problems in environmental engineering and geosciences with artificial neural networks. Mit Press (1995)
5. Esposito, A.M., Giudicepietro, F., D’Auria, L., Scarpetta, S., Martini, M.G., Coltell, M., Marinaro, M.: Unsupervised neural analysis of very-long-period events at Stromboli volcano using the self-organizing maps. *Bull. Seismol. Soc. Am.* **98**(5), 2449–2459 (2008)
6. Giudicepietro, F., Esposito, A.M., Ricciolini, P.: Fast discrimination of local earthquakes using a neural approach. *Seismol. Res. Lett.* **88**(4), 1089–1096 (2017)
7. Goswami, S., Chakraborty, S., Ghosh, S., Chakrabarti, A., Chakraborty, B.: A review on application of data mining techniques to combat natural disasters. *Ain Shams Eng. J.* **9**(3), 365–378 (2018)
8. Huffman, W. S.: Geographic information systems, expert systems and neural networks: disaster planning, mitigation and recovery. *WIT Trans. Ecol. Environ.* **50** (2001)
9. Kohonen T., Hynninen, J., Kangas, J., Laaksonen, J.: SOM\_PAK: the self-organizing map program package, Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo, Finland (1996) ([http://www.cis.hut.fi/research/som\\_lvq\\_pak.shtml](http://www.cis.hut.fi/research/som_lvq_pak.shtml))

10. Kohonen, T.: Self-Organizing Maps. Series in information sciences, vol. 30, 2nd edn. Springer, New York (1997)
11. Pappalardo, L., Mastrolorenzo, G.: Rapid differentiation in a sill-like magma reservoir: a case study from the Campi Flegrei caldera. *Sci. Rep.* **2**, 712 (2012)
12. Peccerillo, A., De Astis, G., Faraone, D., Forni, F., Frezzotti, M.L.: Compositional variations of magmas in the Aeolian arc: implications for petrogenesis and geodynamics. In: From: Lucchi, F., Peccerillo, A., Keller, J., Tranne, C. A. Rossi, P.L. (eds.). *The Aeolian Islands Volcanoes*. Geological Society, vol. 37, pp. 491–510. London, Memoirs (2013)
13. Yu, M., Yang, C., Li, Y.: Big data in natural disaster management: a review. *Geosciences* **8**(5), 165 (2018)

## Chapter 7

# Toward an Automatic Classification of SEM Images of Nanomaterials via a Deep Learning Approach



**Cosimo Ieracitano, Fabiola Pantó, Nadia Mammone, Annunziata Paviglianiti, Patrizia Frontera and Francesco Carlo Morabito**

**Abstract** Nanofibrous materials produced by electrospinning process may exhibit characteristic localized defects and anomalies (i.e., beads, speck of dust) that make the nanostructure a network of nonhomogeneous nanofibers, unsuitable for industrial production at large scale of the nanoproducts. Therefore, monitoring and controlling the quality of nanomaterials production has become increasingly important and intelligent anomalies detection systems have been emerging. In this study, we propose an innovative framework based on machine (deep) learning for automatic anomaly detection. Specifically, a deep convolutional neural network (CNN) is proposed to automatically classify scanning electron microscope (SEM) images of homogeneous (HNF) and nonhomogeneous nanofibers (NHNF), interpreted as two different categories. The proposed approach has been validated on experimental SEM images acquired through SEM images analyzer on polyvinylacetate (PVAc) nanofibers produced by electrospinning process. Experimental results showed that the designed deep CNN achieved an accuracy rate up to 80% and average precision, recall, F-score of, 78.5, 79, and 78.5%, respectively. These promising results encourage the use of this effective technique in industrial production.

---

C. Ieracitano (✉) · F. Pantó · P. Frontera · F. C. Morabito  
DICEAM Department, University Mediterranea of Reggio Calabria,  
89124 Reggio Calabria, Italy  
e-mail: [cosimo.ieracitano@unirc.it](mailto:cosimo.ieracitano@unirc.it)

N. Mammone  
IRCCS Centro Neurolesi Bonino-Pulejo, Via Palermo c/da Casazza,  
SS. 113 98124 Messina, Italy

A. Paviglianiti  
DIIES Department, University Mediterranea of Reggio Calabria,  
89124 Reggio Calabria, Italy

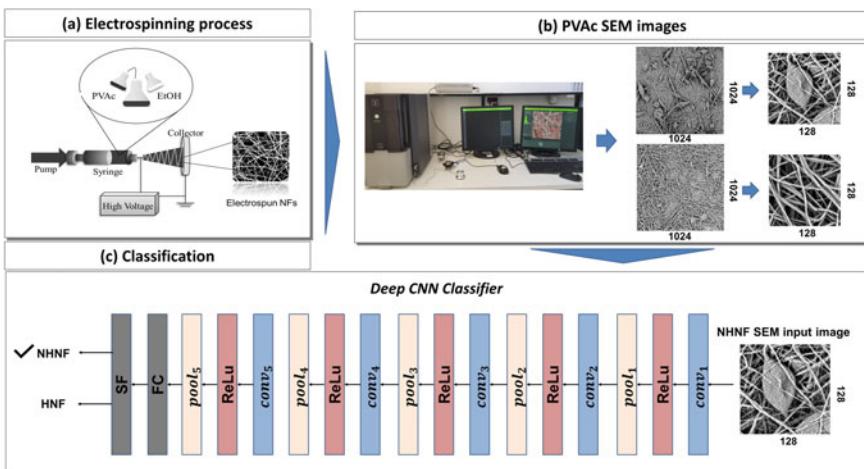
## 7.1 Introduction

Nanofibrous materials are ultra-fine fibers with diameters lower than  $10^2$  nm, typically produced by an effectiveness and efficient technique called *electrospinning* (or electrostatic spinning) [1]. In recent years, the applications of such nanostructures have attracted a great deal of interest in fields as biomedicine [2], electronics [3], drug delivery, and tissue engineering [4]. However, the production of nanofibers (NF) is still difficult to monitor as several parameters (i.e., applied voltage, polymeric concentration, temperature, etc.) may cause structural defects during the electrospinning process [5]. The most common defects are *beads*, namely micro- or nanopolymeric particles, obtained mainly when the concentration of the solution is very low. These anomalies reduce the large surface area per unit volume, which is a typical advantage of electrospinning over competing techniques of production and influence time and cost of production. The most efficient approach of monitoring the electrospun material is acquiring scanning electron microscope (SEM) images from the nanofibrous sample and analyzing its structure. As the expert visual inspection is not the most effective method to identify defects, also for practical reasons, there is an increasing interest in developing automatic anomalies detection systems (ADS) through the analysis of SEM images. However, a few works are reported in the literature. Carrera et al. [6] used the so-called *novelty detection* algorithm also known as a one-class classification to address the issue of anomaly detection within SEM images. The dictionary of only normal (anomaly-free) images patches previously developed by Boracchi et al. [7] was employed. This dictionary was used during the test to identify defects in a patch-wise modality. Experimental results showed that the proposed method was able to outperform state-of-art algorithms (i.e., STSIM, coding) and provided good performance in detecting small defects. Recently, advanced machine learning techniques known as *deep learning* (DL) have been also employed in this context [8]. It has been proved that DL algorithms achieve human-level performances in several real-world applications (i.e., speech recognition [9], biomedicine [10, 11], cybersecurity [12], nondestructive testing and evaluation [13]), so DL-based systems for anomalies detection in SEM images have been emerging. Specifically, Carrera et al. [14] proposed a DL method based on convolutional neural network (CNN) for automatic detection and localization of defects. They used the ResNet-18 network pre-trained on scene and object images defined by the ILSVRC 2015 competition and built a dictionary of features vectors extracted from normal patches. Abnormal patches were evaluated through similarity between a testing patch and an anomaly-free patch of the dictionary. The performances were measured in terms of ROC curve and coverage factor. Experimental results show that the proposed framework outperformed the state of the art of about 5%. However, similarly to their previous works, the authors developed the DL system through a one-class classifier. In this paper, an automatic classification of SEM images of nanomaterials via a deep learning approach is proposed. Specifically, a deep convolutional neural network is developed to detect SEM images of homogeneous nanofibers (HNF, anomalies-free) and SEM images of nonhomogeneous nanofibers (NHNF, with defects). A dataset of 160

SEM images (85 NHNF and 75 HNF) of polyvinylacetate (PVAc) nanofibers were manually collected after electrospinning process at the *Materials for Environmental and Energy Sustainability Laboratory* of the University Mediterranea of Reggio Calabria (Italy). Experimental results (Table 7.2) showed that the 2-way deep CNN classifier achieved accuracy up to 80%. The rest of this paper is organized as follows. Section 7.2 describes the proposed methodology, including the electrospinning process, the experimental setup and the proposed deep CNN classifier. Section 7.3 discusses the results achieved. Section 7.4 concludes this paper.

## 7.2 Methodology

The flowchart of the procedure is illustrated in Fig. 7.1. Firstly, the polymeric solution of PVAc dissolved in EtOH solvent is prepared and used to produce PVAc nanofibers by electrospinning process (Fig. 7.1a). The morphology of the electrospun nanofibers is examined by a scanning electron microscope, and SEM images sized  $1024 \times 1024$  of homogeneous and not-homogeneous nanofibers are manually collected. Afterward, given the  $j$ th SEM image, a subregion (sized  $128 \times 128$ ) is selected (Fig. 7.1b). This operation has been necessary due to the limitations of the available processor. Finally, a deep learning classifier based on CNN is employed to identify HNF and NHNF images (Fig. 7.1c).



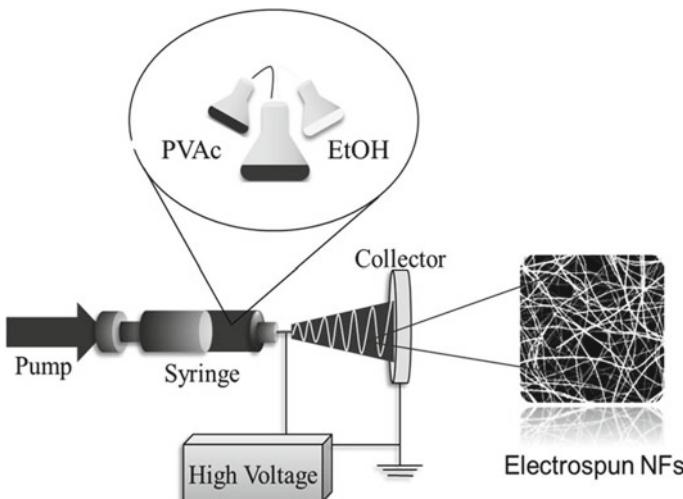
**Fig. 7.1** Flowchart of the method proposed

### 7.2.1 Electrospinning Process

The electrospinning apparatus is schematized in Fig. 7.2. It includes three basic components: a high voltage supply, an extruder, and a metallic collector screen. The polymeric solution is initially placed into a glass syringe and pushed through the metallic needle by the injection pump, which allows controlling the flow rate. A high voltage is applied between the needle (anode) and the collector (cathode), which are electrostatically charged to a different electric potential. As the electric field increases, the formed droplet loses surface tension and takes the form of a cone, referred to as Taylor cone. When the electrostatic force exceeds the surface tension, the polymeric jet is stretched within the high electric field; meanwhile, the solvent evaporates and is deposited on the collector in the form of nanofibers.

The viscosity has a great influence on the jet and diameter of nanofibers. Specifically, when the concentration of the solution is very low, micro- or nanopolymeric particles are obtained and the electrospray phenomenon is observed [15]. However, applied voltage, tip-collector distance, and flow rate have also remarkable effects on the fibers [16].

**Material** In this study, polyvinylacetate (PVAc; average molecular weight (Mw): 170,000) and ethanol (EtOH) were used as polymer and solvent, respectively. The spinnable solution was prepared by dissolving PVAc in EtOH and stirring until a clear solution was obtained. The spinning process was carried out at  $20 \pm 1$  °C temperature and 40% relative air humidity, using a CH-01 Electro-spinner 2.0 (Linari Engineering s.r.l.) and a 20 mL syringe, equipped with a 40 mm long 0.8 mm gauge stainless steel needle. All reactants were supplied by Sigma-Aldrich. The effects of concentration



**Fig. 7.2** Electrospinning setup

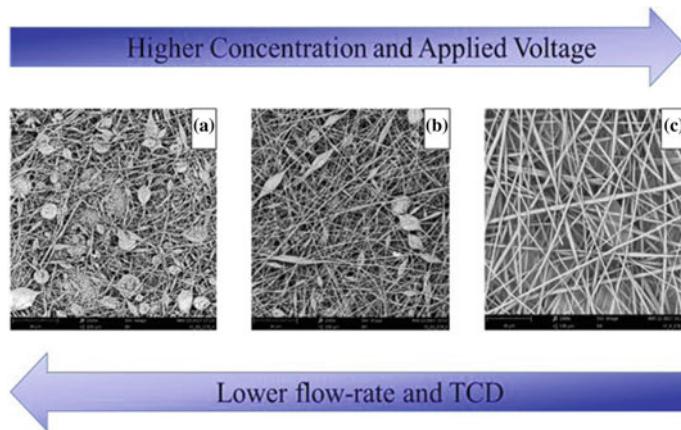
$(v_1)$ , applied voltage ( $v_2$ ), flow rate ( $v_3$ ), and tip-collector distance ( $v_4$ ) were evaluated to study the nanofibers production process. The Phenom Pro-X scanning electron microscope (SEM), equipped with an energy-dispersive X-ray (EDX) spectrometer, examined the morphology of the electrospun fibers. Finally, the average diameter, the distribution of the nanofibers, and the detection of beads were obtained by using Fibermetric software (an SEM image analyzer).

**Experimental Setup and Dataset Description** Sixteen experiments ( $\xi_i$ ,  $i = 1, 2, \dots, 16$ ) were carried out by changing one parameter at time in the following well-defined range of working: 10–25 wt.% concentration; 10–17.5 kV applied voltage; 100–300  $\mu\text{L}/\text{min}$  flow rate; 10–15 cm TCD. Table 7.1 reports the details of the experiments. However, since the purpose of this study was the development of a classification system based on SEM images, the average diameter was not taken into account and it is not reported in Table 7.1.

Given the  $i$ th material sample under analysis, 10 significant and representative areas were arbitrarily selected and evaluated through the SEM images analyzer, in order to augment the dataset. A grand total of 160 SEM images sized  $1024 \times 1024$  was collected. Each image was examined by an expert operator and labeled as SEM image of homogeneous nanofibers (HNF) or nonhomogeneous nanofibers (NHNF). Specifically, 75 images were classified as HNF and 85 as NHNF. Figure 7.3 shows representative SEM images of NHNF and HNF achieved during the experiments. As can be observed, lower concentrations cause low viscosity of the solution, instability

**Table 7.1** Electrospinning setup of the 16 experiments

$\xi$	Concentration ( $v_1$ ) (%wt)	Applied voltage ( $v_2$ ) (kV)	Flow rate ( $v_3$ ) ( $\mu\text{L}/\text{min}$ )	TCD ( $v_4$ ) (cm)
$\xi_1$	10	15	10	100
$\xi_2$	15	10	10	100
$\xi_3$	15	13.5	10	100
$\xi_4$	15	15	10	100
$\xi_5$	15	15	10	200
$\xi_6$	15	15	10	300
$\xi_7$	15	15	12.5	100
$\xi_8$	15	15	13.5	100
$\xi_9$	15	15	15	100
$\xi_{10}$	20	10	10	100
$\xi_{11}$	20	11.5	10	100
$\xi_{12}$	20	13.5	10	100
$\xi_{13}$	20	15	10	100
$\xi_{14}$	20	16	10	100
$\xi_{15}$	20	17.5	10	100
$\xi_{16}$	25	15	10	100



**Fig. 7.3** Effect of the parameters variation on the morphology. **a–b** SEM images of non-homogeneous nanofibers (NHF) due to the presence of beads. **c** SEM image of homogeneous nanofibers (HNF)

of the polymeric jet and consequently beads structures (Fig. 7.3a). With increasing TCD, the electrospun solution is affected by a less intense electric field and causes a mixed morphology of fibers and beads (Fig. 7.3b). Higher applied voltages produce smaller dimensions of beads as they are elongated and stretched by the higher electric field. Therefore, high voltages, together with high concentrations, produce homogeneous networks of nanofibers (Fig. 7.3c).

However, due to the computational limit of the available processor, we did not process the whole SEM sized  $1024 \times 1024$ . Given the  $j$ th SEM image belonging to NHNF or HNF class, it was firstly split into 64 sub-images sized  $128 \times 128$ , and then, one representative sub-image was selected and used as input of the proposed convolutional neural network. It is worth noting that there is no downsampling of the original SEM image to prevent distortion of the size of nanofibers.

### 7.2.2 Convolutional Neural Network

Convolutional neural networks (CNNs) are deep learning architectures able to learn discriminating features directly from raw input data through a deep hierarchical organization. A typical CNN consists of subsequent modules of *convolution*, *activation*, and *pooling* layers. The convolutional layer performs the convolution operation between a set of  $F_j$  learnable filters (or kernels) and the input map  $I_i \in R^{h \times w}$ . The result is the so-called *features map*  $O_j = \sum I_i * F_j + B_j$  where  $B_j$  is the bias term and  $*$  indicates the convolution operator. Each filter convolves with a local area of  $I_i$  and then scans the whole plane with a stride  $s$ , sharing the same values of weights. The  $O_j$  features map is sized  $o_1 \times o_2$  where  $o_1 = \frac{h-f_1+2p}{s} + 1$ ,  $o_2 = \frac{w-f_2+2p}{s} + 1$  and  $p$  is the

zero padding parameter used to control the output size by padding the input edges with zeros. The convolutional layer is typically followed by the “*Rectified Linear Unit*” (ReLU,  $f(z) = \max(0, z)$ ) activation function, as it aids the system in generalization and improves learning time [17]. The pooling layer reduces the input features map through an average (*average pooling*) or maximum (*max-pooling*) operation. In this study, the max-pooling operation is used. The  $F_j$  filter scans the input features map with stride  $\tilde{s}$  producing a subsampled representation of  $O_j$  sized  $\tilde{o}_1 \times \tilde{o}_2$  where  $\tilde{o}_1 = \frac{v_1 - \tilde{f}_1}{\tilde{s}} + 1$  and  $\tilde{o}_2 = \frac{o_2 - \tilde{f}_2}{\tilde{s}} + 1$ . Finally, the learned features are the input of a standard multi-layer neural network (MLP) for the classification task.

**Architecture Proposed and Learning Setup** Fig. 7.1c shows the schematic of the proposed deep CNN. It contains 5 convolutional (*conv*) layers, 5 max pooling (*pool*), 1 fully connected layer (*FC*) with 40 hidden neurons, and 1 softmax (*SF*) output layer for binary classification task (NHNF-HNF). All convolutional layers have filters size of  $3 \times 3$ , stride  $s = 1$  and padding parameter  $p = 1$ ; whereas, all max-pooling layers have filters size of  $2 \times 2$  and stride  $s = 2$ . The rectified linear unit is employed as activation function and batch normalization is applied to the hidden layers to avoid covariate shift phenomenon [18]. The details of architecture configuration are reported in Table 7.2. Learning setup was based on the practical recommendations of [19, 20]. The layers are initialized from a Gaussian distribution with zero mean and standard deviation of  $10^{-2}$ . The stochastic gradient descent (SGD) algorithm with the momentum of 0.9, weight decay of  $10^{-4}$ , learning rate of 0.001 and mini-batch size of 32 shuffled SEM images is used to train the deep neural network. The network was implemented using MATLAB R2017b (The MathWorks, Inc., Natick, MA, USA) and trained for 300 iterations on a single CPU of HP xw4600 workstation with 12 GB RAM.

### 7.3 Experimental Results

The dataset included 160 SEM images (75 HNF and 85 NHNF) sized  $128 \times 128$  (Sect. 7.2.1, *Experimental setup and Dataset Description*). A subset of the dataset was used to train the deep CNN (70% training dataset) and the remaining 30% was used to test the trained model. The performance of the proposed deep CNN was evaluated using standard metrics: precision, recall, F-score, and accuracy (Tables 7.3 and 7.4):

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (7.1)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (7.2)$$

$$\text{F\_score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7.3)$$

**Table 7.2** Layers configuration of the deep CNN proposed

	<i>conv</i> <sub>1</sub>	<i>pool</i> <sub>1</sub>	<i>conv</i> <sub>2</sub>	<i>pool</i> <sub>2</sub>	<i>conv</i> <sub>3</sub>	<i>pool</i> <sub>3</sub>	<i>conv</i> <sub>4</sub>	<i>pool</i> <sub>4</sub>	<i>conv</i> <sub>5</sub>	<i>pool</i> <sub>5</sub>	<i>FC</i> <sub>1</sub>	<i>FC</i> <sub>2</sub>
Input (I)	128 × 128	128 × 128	64 × 64	64 × 64	32 × 32	32 × 32	16 × 16	16 × 16	8 × 8	8 × 8	4 × 4	2048 × 1
Number of filters (F)	16	16	32	32	64	64	96	96	128	128	—	40 × 1
Size of filters (f1 × f2)	3 × 3	2 × 2	3 × 3	2 × 2	3 × 3	2 × 2	3 × 3	2 × 2	3 × 3	2 × 2	—	—
Stride (s)	1	2	1	2	1	2	1	2	1	2	—	—
Padding (p)	1	—	1	—	1	—	1	—	1	—	—	—
Output (O)	128 × 128	64 × 64	64 × 64	32 × 32	32 × 32	16 × 16	16 × 16	8 × 8	8 × 8	4 × 4	40 × 1	2 × 1

**Table 7.3** Accuracy performance of deep CNN with different numbers of hidden layers

CNN architecture	Accuracy (%)
$CNN_1[conv_1 + relu_1 + pool_1] + SF$	50
$CNN_2[CNN_1 + conv_2 + relu_2 + pool_2] + SF$	66
$CNN_3[CNN_2 + conv_3 + relu_3 + pool_3] + SF$	68
$CNN_4[CNN_3 + conv_4 + relu_4 + pool_4] + SF$	70
$CNN_5[CNN_4 + conv_5 + relu_5 + pool_5] + SF$	72
$CNN_5^*[CNN_4 + conv_5 + relu_5 + pool_5] + FC + SF$	80

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (7.4)$$

where true positives (TP) represent the number of SEM images correctly identified as images of nonhomogeneous nanofibers; true negatives (TN) represent the number of SEM images correctly identified as images of homogenous nanofibers (anomalies-free); false positives (FP) are the number of normal images erroneously identified as images of nanofibers nonhomogeneous (with anomalies); and false negatives (FN) are the number of anomalies SEM images misclassified as anomalies-free. The best CNN configuration was chosen by evaluating the effects on the performance of changing the number of hidden layers. Table 7.3 reports the outcome of the experiments on the test set. Experimental results showed that the  $CNN_1$  architecture with only one module of convolution ( $conv_1$ ), ReLu ( $relu_1$ ) and pooling ( $pool_1$ ) layer produced the lowest performance (50% accuracy) and was not capable of discriminating SEM images anomalies-free and with defects; whereas, the best classification result was reached with  $CNN_5$  achieving accuracy rate up to 72%. All the networks ( $CNN_1$ – $CNN_5$ ) included a softmax output layer for binary classification. However, it was observed that the accuracy performance improved of 8% by adding a fully connected layer with 40 hidden neurons and ReLu activation function. As can be observed in Table 7.4, the  $CNN_5 + FC + SF$  architecture achieved accuracy rate up to 80% and average precision, recall, and F\_score of, 78.5, 79, and 78.5%, respectively. To our best knowledge, this is the first work on classification among SEM images of homogeneous (anomaly-free) and nonhomogeneous (with anomaly) PVAc nanofibers (produced by electrospinning process) by using DL techniques. However, it is worth mentioning that, recently, Napoletano et al. [14] proposed a

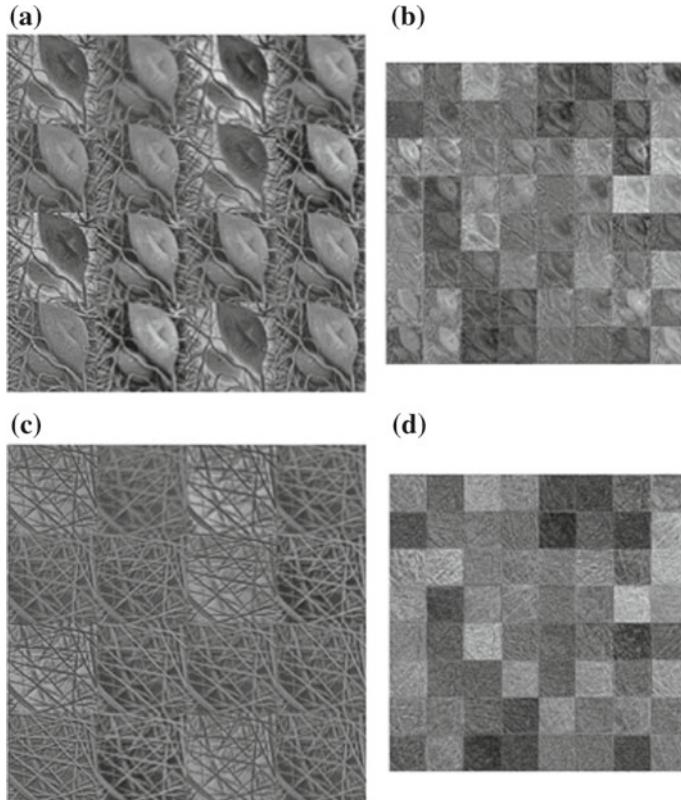
**Table 7.4** Precision, Recall, and F\_score of the deep  $CNN_5^*$  classifier

SEM image	Precision (%)	Recall (%)	F_score (%)
NHNF	84	82	83
HNF	73	76	74
Average	78.5	79	78.5

per-pixel one-class classification based on CNN but for detection and localization of defects within SEM images. Specifically, the proposed patch-wise-based method identifies defects through visual similarity between the test patch under analysis and a reference dictionary of normal subregions.

## 7.4 Conclusions

In this paper, we proposed a deep learning-based system to detect anomalies in scanning electron microscope (SEM) images of nanofibrous materials produced by the electrospinning process. Specifically, a deep convolutional neural network was developed to classify SEM images of homogeneous (HNF) and nonhomogeneous



**Fig. 7.4** Features maps learned by  $conv_1$  and  $conv_3$  on a SEM image input of homogeneous (HNF) and nonhomogeneous nanofibers (NHNF). **a–c** 16 feature maps sized  $128 \times 128$  learned by  $conv_1$  of NHNF and HNF, respectively. **b–d** 64 feature maps sized  $32 \times 32$  learned by  $conv_3$  of NHNF and HNF, respectively

nanofibers (NHNF). The polyvinylacetate (PVAc, Mw 170,000) dissolved in ethanol solvent was electrospun at the *Materials for Environmental and Energy Sustainability Laboratory* of the University Mediterranea of Reggio Calabria (Italy), and 16 experiments were carried out under different experimental conditions. A total of 160 images of PVAc nanofibers was extracted from the scanning electron microscope according to the procedure described in Sect. 7.2 (*Experimental setup and Dataset Description*). A deep CNN was employed to learn the most relevant features from the raw SEM input data and classify PVAc HNF and NHNF images. Figure 7.4 shows examples of the features learned by  $conv_1$  and  $conv_3$  layer on a NHNF and HNF SEM image sized  $128 \times 128$ . Experimental results showed that the proposed deep CNN was able to correctly discriminate HNF and NHNF with accuracy up to 80%. However, it is to be noted that this study supposes to be a preliminary work for a more accurate and versatile system. Future works will address the limit of SEM image processing sized  $1024 \times 1024$  (and above). Moreover, a larger number of experiments on PVAc polymer will be performed through the electrospinning process. In addition, in the future, we intend to integrate the work here presented with the methodology presented in [14] for detection and localization of defects in SEM images.

## References

1. Huang, Z.M., Zhang, Y.Z., Kotaki, M., Ramakrishna, S.: A review on polymer nanofibers by electrospinning and their applications in nanocomposites. *Compos. Sci. Technol.* **63**(15), 2223–2253 (2003)
2. Agarwal, S., Wendorff, J.H., Greiner, A.: Use of electrospinning technique for biomedical applications. *Polymer* **49**(26), 5603–5621 (2008)
3. Miao, J., Miyauchi, M., Simmons, T.J., Dordick, J.S., Linhardt, R.J.: Electrospinning of nano-materials and applications in electronic components and devices. *J. Nanosci. Nanotechnol.* **10**(9), 5507–5519 (2010)
4. Sill, T.J., von Recum, H.A.: Electrospinning: applications in drug delivery and tissue engineering. *Biomaterials* **29**(13), 1989–2006 (2008)
5. Deitzel, J.M., Kleinmeyer, J., Harris, D., Tan, N.B.: The effect of processing variables on the morphology of electrospun nanofibers and textiles. *Polymer* **42**(1), 261–272 (2001)
6. Carrera, D., Manganini, F., Boracchi, G., Lanzarone, E.: Defect detection in sem images of nanofibrous materials. *IEEE Trans. Ind. Inform.* **13**(2), 551–561 (2017)
7. Boracchi, G., Carrera, D., Wohlberg, B.: Novelty detection in images by sparse representations. In: 2014 IEEE Symposium on Intelligent Embedded Systems (IES), pp. 47–54. IEEE (2014)
8. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
9. Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A.r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T.N., et al.: Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Process. Mag.* **29**(6), 82–97 (2012)
10. Ieracitano, C., Mammone, N., Bramanti, A., Hussain, A., Morabito, F.C.: A convolutional neural network approach for classification of dementia stages based on 2d-spectral representation of eeg recordings. *Neurocomputing* **323**, 96–107 (2019)
11. Gasparini, S., Campolo, M., Ieracitano, C., Mammone, N., Ferlazzo, E., Sueri, C., Tripodi, G.G., Aguglia, U., Morabito, F.C.: Information theoretic-based interpretation of a deep neural network approach in diagnosing psychogenic non-epileptic seizures. *Entropy* **20**(2), 43 (2018)

12. Ieracitano, C., Adeel, A., Gogate, M., Dashtipour, K., Morabito, F.C., Larijani, H., Raza, A., Hussain, A.: Statistical analysis driven optimized deep learning system for intrusion detection. In: International Conference on Brain Inspired Cognitive Systems. pp. 759–769. Springer (2018)
13. Morabito, C.F.: Independent component analysis and feature extraction techniques for ndt data. Mater. Eval. **58**(1), 85–92 (2000)
14. Napoletano, P., Piccoli, F., Schettini, R.: Anomaly detection in nanofibrous materials by cnn-based self-similarity. Sensors **18**(1), 209 (2018)
15. Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., Whitehouse, C.M.: Electrospray ionization for mass spectrometry of large biomolecules. Science **246**(4926), 64–71 (1989)
16. Lasprilla-Botero, J., Álvarez-Láinez, M., Lagaron, J.: The influence of electrospinning parameters and solvent selection on the morphology and diameter of polyimide nanofibers. Mater. Today Commun. **14**, 1–9 (2018)
17. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th international conference on machine learning (ICML-10). pp. 807–814 (2010)
18. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint [arXiv:1502.03167](https://arxiv.org/abs/1502.03167) (2015)
19. Goyal, P., Dollár, P., Girshick, R., Noordhuis, P., Wesolowski, L., Kyrola, A., Tulloch, A., Jia, Y., He, K.: Accurate, large minibatch sgd: Training imagenet in 1 hour. arXiv preprint [arXiv:1706.02677](https://arxiv.org/abs/1706.02677) (2017)
20. Bengio, Y.: Practical recommendations for gradient-based training of deep architectures. In: Neural networks: Tricks of the trade, pp. 437–478. Springer (2012)

# Chapter 8

## An Effective Fuzzy Recommender System for Fund-raising Management



Luca Barzanti, Silvio Giove and Alessandro Pezzi

**Abstract** In the social economics field that deals with the nonprofit organizations (NPOs), the fund-raising is a crucial activity that requires the management of a great number of quantitative and qualitative information regarding Donors and Contacts (i.e., potential donors). This data is normally stored in a structured database (DB) by each NPO, and it is clear that their effective processing by data science methods significantly improves the performances of the fund-raising campaigns. For this reason, the use of rigorous mathematical methods and decision support systems (DSS) has been playing a very important role in this context. The process of fund-raising is very complex and in part different depending on the characteristics of each organization. However, a common important feature is the role of the Contacts, and therefore, the method for turning the Contacts into actual Donors contextualized in the so-called giving pyramid is crucial from a strategic point of view. Recently, a recommender system (RS) has been proposed to optimize the Contacts' management, by computing the similarity of each Contact with respect to the Donors. In this contribution, we enhance and complete this model by considering both a large DB and two significant extensions of the model, obtaining in this way an effective and whole fuzzy RS. With respect to the DB, the availability of information is effectively exploited. As for the algorithm, a proper similarity measure is defined, based on the specificity of the context. Moreover, a complete estimation of the Contacts' characteristics is taken into account, by considering not only the frequency but the averaged amount of the

---

L. Barzanti (✉)

Department of Mathematics, University of Bologna, Piazza di Porta San Donato 5,  
40126 Bologna, Italy

e-mail: [luca.barzanti@unibo.it](mailto:luca.barzanti@unibo.it)

S. Giove

Department of Economics, Ca' Foscari University of Venice, Cannaregio 873,  
30121 Venice, Italy

e-mail: [sgiove@unive.it](mailto:sgiove@unive.it)

A. Pezzi

School of Economics and Management, University of Bologna, P.le della Vittoria 15,  
47121 Forlì, Italy

e-mail: [a.pezzi@unibo.it](mailto:a.pezzi@unibo.it)

gift as well, in the context of a nonparametric approach. The experimental results show the effectiveness of the proposed system.

## 8.1 Introduction

A crucial support for NPOs, which operate in the context of social economics, is the fund-raising (FR) activity, in which the resources for the mission of the association are collected, see Andreoni [2], Rosso [25]. FR strategies are therefore crucial for the achievement of the mission and, specifically, to reaching the goal of the current campaign, see Nudd [24], Sargeant [27]. Quantitative methods employing DB technologies have been studied and developed in the pertaining literature for making these strategies more effective, see Flory [15, 16], Kercheville [20]. The effective use of the information on Donors and Contacts (i.e., potential donors), which is in fact normally managed by an organized DB, is crucial for optimizing the resources for the campaign by selecting the most promising Donors/Contacts for the considered context. For a few years, the literature in the area of mathematical models and DSS has dealt with the fund-raising problem, determining an evolution, a strengthening, and a specialization of the proposed methods and algorithms. With respect to the first two objectives, see Verhaert [28], Barzanti [6]. Moreover, the consideration of the specific NPO's characteristics and their consequences in the modeling process is developed, e.g., in Barzanti [3, 4]. In these contributions and in the operative literature, see, e.g., Melandri [22, p. 7], it is also documented that the operative word of associations are using the results that have been achieved in this field. Moreover, in Moro [23] it is showed the analogy between the fund-raising process and some bank activities and the consequent correspondence of the employed methodologies.

One of the goals of a loyalty campaign is to involve new people in the mission of the association by their first donation. In operational language, the goal is to make some Contacts going up from the ground of the so-called giving pyramid, see Melandri [22], to the first level. From a strategic point of view, it is particularly important that a Contact becomes Donor at the first gift request, and therefore, it is fundamental to solicit the gift in a campaign that is suitable for that Contact.

In this paper, we focus our attention on such an essential aspect of the fund-raising process, by determining a *recommender system*, see Jananch [18], a branch of *Artificial Intelligence*,<sup>1</sup> that identifies the more suitable Contacts to reach in the current campaign. In fact, reaching a person obviously has a cost (that depends on the type of request) and the budget constraint requires a choice. For this purpose, a suitable *similarity measure* can be used for matching the profile of the Contacts with the profile of all the regular Donors in the association DB who usually donate for similar campaigns. In particular, by using the results obtained in econometrics [13, 21], and more specifically in Cappellari [10], Duffy [12], the variables of the personal profile that influence the gift probability are selected.

---

<sup>1</sup>Refer to Russell [26] for an exhaustive introduction to algorithms and aim of Artificial Intelligence.

The problem of the *Contacts' management* was before considered in Barzanti [5], where first attempt of a recommender system has been implemented. Although a relevant effort has been performed for modeling the process, the numerical experiments were limited to a few Contacts and Donors. Conversely, in the present paper we focus on a large structured DB that implies a relevant improvement of the significance of the results. The large amount of information is also exploited by the consideration of all the Donors' historical data, in particular of the gift volume, in addition to the frequency. This also allows the estimation of an amount for each Contact and the completion of the analysis, through the joined computation and elaboration of both the significant quantities.

Moreover, a specific measure of similarity is developed that takes into account with appropriate weights the most relevant variables of the personal profile, which are significant for the pertaining literature for measuring the gift probability.

The numerical experiments are developed with the contribution of ASSIF, the Italian fund-raiser association, and Philanthropy, a university fund-raising research center, that ensures the reliability of the proposed approach.

## 8.2 Contacts' Characterization

In this section, we consider Contacts' characterization. In Sect. 8.2.1, we analyze the similarity between a Donor and a Contact, based on a set of personal data, namely the financial situation, the age, and the qualification. The most important criterion, the financial situation, is considered as necessary for the similarity, and thus, a novel similarity function is proposed.<sup>2</sup> Subsequently, in Sect. 8.2.2 we compute, for each, an estimation of the Contact's expected gift volume and frequency, using a nonparametric approach based on the similarity between the considered Contact and each Donor in the DB. This nonparametric method is inspired by the fuzzy nearest neighbor techniques, see Aman [1], Keller [19].

### 8.2.1 The Personal Characterization for Donors and Contacts

The pertaining literature suggests that the propensity to gift depends on some personal parameters, like the financial situation (the most significant), the age, the risk aversion, and other, see, e.g., Cappellari [10] and Duffy [12]. In this contribution, among others, we consider as most representatives, the financial situation (*Fin*), the age (*Age*), and the qualification (*Qual*), that are:

---

<sup>2</sup>For general properties and a list of possible similarity measures, the interested reader can refer to Couso [11], Beg [7].

- (1) *Fin* (Real), the global income amount.
- (2) *Age* (real).
- (3) *Qual* (label), with four-ordered values: P (Ph.D.), B (Bachelor), H (High School), and O (Other).

The measure of the other parameters, as for the risk aversion, can indeed be difficult, while their influence on the gift attitude can be debatable (for instance, the presence of children can be source of effects of opposite sign). About the possible influence of such parameters on the propensity to gift, a future and deeper analysis is advisable, but it is beyond the purpose of this contribution. For this reason, in this paper we consider only the three most important parameters, *Fin*, *Age*, *Qual*, even if the inclusion of other variables would not change the methodological framework. The *Fin* parameter is clearly the most significant, given that a strong correlation exists among the personal richness and the gift probability (and amount). Moreover, this represents a necessary condition, given that financial scarcity is normally a serious obstacle for the charity. In the absence of similarity for *Fin* between a Donor and a Contact, the entire similarity is null and cannot be compensated by high values of similarity between *Age* and *Qual*. The personal variables can be collected into a vector, and thus, the *i*th Donor can be represented as the vector:

$$W(i) = \{Fin(i), Age(i), Qual(i), f(i), v(i), r(i)\}$$

being  $(F, V, \rho)$  the frequency, the (average) volume of past gifts and the statistical robustness. The reader can refer to Barzanti [5] for a detailed computation of  $(F, V, \rho)$ . The first three components of the vector  $W_i$  are used to compare the *i*th Donor with a Contact, using a suitable similarity measure, given that the Contact personal data is known. For this purpose, it is convenient to consider a given Contact and, for its value of the *Fin* parameter, to define a fuzzy moving window, centered in the Contact value. The same is done for the *Age* parameter, but for *Fin* the amplitude depends on the value of the variable itself. Let us define the following moving fuzzy window:

$$MOV(a, b, c, d) = \begin{cases} 1, & b \leq x \leq c \\ \frac{x-a}{b-a}, & a \leq x < b \\ \frac{d-x}{d-c}, & c < x \leq d \\ 0, & x < a, \text{ or } x > d \end{cases} \quad (8.1)$$

The fuzzy window for *Fin* is then defined as follows, given that, for the considered *j*th Contact, is  $Fin_j = \lambda$ :

$$\begin{aligned} G_j(x) &= MOV(a_j, b_j, c_j, d_j), \text{ where} \\ a_j &= \lambda(1 - \beta), \quad b_j = \lambda(1 - \alpha), \quad c_j = \lambda(1 + \alpha), \quad d_j = \lambda(1 + \beta) \end{aligned} \quad (8.2)$$

and  $0 < \alpha < \beta < 1$  define the upper and lower amplitude of the moving fuzzy window, both monotonically dependent on the current value of  $Fin_j = \lambda$ . Then,

the  $Fin$  similarity between the  $j$ th Contact and the  $i$ th Donor,  $SimFin(i, j)$ , is  $SimFin(i, j) = G_j(Fin_i)$ . A similar formulation can be applied to the  $Age$  parameter, but, in this case, the amplitude can be considered as a constant. Thus, if  $Age_j = \mu$ :

$$\begin{aligned} H_j(x) &= MOV(e_j, f_j, g_j, h_j) \\ e_j &= \mu - \gamma, \quad f_j = \mu - \delta, \quad g_j = \mu + \delta, \quad h_j = \mu + \gamma \end{aligned} \quad (8.3)$$

and  $SimAge(i, j) = H_j(Age_i)$  with  $\delta < \gamma$ . Finally, with respect to the third variable  $Qual$ , formed by ordered classes, we consider the similarity between two classes as a two entries function:

$$SimQual(i, j) = L_j(Qual_i, Qual_j) \quad (8.4)$$

This requires the assessment of six parameters (all the possible combinations of two different classes).

Finally, the complete similarity between the Donor  $j$  and the Contact  $i$  is:

$$Sim(D_i, C_j) = SimFin(i, j) \frac{(\omega_1 + \omega_2 SimAge(i, j) + \omega_3 SimQual(i, j))}{\omega_1 + \omega_2 + \omega_3} \quad (8.5)$$

where  $\omega_1, \omega_2, \omega_3$  are suitable weights ( $\omega_1, \omega_2, \omega_3 > 0, \omega_1 + \omega_2 + \omega_3 = 1$ ).

### 8.2.2 Nonparametric Estimation of Volume and Frequency

A possible way to produce a numerical estimation of an unknown distribution from a set of data is based on the *nearest neighbor* function approach, see Keller [19], Bhatia [8]. Roughly speaking, the probability estimation is based on an average combination of the occurrences of the items, weighted by the similarity degree between each item and the case. In some cases, a *kernel function* is formed by the set of items that are *similar* to the current case, see Wand [29], Canestrelli [9]. The unknown probability is estimated using, as conditioning parameters (weights), the kernel-based similarities of the case study to each item inside a suitable neighbor. The closer the item, the greater the weight, given that more “importance” (credibility) is given to the most similar items. In Fan [14], the reader can find a complete and detailed explanation of the kernel-based estimation methods, while in Keller [19] a fuzzy  $K$ -nearest neighbor is described. From the quoted literature, if  $B = \{b_1, b_2, \dots, b_n\}$  is a set of input vectors (*incomplete* pattern, see Giove [17]), with corresponding output value  $Y = \{y(b_1), \dots, y(b_n)\}$ , the expected value of the probability distribution of an (other) item  $b$ , conditioned to  $B$ , can be computed as:

$$E(F) = \frac{\sum_{b \in B} Sim(a, b) \cdot y(b)}{\sum_{b \in B} Sim(a, b)} \quad (8.6)$$

We use this formulation to compute both the (expected value of) frequency and volume, using as similarity the function described in Sect. 8.2.1, corrected by the statistical robustness of the Donor. Namely, for the same reason as for the similarity, we give less weight to those Donors who are less robust, in the sense that their own history about the gifts is limited to few requests (and corresponding answers, if any). Thus, let  $F(j)$ ,  $V(j)$  be the estimated values of (average) frequency and volume of the Contact  $j$ :

$$F(j) = \frac{\sum_{i=1}^n \text{Sim}(i, j) \cdot r(i) \cdot f(i)}{\sum_{i=1}^n \text{Sim}(i, j) \cdot r(i)} \quad (8.7)$$

$$V(j) = \frac{\sum_{i=1}^n \text{Sim}(i, j) \cdot r(i) \cdot v(i)}{\sum_{i=1}^n \text{Sim}(i, j) \cdot r(i)} \quad (8.8)$$

Moreover, to avoid a meaningless estimation, we impose that at least  $K$  Donors exist, with similarity and robustness greater than a specified threshold; otherwise, the Contact is classified as *undecidable* and is inserted in a separated list. Then, if we define  $SR(i) \in \{\text{Sim}(i, j) \cdot r(i)\}$  as the ordered similarities multiplied by the robustness, i.e.,  $SR(1) \geq SR(2) \geq \dots \geq SR(n)$ , the rule classifies a Contact as undecidable if:

$$\sum_{i=1}^K SR(i) < \sigma \quad (8.9)$$

being  $\sigma > 0$  a specified threshold (i.e.,  $\sigma = K$ ). Finally, for all the promising Contacts, a proxy of the estimation's robustness is computed as the sum of all the similarities:

$$R(j) = \sum_{i=1}^n \text{Sim}(i, j) \cdot r(i) \quad (8.10)$$

The considered Contact is thus characterized by the triple:  $F(j)$ ,  $V(j)$ ,  $R(j)$ . This information can be used for the Contacts' evaluation and/or screening, based on the system's user own preferences, and thus constitutes the engine of a decision support system in the field of the nonprofit organizations. The numerical results show the good performances of the proposed approach.

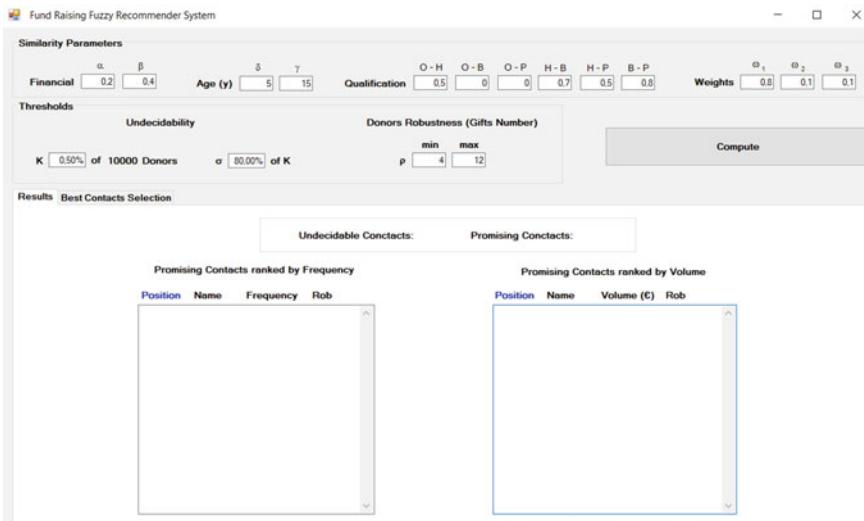
### 8.3 Computational Results

The numerical experiments have been performed in collaboration with ASSIF<sup>3</sup> and Philanthropy Centro Studi<sup>4</sup>; this way, a real context is considered, in particular with respect to the simulated DB construction, implemented by the criteria of the giving pyramid for a medium-sized organization. A SQL Server DB with 10,000 Donors and 2000 Contacts is considered, while the system is designed in Visual Basic, using SQL language.

The decision maker (DM) sets up the similarity parameters, with reference to the moving fuzzy windows in formulas (8.2), (8.3), the assessment required by (8.4), and the weights in (8.5). The thresholds of undecidability [formula (8.9)] and robustness are also set (see below). Figure 8.1 shows the whole graphical interface.

The parameters are set like in Fig. 8.1. In particular, notice that, for the undecidability,  $K$  is set as (the integer part of) a percentage of the DB's numerosity and  $\sigma$  is a specified percentage of  $K$ , while for the Donors' robustness  $\rho$ , the thresholds are set in terms of gifts number, referring to Barzanti [5], as pointed out in Sect. 8.2.1.

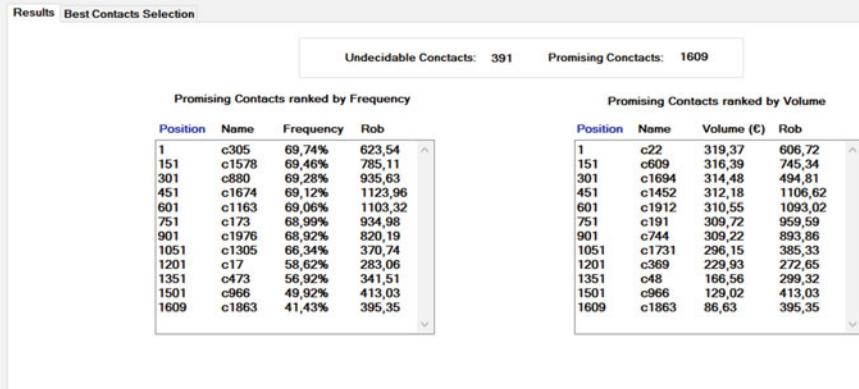
The results of the estimated Contacts' gift frequencies and volumes [with reference to formulas (8.7) and (8.8)] and of the robustness index ( $Rob$ ) for the estimation [formula (8.10)] are presented in Fig. 8.2.



**Fig. 8.1** User interface

<sup>3</sup>The Italian fund-raiser association.

<sup>4</sup>Research center on non profit, fund-raising and social responsibility operative in the University of Bologna.

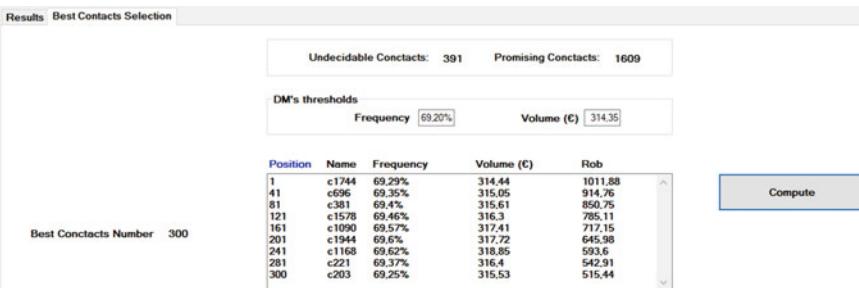


**Fig. 8.2** Rankings of the estimated gift frequencies and volumes, with the corresponding robustness index, for the promising Contacts. Number of undecidable Contacts

For both the frequency and the volume, a ranking of promising Contacts is showed. Moreover, the number of undecidable Contacts is exhibited. Note that the DM can choose how deep to explore the rankings.

In Fig. 8.3, the result of a composite selection query (ordered by *Rob*) is displayed, where those Contacts are selected, that exceed both the two prefixed thresholds by the DM, one for each estimated characteristic (frequency and volume). For the sake of brevity, these Contacts are called “best Contacts.”

Notice that the DM can set the thresholds in order to obtain a prefixed number of Contacts to reach (a unit cost is involved, see Introduction), in this case 300, displayed in Figure with a step of 40, in order to satisfy the budget constraint of the considered fund-raising campaign.



**Fig. 8.3** Best contacts

## 8.4 Conclusion

In this paper, we developed an effective recommender system for fund-raising management, which extends into an actual context a previous approach, by using a large structured DB. This extension involves a significant improvement of the basic method, by considering both a proper similarity measure and the entire available information (including the volume), in the context of a nonparametric approach. The numerical study, realized in collaboration with two Italian operative research centers on fund-raising, gives evidence of the effectiveness of the proposed approach.

## References

1. Aman, K., Singh, M.D.: A review of data classification using K-nearest neighbor algorithm. *Int. J. Emerg. Technol. Adv. Eng. Website* **3**(6), 354–360 (2013)
2. Andreoni, J.: Philanthropy. In: Kolm, S.C., Ythier, J. (ed.) *Handbook the Economics of Giving, Altruism and Reciprocity* 2, pp. 1201–1269. Elsevier, Amsterdam (2006)
3. Barzanti, L., Gaspari, M., Saletti, D.: Modelling decision making in fund raising management by a fuzzy knowledge system. *Expert Syst. Appl.* **36**, 9466–9478 (2009)
4. Barzanti, L., Giove, S.: A decision support system for fund raising management based on the Choquet integral methodology. *Expert Syst.* **29**(4), 359–373 (2012)
5. Barzanti, L., Giove, S., Pezzi, A.: A Recommender system for fund raising management (submitted)
6. Barzanti, L., Mastroleo, M.: An enhanced approach for developing an expert system for fund raising management. In: Segura, J.M., Reiter, A.C. (ed.) *Expert System Software: Engineering, Advantages and Applications*, pp. 131–156. NYC Nova Science Publishers (2013)
7. Beg, I., Ashraf, S.: Similarity measures for fuzzy sets. *Appl. Comput. Math.* **2**(8), 192–202 (2009)
8. Bhatia, N., Vandana, A.: Survey of nearest neighbor techniques. *Int. J. Comput. Sci. Inform. Secur.* **8**(2), 302–305 (2010)
9. Canestrelli, E., Canestrelli, P., Corazza, M., Filippone, M., Giove, S., Masulli, F.: Local learning of tide level time series using a fuzzy approach. In: *Proceedings of International Joint Conference on Neural Networks*, Orlando, Florida, USA, 12–17 Aug 2007
10. Cappellari, L., Ghinetti, P., Turati, G.: On time and money donations. *J. Socio-Economics* **40**(6), 853–867 (2011)
11. Couso, I., Garrido, L., Sánchez, L.: Similarity and dissimilarity measures between fuzzy sets: a formal relational study. *Inf. Sci.* **229**, 122–141 (2013)
12. Duffy, J., Ochs, J., Vesterlund, L.: Giving little by little: dynamic voluntary contribution game. *J. Public Econ.* **91**(9), 1708–1730 (2007)
13. Duncan, B.: Modeling charitable contributions of time and money. *J. Public Econ.* **72**, 213–242 (1999)
14. Fan, J., Gijbels, I.: *Local Polynomial Modelling and Its Applications*. Chapman & Hall, London (1996)
15. Flory P.: Fundraising databases. DSC London (2001)
16. Flory P.: Building a fundraising database using your PC. DSC London (2001)
17. Giove, S., Pellizzari, P.: Time series filtering and reconstruction using fuzzy weighted local regression. In: *Soft Computing in Financial Engineering*, pp. 73–92. Physica-Verlag, Heidelberg (1999)
18. Janach D., Zanker M., Felfernig A., Friedric G.: *Recommender Systems: An Introduction*. Cambridge University Press, Cambridge (2001)

19. Keller, J.M., Gray, M.R., Givens Jr., J.A.: A fuzzy K-Nearest neighbor algorithm. *IEEE Trans. Syst. Man Cybern. Smc-15*(4), 580–585 (1985)
20. Kercheville, J., Kercheville, J.: The effective use of technology in nonprofits. In Tempel, E. (ed.) Hank Rosso's Achieving Excellence in Fund Raising, pp. 366–379. Wiley, New York (2003)
21. Lee, L., Piliavin, J.A., Call, V.R.: Giving time, blood and money: similarities and differences. *Soc. Psychol. Quart.* **62**(3), 276–290 (1999)
22. Melandri, V.: Fundraising. Civil Sector Press, Toronto (2017)
23. Moro, S., Cortez, P., Rita, P.: A divide-and conquer strategy using feature relevance and expert knowledge for enhancing a data mining approach to bank telemarketing. *Expert Syst.* 1–13 (2017). <https://doi.org/10.1111/exsy.12253>
24. Nudd, S.P.: Thinking strategically about information. In Tempel, E. (ed.) Hank Rosso's Achieving Excellence in Fund Raising, pp. 349–365. Wiley, New York (2003)
25. Rosso, H., Tempel, R., Melandri, V.: The Fund Raising Book. ETAS, Bologna (2004) (in Italian)
26. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach, 2nd edn. Prentice Hall, New York (2003)
27. Sargeant, A.: Using Donor Lifetime Value to Inform Fundraising Strategy. *Nonprofit Manage. Leadersh.* **12**(1), 25–38 (2001)
28. Verhaert, G.A., Van den Poel, D.: The role of seed money and threshold size in optimizing fundraising campaigns: Past behavior matters! *Expert Syst. Appl.* **39**, 13075–13084 (2012)
29. Wand, M.P., Jones, M.C.: Kernel Smoothing. Chapman & Hall, London (1995)

# Chapter 9

## Reconstruction, Optimization and Quality Check of Microsoft HoloLens-Acquired 3D Point Clouds



Gianpaolo Francesco Trotta, Sergio Mazzola, Giuseppe Gelardi,  
Antonio Brunetti, Nicola Marino and Vitoantonio Bevilacqua

**Abstract** In the context of three-dimensional acquisition and elaboration, it is essential to maintain a balanced approach between model accuracy and required resources. As a possible solution to this problem, the present paper proposes a method to obtain accurate and lightweight meshes of a real environment using the Microsoft® HoloLens™ as a device for point clouds acquisition. Firstly, we describe an empirical procedure to improve 3D models, with the use of optimal parameters found by means of a genetic algorithm. Then, a systematic review of the indexes for evaluating the quality of meshes is developed, in order to quantify and compare the goodness of the obtained outputs. Finally, in order to check the quality of the proposed approach, a reconstructed model is tested in a virtual scenario implemented in Unity® 3D Game Engine.

### 9.1 Introduction

The need for an accurate collection of three-dimensional (3D) data for generating 3D models goes opposite to the time and computational cost needed for obtaining and generating the desired model. This is even more emphasized if mobile and real-time contexts are concerned: For this reason, this issue is often reduced to the search of the best trade-off between the accuracy of the generated 3D model and its computational cost, in terms of required hardware, time and computational resources. Nowadays, more and more affordable and powerful 3D scanning devices are coming on the market, allowing to obtain high-resolution spatial data of real-world objects and environments, making them available to the community of 3D developers.

---

G. F. Trotta

Department of Mechanics, Mathematics and Management Engineering,  
Polytechnic University of Bari, Bari, Italy

S. Mazzola · G. Gelardi · A. Brunetti · N. Marino · V. Bevilacqua (✉)

Department of Electrical and Information Engineering,  
Polytechnic University of Bari, Bari, Italy  
e-mail: [vitoantonio.bevilacqua@poliba.it](mailto:vitoantonio.bevilacqua@poliba.it)

Several research fields require efficient elaboration of raw 3D data acquired with different devices; in fact, previous works have been done using optimized 3D models in different contexts, such as medical [1], cultural heritage [2, 3] and industry or education training [4, 5].

Currently, several techniques for extracting meshes from acquisitions of real environments have been developed; in this work, we are going to consider the acquisitions performed using the sensors integrated into the Microsoft® HoloLens™. In particular, we will demonstrate how much a rough mesh acquired with the considered device may be optimized, to make it usable in more challenging contexts, such as mobile or virtual reality applications.

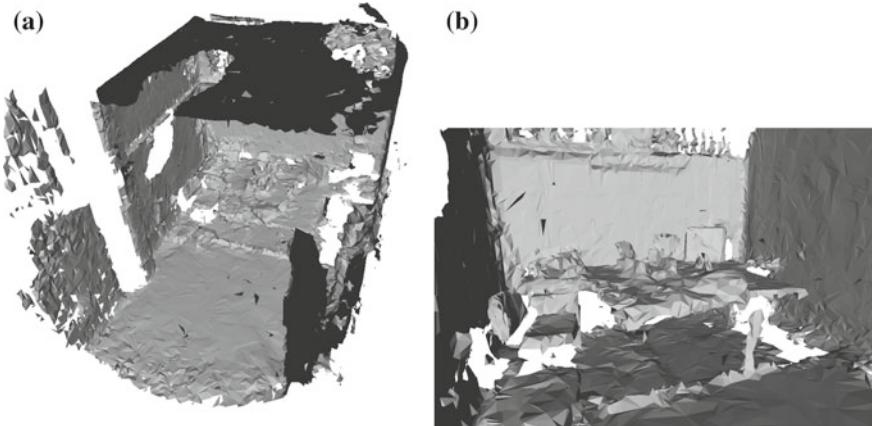
A relevant contribution for improving the optimization of meshes has been made by the genetic approach employed to find an optimal set of parameters for the surfaces subdivision algorithm, a fundamental step of the mesh reconstruction procedure. In fact, genetic algorithms, or evolutionary approaches in general, demonstrated to be reliable in finding the optimal solution to specific problems, having a suitable design of the fitness function and genetic operators [6, 7].

In [2], the authors propose a 3D point cloud reconstruction method based on an optimization procedure using a multi-objective genetic algorithm (MOGA) to improve the mesh obtained by low-cost acquisition devices. In fact, a photogrammetric technique optimized with the genetic approach has been used to build virtual scenarios, thus lowering the costs for the acquisition of the environment and the computational complexity required, in the scope of cultural heritage safekeeping and preservation. Moreover, in order to perform the comparison between an object acquired with the Artec Eva™ 3D scanner and with a typical smartphone camera, reconstructed using a photogrammetric procedure optimized with the evolutionary procedure, the Hausdorff distance [8] between the two objects has been computed, considering the model obtained from the 3D scanner as the ground truth.

The rest of the paper is organized as follows: In Sect. 9.2, an empirical procedure which combines well-known techniques to improve 3D models is presented, exploiting the parameters obtained using the genetic algorithm; then, a review of indexes for evaluating the quality of meshes is performed; Sect. 9.3 describes some applications of optimized reconstruction of 3D model. Furthermore, a quality check is performed considering a test mesh, in terms of both quality indexes and practical use in Unity® 3D Game Engine. Finally, results and conclusions are pointed out in Sect. 9.4.

## 9.2 Materials

In order to perform point clouds acquisition, several devices could be found in the market. In this work, we investigate the Microsoft® HoloLens™ as a valid option to perform this task. Despite its cost is prohibitive for the general market, it is very likely that, in the near future, this device, together with similar products of the same category, will become available to everyone.



**Fig. 9.1** An example of 3D acquisition performed using HoloLens<sup>TM</sup>. **a** 3D model of a room acquired with the device; **b** a detailed view of the models, containing the interiors of the room

Microsoft® HoloLens<sup>TM</sup> is provided with a depth camera which is able to sense the three-dimensional information of the environment around the user wearing it. Although the positive aspects make it a landmark into the Mixed Reality (MR) community, i.e., the complete portability, it suffers from multiple limitations. Specifically, the quality of the real-time data coming from the depth sensor is very low [9]. An example of HoloLens<sup>TM</sup> acquisition is represented in Fig. 9.1. The displayed mesh is made up of 38,170 triangular faces, the model has a lot of holes, the shape of objects and walls is not precise, and, in general, it doesn't have a good visual effect, thus making the acquisition not usable.

### 9.2.1 Mesh Reconstruction

In order to improve the acquisition performed by using the HoloLens<sup>TM</sup> depth sensor, there are several techniques.

Thanks to the evolutionary approach described in [2], the obtained mesh was filtered using a surface subdivision algorithm, configured with the optimized parameters calculated by the genetic algorithm. The filtering procedure makes the point cloud much denser than the initial condition; thus, after the reconstruction, an accurate decimation of the number of points is needed. The software used for this process is MeshLab, an open source system for processing and editing 3D triangular meshes [10].

**Cleaning filters** Since the depth sensor of HoloLens<sup>TM</sup> is quite noisy, isolated artefacts are generated during the scan; therefore, it is necessary to remove all the isolated pieces which are smaller than a set dimension. Thanks to MeshLab, the removal

of isolated pieces is performed using the procedure *Remove isolated pieces (wrt Diameter)*, specifying the required threshold. Regarding the example reported in Fig. 9.1, a diameter threshold of 6% of the diameter of the biggest isolated object in the model was set. Moreover, the reconstruction filter requires the application of other cleaning filters, which are: *Remove faces from non-manifold edges*, *Remove unreferenced vertices* and *Remove zero area faces*.

In fact, the application of the surfaces subdivision algorithm requires that the mesh input to the algorithm has to be “manifold” [11]; to make it meet the requirements, non-manifold edges can be easily identified and removed with the filter mentioned above. The other two filters are required by the surface reconstruction algorithm, described in the following paragraphs, which uses the normal of each face for its calculations [12]. For this reason, faces with a null normal make it impossible to use the filter and need to be removed; these include unreferenced vertices and zero area faces.

**Surface subdivision** After cleaning the mesh, the subdivision algorithm can be applied. Referring to the results reached in [2], which are synthesized in Table 9.1, we have used the algorithm which performed the best in terms of minimization of Hausdorff distance [8, 13], the Butterfly algorithm [14], with 4 iterations and a threshold of 0.125%. It can be applied to the input mesh through the MeshLab filter *Subdivision Surfaces: Butterfly Subdivision*.

As expected, the application of Butterfly algorithm to the input mesh dramatically increased its faces number. In fact, for the considered object represented in Fig. 9.1, it raised from 38,170 (and 23,615 vertices) to 4,484,307 (and 2,267,738), that is, an increase of the 11,648.25%. Since the obtained mesh is too heavy to be elaborated quickly by any software for 3D rendering, the 3D model is still not usable [15]. In the following sections, we are going to address this problem.

**Smoothing and holes closure** In order to achieve a better final result in terms of visual effect, other minor fixes can be applied. For the considered example, we have applied the *HC Laplacian Smoothing* filter, based on [16]. Moreover, due to the noisy sensor, several small holes are disseminated throughout the model; they can be closed with the filter *Close holes*, opportunely tuning the parameter in order to avoid damaging filtering actions.

**Surface reconstruction** Finally, the mesh filtered, as described in the previous steps, can be reconstructed by means of a surface reconstruction algorithm. In this work, the Poisson algorithm, described in [12], has been applied; in MeshLab, it can be

**Table 9.1** Best individual from the MOGA application obtained in [2]

Algorithm	Vertices	Iterations	Threshold	Hausdorff distance	
				Mean	RMS
Butterfly [14]	75,137 (+441.281%)	3	0.5%	3.778014 (-0.015%)	5.043389 (+0.004%)

found under the *Screened Poisson Surface Reconstruction* procedure. Considering the example reported in this work, the reconstruction algorithm has been applied increasing the *Minimum number of samples* parameter to 10, due to the noise of the mesh. Moreover, in some cases, it could be advisable to clean one more time the mesh with the previous filters.

As a result of the filtering and reconstruction procedure, it could be observed that the number of faces of the considered mesh was decreased, from 4,484,307 of faces to 440,947 triangles, which need to be further reduced for this object, thus obtaining a better trade-off between quality and number of faces.

### 9.2.2 *Quality Indexes*

To carry out a valid improvement of an input mesh, the number of faces must not be reduced indiscriminately, but a measure of quality using acceptable indexes appropriately measuring the trade-off between rendering quality and computational cost need to be adopted. These indexes will also demonstrate how the reconstructed and optimized mesh is better than the input one.

As reported in [17], quality is relative, and the solution of a quality check, by definition, is approximative and depends on the case under examination. However, in order to establish an objective criterion for quality evaluation is to consider a mesh composed only of ideal faces as gold standard. For example, in the case of triangles, the ideal polygonal face is the equilateral triangle, as stated in [18, 19]. In particular, the more the faces of a considered mesh are similar to an equilateral triangle, the higher is the quality of the mesh itself; therefore, the quality parameters have to measure how far a given element deviates from the ideal shape. According to this, five quality indexes were selected from [17]: aspect ratio, skewness, maximum angle, minimum angle and maximum size. It is noteworthy that all the metrics refers to individual faces of the mesh. Thanks to empirical evaluations based on the considered object, appropriate thresholds for every index were identified and the percentage of faces which fitted within all the defined constraints was used as quality measure. The computation of the indexes was performed using HyperMesh®, a toolbox from the Altair® HyperWorks® suite.

Considering the input mesh, the *Aspect Ratio* critical value was set to 5, meaning that every face with an aspect ratio smaller than 5 is considered acceptable. Regarding *Skewness*, a skew smaller than 40° is considered acceptable. The acceptable *Included Angles* were those included between 30° and 120°, meaning that, for every face of the mesh, all its three angles are checked: If any angle is smaller than 30° or bigger than 120°, the face is labeled as a fail. Finally, a *Maximum Size* threshold of 0.5 was set.

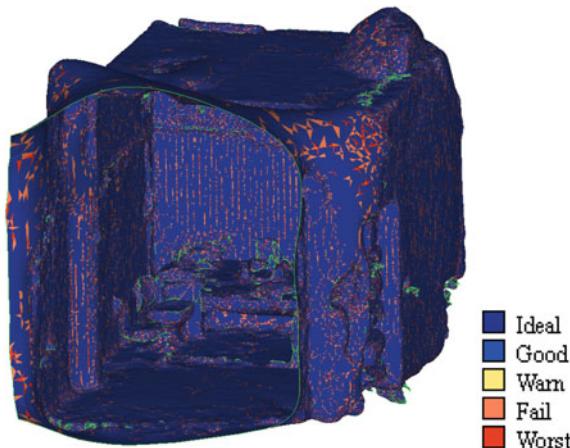
### 9.3 Methods

#### 9.3.1 Performances Optimization

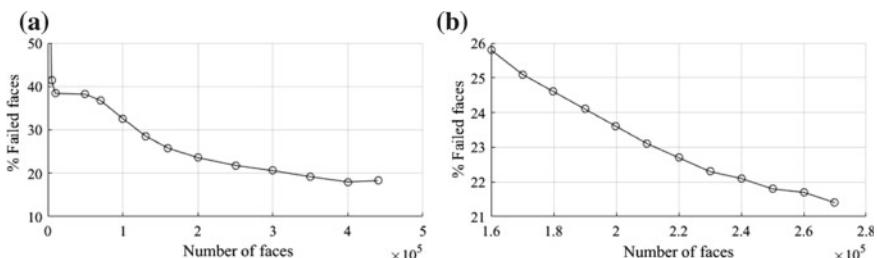
Based on the quality indexes reported in the previous section, the result of the quality analysis performed on the reconstructed mesh is represented in Fig. 9.2.

As previously said, the considered model is composed of 440947 faces. According to HyperMesh®, the 18.3% of its faces (called *failed faces*) violates the imposed thresholds. Using the *Simplification: Quadric Edge Collapse Decimation* filter in MeshLab [20], the number of the faces was gradually reduced, keeping track of the percentage of failed faces. In order to avoid high distortion of the details, the parameter *Preserve normal* has been activated. The result is a plot of the percentage of failed faces as a function of the number of faces, shown in Fig. 9.3a.

Observing the plot, it is clear that a convenient trade-off would be around 200,000 faces: A deeper analysis (Fig. 9.3b) shows that between 240,000 and 230,000 faces,



**Fig. 9.2** Room mesh quality map



**Fig. 9.3** In a plot of the quality check during faces decimation; in b a zoom on the same plot around the 200,000  $x$ -coordinate

for a great decrease of the number of faces, the quality decreases slowly. After 230,000, the loss of quality seems to become linear and its drop speeds up. For this reason, it is convenient to choose the value of 230,000 faces for our final mesh. The corresponding percentage of failed faces is 22.3%, that is, with just 4% points more respect to the original interpolated mesh, we have almost halved the number of faces.

As a final result, we have assumed that a reduction in faces of a factor of 0.52 is optimal and has a general validity, as can be demonstrated through further tests with other HoloLens™ acquisitions. This parameter can be set directly in MeshLab when applying the quadric decimation filter, under the name *Percentage reduction*, or included in an automated script which performs both the reconstruction (as described in Sect. 9.2.1) and the optimization.

### 9.3.2 *Quality Check*

After the process of reconstruction and optimization, a comparison between the final result and the rough model is reported. This comparison was carried out in two ways: firstly, using the HyperMesh® quality check tool, which checks the quality in a theoretical way measuring the deviation from the ideal shapes, and then, using Unity® 3D, whose analysis is more practical and related to the real use of the mesh in a 3D rendering software.

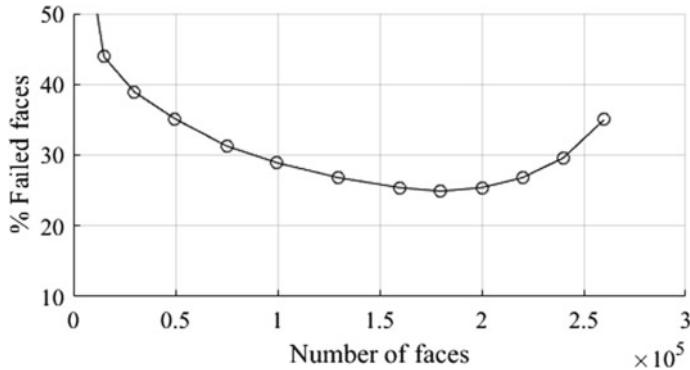
The comparison is carried out considering another HoloLens™ 3D acquisition (a room bigger than the one reported in Fig. 9.1), using the previously explained method of reconstruction and optimization.

Considering Fig. 9.4, it is possible to see that the found factor 0.52 is optimal for this mesh too. The number of faces of the reconstructed (and non-optimized) mesh is indeed 266,806, which multiplied by 0.52 equals about 138,740. As expected, it corresponds with the exact point where the faces number reaches the minimum value before the percentage of failed faces starts increasing. With these premises, a quality check of the mesh is performed.

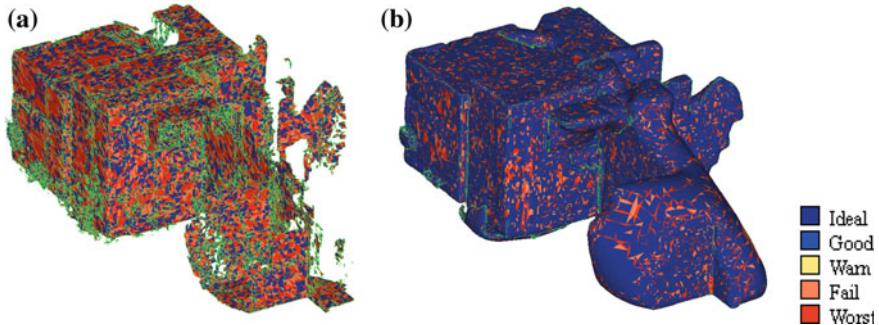
**HyperMesh® Quality Check** HyperMesh® quality check is a measure of how much the faces of a mesh deviate from their ideal shape. For this reason, this check is to be considered a theoretical measure of the goodness of the mesh, founded on objective criteria.

In Fig. 9.5 is shown a comparison between the quality maps of the rough and the optimized mesh (138,738 faces): A significant increase in the quality occurred. More specifically, the quality indexes of individual faces improved from the rough model to the optimized one, accumulating around the same value near the ideal one.

**Unity® 3D Quality Check** Unity® 3D, a game engine used for creating three-dimensional and two-dimensional games as well as simulations for several platforms, can help in obtaining a more practical-focused quality check. In fact, accurate and lightweight models of the physical spaces are crucial for creating interactive virtual environments. To test the optimized mesh following this approach, the navi-



**Fig. 9.4** Plot of the quality check during faces decimation

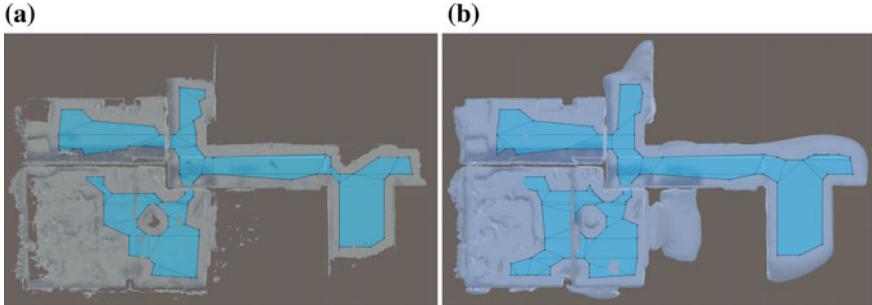


**Fig. 9.5** **a** Rough mesh, with the 62.7% of failed faces; **b** optimized mesh, with the 26.3% of failed faces

gation system functionality provided by the game engine was used. Specifically, the navigation system empowers a game object to move around the scene intelligently, following efficiently real-time calculated paths. To do so, the system computes a so-called NavMesh of the scene [21], which is a layer that can be stepped on by a set character, called NavMesh Agent.

To estimate the quality improvement of the refined mesh, we have referred to the surface covered by the NavMesh, the average length of random paths and the percentage of failed computations of paths between two random points. They are to be considered valid since they provide a good way to approximate all the potential movements of an object inside the model and show which surfaces the NavMesh Agent can interact with. The dimension of the NavMesh Agent is important in this analysis: In fact, it affects the NavMesh, defining the zones that can be stepped on and the spaces which are too narrow to let it in. For the purpose of this work, an average-sized Agent has been chosen, representing a character walking around our room. The resulting NavMeshe are shown in Fig. 9.6.

The difference between the two areas can be calculated: The NavMesh of the refined mesh covers 19% more surface with respect to the NavMesh of the rough



**Fig. 9.6** **a** NavMesh for the rough model; **b** NavMesh for the optimized model

**Table 9.2** Results for the rough model and for the refined one with 200,000 iterations

	Calculated paths	Failed paths	Average paths length	Failed paths percentage (%)
Rough mesh	59,380	140,620	3.808226	70.31
Fine mesh	80,308	119692	3.889201	59.846

one. Moreover, considering random couples of points placed in both the meshes at the floor level, Table 9.2 reports the results obtained for the metrics previously introduced.

In details, considering a mesh which has no possibility of failing the path computation between two points, a lower average paths length would mean a higher quality of the model. Indeed, in the higher-quality mesh, there would be fewer obstacles related to noise, and new shortcuts between the same points may open. However, the considered rough mesh has a high percentage of failed paths (70.3%), and it decreases (by about 10.5%) for the refined model: This means that several points were not connected in the rough mesh, and their number decreased with the reconstruction of the model. From this point of view, the fact that the average length of the paths remained fairly constant means that the refined mesh is smoother and cleaner, thus allows shorter paths between random points; moreover, as confirmed by the decrease in the failed paths percentage, more points of the floor are connected, even the most distant ones. Both phenomena contribute to modify the average length of the paths in such a way that they balance themselves, resulting in an almost unchanged average length.

## 9.4 Results and Conclusion

In this work, an empirical procedure to improve 3D models, with the use of optimal parameters, found by means of a genetic algorithm, and processing filters have been introduced. Then, a review of the indexes for evaluating the quality of meshes has been

**Table 9.3** Results of the first and second model (.stl files)

	Rough mesh	Reconstructed mesh	Optimized mesh
First model	38,170 faces—1864 KB	439,647 faces—21,468 KB	228,616 faces—11,163 KB
Second model	132,046 faces—6448 KB	267,836 faces—13,079 KB	139,272 faces—6801 KB

performed, and several acquisitions of real environments have been reconstructed and evaluated based on the considered indexes.

The reported results demonstrate how much a rough mesh acquired by the means of the Microsoft® HoloLens™ can be improved in terms of accuracy and computational cost, to make it usable in different contexts, ranging from medicine to cultural heritage preservation.

Furthermore, in addition to the results shown in Sect. 9.3.2 obtained with the use of HyperMesh® and Unity® 3D, it is possible to use the weight of the models, even in terms of storage size, to understand how much the original mesh has been lightened to make it more interactive in virtual scenarios of real applications. In fact, the results reported in Table 9.3 for both the considered models confirm the goodness of the optimization procedure.

The results are very promising: The optimized meshes are lightweight enough to be elaborated easily and a substantial improvement in quality occurred. Not only the reconstruction increases the visible quality of the models, but also it contributes to augment the potential interactivity of the acquired environment.

## References

1. de Tommaso, M., Ricci, K., Delussi, M., Montemurno, A., Vecchio, E., Brunetti, A., Bevilacqua, V.: Testing a novel method for improving wayfinding by means of a p3b virtual reality visual paradigm in normal aging. Springerplus **5**(1), 1297 (2016)
2. Bevilacqua, V., Trotta, G.F., Brunetti, A., Buonamassa, G., Bruni, M., Delfine, G., Riezzo, M., Amadio, M., Bellantuono, G., Magaletti, D., Verrino, L., Guerriero, A.: Photogrammetric meshes and 3d points cloud reconstruction: a genetic algorithm optimization procedure. In: Rossi, F., Piotto, S., Concilio, S. (eds.) Advances in Artificial Life, Evolutionary Computation, and Systems Chemistry, pp. 65–76. Springer International Publishing, Berlin (2017)
3. Voulodimos, A., Doulamis, N., Fritsch, D., Makantasis, K., Doulamis, A., Klein, M.: Four-dimensional reconstruction of cultural heritage sites based on photogrammetry and clustering. J. Electron. Imaging **26**(1), 011013 (2016)
4. Zhang, H.: Head-mounted display-based intuitive virtual reality training system for the mining industry. Int. J. Mining Sci. Technol. **27**(4), 717–722 (2017)
5. Uva, A.E., Fiorentino, M., Gattullo, M., Colaprico, M., de Ruvo, M.F., Marino, F., Trotta, G.F., Manghisi, V.M., Boccaccio, A., Bevilacqua, V., Monno, G.: Design of a projective ar workbench for manual working stations. In: De Paolis, L.T., Mongelli, A. (eds.) Augmented Reality, Virtual Reality, and Computer Graphics, pp. 358–367. Springer International Publishing, Berlin (2016)

6. Bevilacqua, V., Mastronardi, G., Menolascina, F., Pannarale, P., Pedone, A.: A novel multi-objective genetic algorithm approach to artificial neural network topology optimisation: the breast cancer classification problem. In: The 2006 IEEE International Joint Conference on Neural Network Proceedings, pp. 1958–1965. IEEE (2006)
7. Bevilacqua, V., Ivona, F., Cafarchia, D., Marino, F.: An evolutionary optimization method for parameter search in 3d points cloud reconstruction. In: Huang, D.-S., Bevilacqua, V., Figueroa, J.C., Premaratne, P. (eds.) Intelligent Computing Theories, pp. 601–611. Springer, Berlin (2013)
8. Aspert, N., Santa-Cruz, D., Ebrahimi, T.: Mesh: Measuring errors between surfaces using the hausdorff distance. In: 2002 IEEE International Conference on Multimedia and Expo (ICME'02), vol. 1, pp. 705–708. IEEE (2002)
9. Garon, M., Boulet, P.-O., Doironz, J.-P., Beaulieu, L., Lalonde, J.-F.: Real-time high resolution 3d data on the hololens. In: 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), pp. 189–191. IEEE (2016)
10. Cignoni, P., Callieri, M., Corsini, M., Dellepiane, M., Ganovelli, F., Ranzuglia, G.: Meshlab: an open-source mesh processing tool. In: Eurographics Italian Chapter Conference, pp. 129–136 (2008)
11. Luebke, D.P.: A developer’s survey of polygonal simplification algorithms. *IEEE Comput. Graph. Appl.* **21**(3), 24–35 (2001)
12. Kazhdan, M., Hoppe, H.: Screened poisson surface reconstruction. *ACM Trans. Graph. (TOG)* **32**(3), 29 (2013)
13. Cignoni, P., Rocchini, C., Scopigno, R.: Metro: measuring error on simplified surfaces. In: Computer Graphics Forum, vol. 17, pp. 167–174. Blackwell Publishers (1998)
14. Dyn, N., Levine, D., Gregory, J.A.: A butterfly subdivision scheme for surface interpolation with tension control. *ACM Trans. Graph.* **9**(2), 160–169 (1990). April
15. Argelaguet, F., Andujar, C.: A survey of 3d object selection techniques for virtual environments. *Comput. Graph.* **37**(3), 121–136 (2013)
16. Vollmer, J., Mencl, R., Muller, H.: Improved laplacian smoothing of noisy surface meshes. *Comput. Graph. Forum* (1999)
17. OptiStruct User’s Guide Hyperworks. 13.0. User Manual, Altair Engineering Inc., Troy MI (2014)
18. Paul Chew, L.: Guaranteed-quality triangular meshes. Technical report, Cornell University (1989)
19. Heckbert, P.S., Garland, M.: Survey of polygonal surface simplification algorithms. Technical report, Carnegie-Mellon Univ Pittsburgh PA School of Computer Science (1997)
20. Tarini, M., Pietroni, N., Cignoni, P., Panozzo, D., Puppo, E.: Practical quad mesh simplification. *Comput. Graphics Forum* **29**(2), 200 (Special Issue of Eurographics 2010 Conference)
21. Cui, X., Shi, H.: An overview of pathfinding in navigation mesh. *IJCSNS* **12**(12), 48–51 (2012)

# Chapter 10

# The “Probabilistic Rand Index”: A Look from Some Different Perspectives



Stefano Rovetta , Francesco Masulli and Alberto Cabri

**Abstract** One crucial tool in machine learning is a measure of partition similarity. This study focuses on the “probabilistic Rand index”, a variant of the Rand index. We look at this measure from different perspectives: probabilistic, information-theoretic, and diversity-theoretic. These give some insight, reveal relationships with other types of measures, and suggest some possible alternative interpretations.

## 10.1 Introduction

As is well known, the current power of computational architectures has enabled an exponentially growing number of big-data applications of machine learning, with artificial neural networks arguably being the most successful approach. At the same time, interest has grown in both unsupervised learning, which allows useful work to be performed without expensive ground-truth collection, and ensemble methods, which are scalable in that they use multiple, simpler learners.

One crucial tool in both these fields is a measure of *partition similarity*, where partitions may be either clustering of data or classification outputs.

Similarity measures between partitions is a well-studied problem. A recent extensive survey [1] cites 76 measures of similarity or dissimilarity developed over the last century. The same problem can be cast as measuring diversity among classifiers or clusterings, binary string similarity, and categorical feature similarity.

---

S. Rovetta ( ) · F. Masulli · A. Cabri  
DIBRIS, University of Genova, Via Dodecaneso 35, 16146 Genoa, Italy  
e-mail: [stefano.rovetta@unige.it](mailto:stefano.rovetta@unige.it)

F. Masulli  
e-mail: [francesco.masulli@unige.it](mailto:francesco.masulli@unige.it)

A. Cabri  
e-mail: [alberto.cabri@dbris.unige.it](mailto:alberto.cabri@dbris.unige.it)

F. Masulli  
Sbarro Institute for Cancer Research and Molecular Medicine,  
Temple University, Philadelphia, PA, USA

A more recent trend has been to incorporate more information than just the coincidence of binary/categorical attributes; this includes for instance the development of fuzzy variants [2, 3].

As a contribution to the study of this class of measures, this work focuses on the “probabilistic Rand index”, a variant of the well-known Rand index of partition similarity [4] that was proposed in [5]. We look at this measure from several perspectives and suggest some possible, alternative interpretations, which may prompt new and improved ways to look at partition similarity in machine learning.

## 10.2 Co-Association Statistics

Let  $\Pi^A = \{\pi_1^A, \pi_2^A, \dots, \pi_{m_A}^A\}$  and  $\Pi^B = \{\pi_1^B, \pi_2^B, \dots, \pi_{m_B}^B\}$  be two partitions of the same set of  $P$  data items  $X = \{x_1, \dots, x_P\} \subset \mathcal{X}$  into, respectively,  $m_A$  and  $m_B$  parts (clusters).

As in [6], we indicate that partition  $\Pi^A$  puts two data items  $x_l$  and  $x_m$  in the same cluster by writing the indicator function  $x_l \sim_A x_m$ . The barred symbol  $x_l \not\sim_A x_m$  expresses negation. In the following, we will simplify the notation by dropping the superscript  $A$  or  $B$  when stating facts that hold for both.

**Definition 1** (*Association model*). Given a partition  $\Pi$  of a set  $X$ , the *association model* under  $\Pi$  is the set of probabilities  $\Pr(x_i \in \pi_k | x_j \in \pi_k) = \Pr(x_j \in \pi_k | x_i \in \pi_k) = \Pr(x_i \sim x_j)$  for all pairs  $(x_i, x_j) \in X^2$ .

The association model gives the probability that any two objects are put in the same cluster by partition  $\Pi$ . When this is 0 or 1, we consider all possible  $N = \binom{P}{2} = P(P - 1)/2$  pairs of data items  $(x_l, x_m)$ ,  $l < m$ . Under partition  $\Pi$ , we build the association vector  $\mathbf{u}$  whose generic element  $u_k$ , with  $k = (m - 1)(m - 2)/2 + l$ , indicates whether  $x_l$  and  $x_m$  are in the same cluster of  $\Pi$  or not:

$$u_k = [x_l \sim x_m] \quad (10.1)$$

using the “Iverson bracket” notation for indicator functions.

**Definition 2** (*Co-association model*). Given two association models corresponding to two partitions  $\Pi^A, \Pi^B$  of a set  $X$ , the corresponding *co-association model* is the set of the probabilities  $p = \{p_0, p_1, p_2\}$  of the following three events:

- 0, negative agreement:  $x \not\sim_A y$  and  $x \not\sim_B y$
- 1, disagreement:  $(x \not\sim_A y \text{ and } x \sim_B y) \text{ or } (x \sim_A y \text{ and } x \not\sim_B y)$
- 2, positive agreement:  $x \sim_A y$  and  $x \sim_B y$

Several partition similarity measures, or cluster similarity indexes, are based on the comparison of two co-association vectors  $\mathbf{u}^A$  and  $\mathbf{u}^B$ ; many of them [7] use the contingency matrix

$$\begin{pmatrix} \mathbf{C}_{00} & \mathbf{C}_{01} \\ \mathbf{C}_{10} & \mathbf{C}_{11} \end{pmatrix} \quad (10.2)$$

defined by

$\mathbf{C}_{00}$  = number of items s.t.  $x_l \not\sim_A x_m$  and  $x_l \not\sim_B x_m$

$\mathbf{C}_{01}$  = number of items s.t.  $x_l \not\sim_A x_m$  and  $x_l \sim_B x_m$

$\mathbf{C}_{10}$  = number of items s.t.  $x_l \sim_A x_m$  and  $x_l \not\sim_B x_m$

$\mathbf{C}_{11}$  = number of items s.t.  $x_l \sim_A x_m$  and  $x_l \sim_B x_m$

In classification, one partition is the ground truth that acts as a reference, and the other is under test. As a consequence, disagreements are asymmetrical and may be weighted differently. However, to simplify the analysis here we implicitly assume the zero-one loss, so all similarity indices will be symmetrical with respect to the exchange of disagreements. In particular, throughout this work we will only deal with indices such that

$$\text{index}(\mathbf{C}) = \text{index}(\mathbf{C}^T) \quad (10.3)$$

where superscript T is transpose. In other words, we will take into account only symmetric contingency matrices of the form

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_0 & \mathbf{C}_1 \\ \mathbf{C}_1 & \mathbf{C}_2 \end{pmatrix}. \quad (10.4)$$

In terms of association vectors, the contingency matrix can be expressed as

$$\begin{aligned} \mathbf{C}_0 &= \sum_k [u_k^A = 0 \wedge u_k^B = 0] = \| (1 - \mathbf{u}^A) \wedge (1 - \mathbf{u}^B) \|_1 \\ \mathbf{C}_1 &= \sum_k [u_k^A \neq u_k^B] = \frac{1}{2} (\| (1 - \mathbf{u}^A) \wedge \mathbf{u}^B \|_1 + \| \mathbf{u}^A \wedge (1 - \mathbf{u}^B) \|_1) \\ \mathbf{C}_2 &= \sum_k [u_k^A = 1 \wedge u_k^B = 1] = \| \mathbf{u}^A \wedge \mathbf{u}^B \|_1 \end{aligned}$$

where  $\wedge$  indicates the “and” operation, i.e. the Hadamard product of the indicators, and  $\| \cdot \|_1$  the entry-wise 1-norm.

We will also refer to the normalized contingency matrix

$$\mathbf{F} \triangleq \frac{1}{N} \mathbf{C} = \begin{pmatrix} \mathbf{F}_0 & \mathbf{F}_1 \\ \mathbf{F}_1 & \mathbf{F}_2 \end{pmatrix} \quad (10.5)$$

which represents frequencies rather than counts and is, therefore, an estimate of a co-association model:

$$\mathbf{F} \approx \begin{pmatrix} p_0 & p_1 \\ p_1 & p_2 \end{pmatrix} \quad (10.6)$$

### 10.3 The Probabilistic Rand Index

The Rand index [4] is defined as

$$\text{RI} = \frac{\mathbf{C}_0 + \mathbf{C}_2}{\mathbf{C}_0 + 2\mathbf{C}_1 + \mathbf{C}_2} = \mathbf{F}_0 + \mathbf{F}_2 \quad (10.7)$$

The Rand index is known to have a higher sensitivity (lower false negative rate) than specificity (higher false positive rate). This is because the index does not incorporate a priori assumptions on any given null hypothesis, therefore it is not able to distinguish false negatives from true negatives. As a result, while the index is expected to output the value 1 for identical partitions, it will not necessarily output the value 0 for “maximally different” partitions, a concept that is in itself not well defined. To cope with this known issue, a modified version of the Rand index was proposed by Hubert and Arabie [8] incorporating a “correction for chance” which provides the ability to compare partition diversity with the null model, a generalized hypergeometric data assumption:

$$\text{ARI} = \frac{\mathbf{C}_0 \mathbf{C}_2 + \mathbf{C}_1^2}{2(\mathbf{C}_0 + \mathbf{C}_1)(\mathbf{C}_1 + \mathbf{C}_2)} \quad (10.8)$$

The adjusted Rand index is another popular choice for comparing partitions, more frequently adopted than the original RI. It reduces the specificity issue by comparing the Rand index with its expected value. This makes it a bipolar index, with values in  $[-1, 1]$  rather than  $[0, 1]$  as the other indexes considered here.

In [5], another avenue was chosen to cope with the specificity problem. A weighted version of the Rand index was defined by taking into account directly the a priori probability (*prior* to observing the data) of the events of interest  $h \in \{0, 1, 2\}$ , namely the probabilities  $p_0, p_1, p_2$  that a random pair of data items  $(x, y) \in \mathcal{X}^2$  are in negative agreement / disagreement / positive agreement, respectively.

These probabilities were computed under a set of maximum uncertainty (maximum entropy) clustering hypotheses:

- partitions are independent of each other;
- all clusters are equiprobable;
- no spatial constraint (e.g., metric data) is known;
- points are uniformly sampled from  $\mathcal{X}$ .

In this situation, the probability of assigning a point to a cluster under  $\Pi^A$  does not depend on its spatial location nor on its assignment under  $\Pi^B$ . The probability of an item being assigned to a given cluster is  $1/m$  and that of two items in the same cluster is therefore  $1/m^2$ ; averaged over all clusters, this gives

$$\Pr(x \sim y) \triangleq Z = \frac{1}{m} \sum_{i=1}^m \frac{1}{m^2} = \frac{1}{m}, \quad (10.9)$$

and therefore,

$$p_0 = (1 - Z^A)(1 - Z^B) = \frac{(m_A - 1)(m_B - 1)}{m_A m_B}; \quad (10.10)$$

$$p_1 = \frac{1}{2} ((1 - Z^A)Z^B + Z^A(1 - Z^B)) = \frac{m_A + m_B}{2m_A m_B} - 1; \quad (10.11)$$

$$p_2 = Z^A Z^B = \frac{1}{m_A m_B}. \quad (10.12)$$

Given the probability  $p_h$  of event  $h$ , a corresponding weight is defined:

$$w_h = -\log p_h. \quad (10.13)$$

The *probabilistic Rand index* is then:

$$\text{PRI} = \frac{w_0 \mathbf{C}_0 + w_2 \mathbf{C}_2}{w_0 \mathbf{C}_0 + 2w_1 \mathbf{C}_1 + w_2 \mathbf{C}_2} = \frac{w_0 \mathbf{F}_0 + w_2 \mathbf{F}_2}{w_0 \mathbf{F}_0 + 2w_1 \mathbf{F}_1 + w_2 \mathbf{F}_2} \quad (10.14)$$

## 10.4 Analysis of the Probabilistic Rand Index

### 10.4.1 A Probabilistic Perspective

The Rand index has a straightforward probabilistic interpretation in terms of probabilities of co-association [8]. We can compute maximum likelihood a posteriori estimates (given the data) of the probability of each of the four events of interest by approximating them with the observed relative frequencies:

$$q_h \approx \mathbf{F}_h. \quad (10.15)$$

So, if frequencies  $\mathbf{F}_0, \mathbf{F}_1, \mathbf{F}_2$  are interpreted as the probabilities *after* observing the data  $q_0, q_1, q_2$ , the expressions of RI (10.7) and ARI (10.8) in terms of  $\mathbf{F}$  make it clear that these two versions of the Rand index are (estimates of) the probability of agreement between two partitions, after observing the data: the Rand index is an absolute estimate, while the Hubert and Arabie index is relative, being the normalized difference with respect to the expected value.

What is, then, the probabilistic Rand index? The original work [5] introduces weights as per Eq. (10.13) without clear justifications. But the following section will provide a possible explanation.

### 10.4.2 An Information-Theoretic Perspective

In Eq. (10.14), we can revert the weights  $w_h$  to their explicit formulation, Eq. (10.13):

$$\text{PRI} = \frac{q_0 \log p_0 + q_2 \log p_2}{q_0 \log p_0 + 2q_1 \log p_1 + q_2 \log p_2} \quad (10.16)$$

We can observe that the weight of event  $h$  is defined as the self-information [9] of  $h$  itself. So the PRI index does not deal with probabilities, but with measures of information. In this sense, we can more accurately state that the original Rand index RI and the Hubert–Arabie index ARI are probabilistic, while the “probabilistic” Carpineto–Romano index PRI index is actually information-theoretic.

The co-association model  $p$  is an a priori distribution which only depends on the partitions; in [5], it only takes into account model size, i.e. the number of clusters in the two clusterings. Then we also have  $\{q_h\}$  ( $q$  for short), the a posteriori co-association model which also depends on the observed data.

The PRI index is expressed in terms of cross-entropies  $-q_h \log p_h$ . The sum of these terms is the total cross-entropy of the a posteriori model  $q$  with respect to the a priori model  $p$ :

$$H_x(q|p) = - \sum_h q_h \log p_h. \quad (10.17)$$

The cross-entropy can be rewritten as follows:

$$-\sum_h q_h \log p_h = -\sum_h q_h \log q_h + \sum_h q_h \log \frac{q_h}{p_h} \quad (10.18)$$

or

$$H_x(q|p) = H(q) + D(q|p) \quad (10.19)$$

where

$$H(q) = - \sum_h q_h \log q_h \quad (10.20)$$

is the entropy of  $q$ , and

$$D(q|p) = \sum_h q_h \log \frac{q_h}{p_h} \quad (10.21)$$

is the Kullback–Leibler divergence of  $q$  with respect to  $p$  [10].

The cross-entropy expresses the amount of information needed to represent the co-association model  $p$  in terms of model  $q$ ; roughly speaking, the quantity of information that is necessary to specify  $p$  as a modification over  $p$ .

Therefore,  $H_x(q|p)$  measures how divergent is the observed co-association structure  $p$  with respect to  $q$ , the one expected in the absence of information. This is a

property of the partition, i.e. of the cluster representation and of the data: after having observed the data, the a priori guess is corrected, and  $H_x(q|p)$  is the amount of information needed for this correction.

Let us also consider the restriction of  $h$  to the agreements, namely  $h \in \mathcal{A}$ , obtaining the probability subsets  $p_{\mathcal{A}} = \{p_0, p_1\}$  and  $q_{\mathcal{A}} = \{q_0, q_1\}$ . The cross-entropy of this set of events is

$$H_x(q_{\mathcal{A}} | p_{\mathcal{A}}) = H(q_{\mathcal{A}}) - D(q_{\mathcal{A}} | p_{\mathcal{A}}) \quad (10.22)$$

The overall expression of PRI is therefore equivalent to

$$\text{PRI} = \frac{H_x(q_{\mathcal{A}} | p_{\mathcal{A}})}{H_x(q|p)} = \frac{H(q_{\mathcal{A}}) - D(q_{\mathcal{A}} | p_{\mathcal{A}})}{H(q) - D(q | p)} \quad (10.23)$$

and measures the contribution given by the agreements ( $h \in \mathcal{A}$ ) to the cross-entropy, expressed as a fraction of the total cross-entropy. When the agreements contribute to a high degree to the total cross-entropy, the index value approaches 1. However, this is counterbalanced by the entropy of the a priori model  $H(q)$ : if this is high, the index has a lower value.

Note that PRI, as originally defined, is based on equal a priori cluster assignment probabilities, where the probability of being assigned to any of the  $m$  clusters in  $\Pi$  is  $1/m$ . This is the maximum entropy case.

Summing up, PRI measures the fraction of the cross-entropy with respect to the a priori model explained by the observed agreements. In contrast, RI only measures the fraction of probability due to the observed agreements, without accounting for the a priori model. The improved index therefore measures the *significance* of the observed partition configuration with respect to the assumed model, reducing its sensitivity to agreements that are already explained by the a priori model. The counterbalancing effect of the model entropy is the main contribution of the index proposed in [5].

### 10.4.3 A Diversity-Theoretic Perspective

The concept of (bio)diversity arose naturally in ecology. In that context, diversity is a property of biological communities related to the number of different species and their respective populations.

Under the framework proposed by Hill [11] to introduce a unifying notation for the *diversity* of partitions, we define a diversity index in terms of the empirical (observed) relative frequency  $|\pi_i|/P$  of its elements in clusters  $\pi_i$ :

$$^rD = \sqrt[r-1]{\sum_{i=1}^m \frac{|\pi_i|}{P} \left(\frac{|\pi_i|}{P}\right)^{r-1}} = \sqrt[r-1]{\sum_{i=1}^m \left(\frac{|\pi_i|}{P}\right)^r} \approx \left(\sum_{i=1}^m \left(\frac{|\pi_i|}{P}\right)^r\right)^{1/(1-r)}. \quad (10.24)$$

This expression is the inverse generalized mean of the cluster relative cardinalities, with exponent  $r - 1$ , weighted by the proportions themselves. In the ecological domain many, frequently used indices can be cast under this form or simple transformations of it [12].

The value of  $r$  regulates the weight given to relative proportions, with  $r = 0$  disregarding it completely. If we define the two inverse diversities,  ${}^r z_A = \frac{1}{{}^0 D^A}$  and  ${}^r z_B = \frac{1}{{}^0 D^B}$ , and from them, we define the matrix

$${}^r Z = \begin{pmatrix} (1 - {}^r z_A)(1 - {}^r z_B) & {}^r z_A(1 - {}^r z_B) \\ (1 - {}^r z_A){}^r z_B & {}^r z_A {}^r z_B \end{pmatrix}, \quad (10.25)$$

then we can obtain weights parameterized by the diversity order  $r$ :

$${}^r W = -\log {}^r Z. \quad (10.26)$$

In particular, the Carpineto-Romano weight definition is obtained for  $r = 0$ , since the diversity of order 0 is

$${}^0 D = \sum_{i=1}^m \left( \frac{|\pi_i|}{P} \right)^0 = m. \quad (10.27)$$

This diversity index is termed “richness” in the ecological literature and, as noted, only takes into account the quantity of clusters.

By increasing  $r$ , proportions have an increasing effect on the overall index. In particular, two other interesting diversity indices correspond to  $r = 1$  and  $r = 2$ . For  $r = 1$ , the definition  $\lim_{r \rightarrow 1} {}^r D \triangleq {}^1 D$  is used, which leads to an entropy-based measure termed Shannon-Wiener index. For  $r = 2$ , we obtain the inverse Gini-Simpson index [13, 14].

## 10.5 Experimental Simulations

This section presents some simulated data to gain insight into the actual behaviour of the indexes discussed here. The probabilistic Rand index (PRI) is compared with both the Rand index (RI) and Hubert and Arabie’s adjusted Rand index (ARI) by controlled experiments on artificial data.

As per its definition, ARI takes values in the range  $[-1, 1]$ . However, for comparison purposes, here it is rescaled in the same  $[0, 1]$  range as the other two indexes.

**Table 10.1** Correlation between values of indexes on four different data sets

Kmeansdata			Iris						
		$\tau$			$\tau$				
		RI	ARI	PRI	RI				
$\rho$	RI	–	<b>1.000</b>	<b>1.000</b>	$\rho$	RI	–	<b>1.000</b>	<b>1.000</b>
	ARI	0.999	–	<b>1.000</b>		ARI	0.997	–	<b>1.000</b>
	PRI	0.991	0.990	–		PRI	0.997	0.989	–
New-thyroid			Wine						
		$\tau$			$\tau$				
		RI	ARI	PRI	RI				
$\rho$	RI	–	<b>1.000</b>	<b>1.000</b>	$\rho$	RI	–	<b>1.000</b>	<b>1.000</b>
	ARI	0.987	–	<b>1.000</b>		ARI	0.998	–	<b>1.000</b>
	PRI	0.999	0.981	–		PRI	0.999	0.993	–

Bold and italics have been used to graphically distinguish the upper triangular part of tables from the lower triangular part, which have different meanings. Upper triangle contains values for index  $\tau$ , lower triangle contains values for index  $\rho$ .

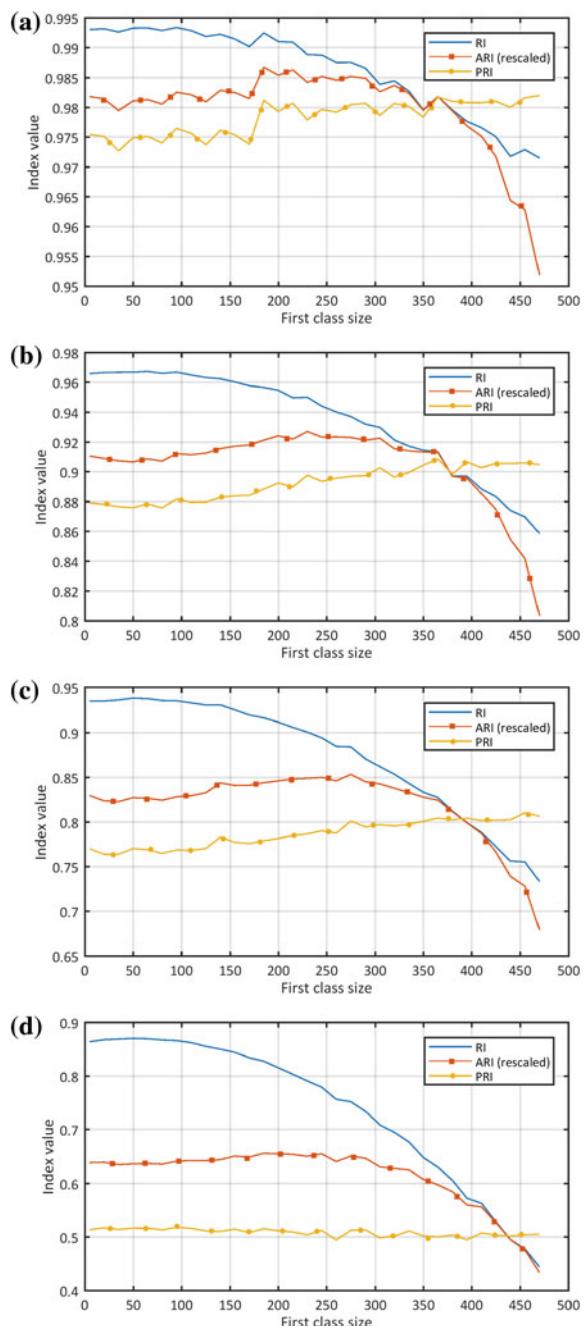
Table 10.1 constitutes a trend check. Correlation between indexes is computed for several runs of a clustering algorithm (a fuzzy  $c$ -means variant) on four different data sets, one synthetic composed of four well-separated, Gaussian 2D clusters (“Kmeansdata”), and three from the UCI repository [15], namely “Iris”, “New-thyroid”, “Wine”. Correlation between indexes is evaluated by means of Pearson’s correlation coefficient  $\rho$  and Kendall’s rank correlation coefficient  $\tau$ .

While rank correlation  $\tau$  is perfect in all cases, value correlation  $\rho$  is overall high as expected, but in general that of PRI with RI tends to be higher than that with ARI. However, this is not observed in all experiments.

In Fig. 10.1, indexes are compared on partitions with different numbers of parts. Partition 1 has 3 parts and partition 2 has 10 parts. The data set contains 515 points. Experiments are done for four numbers of mismatches: 10 (equivalent to an accuracy of 98%), 50 (90%), 100 (80%), and 250 (48%). In each experiment, several data sets are generated with one class sweeping from 5 to 470 points ( $x$  axis in the graphs). The remaining points are assigned with equal proportions to the remaining classes. Each experiment is run 10 times, and average results are presented.

It is interesting to see how PRI is the only index whose value reflects the accuracy irrespective of the imbalance in the data set. Apart from statistical oscillations, its value is very close to the accuracy in each of the four cases and does not vary with class proportions.

**Fig. 10.1** Comparison on indexes for partitions with different numbers of parts. Partition 1 has 3 parts and partition 2 has 10 parts. The number of mismatches is:  
**a** 10, **b** 50, **c** 100, **d** 250



## 10.6 Conclusion

In this paper, we have analysed the probabilistic Rand index to gain some insight into its properties. An information-theoretic perspective shows that the index is the fraction of cross-entropy due to agreements over the total cross-entropy, while a diversity-theoretic analysis suggests that the index is a member of a wider family whose sensitivity to expected proportions between parts is regulated by a single parameter  $r$ .

Future work will elaborate on the latter observation and will provide a flexible framework to be applied in the contexts of clustering and classification.

## References

1. Choi, S.S., Cha, S.H., Tappert, C.C.: A survey of binary similarity and distance measures. *J. Syst. Cybern. Inform.* **8**, 43–48 (2010)
2. Rovetta, S., Masulli, F.: An experimental validation of some indexes of fuzzy clustering similarity. In: Di Gesù, V., Pal, S.K., Petrosino, A. (eds.) 8th International Workshop on Fuzzy Logic and Applications (WILF 2009), Palermo, Italy, 9–12 June 2009. Lecture Notes in Computer Science, vol. 5571, pp. 132–139. Springer, Berlin (2009)
3. Campello, R.J.G.B.: Generalized external indexes for comparing data partitions with overlapping categories. *Pattern Recogn. Lett.* **31**, 966–975 (2010). <https://doi.org/10.1016/j.patrec.2010.01.002>
4. Rand, W.: Objective criteria for the evaluation of clustering methods. *J. Am. Statistical Assoc.* **66**, 846–850 (1971)
5. Carpineto, C., Romano, G.: Consensus clustering based on a new probabilistic rand index with application to subtopic retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(12), 2315–2326 (2012)
6. Rovetta, S., Masulli, F.: Visual stability analysis for model selection in graded possibilistic clustering. *Inform. Sci.* **279**, 37–51 (2014). <https://doi.org/10.1016/j.ins.2014.01.031>
7. Anderson, D.T., Bezdek, J.C., Popescu, M., Keller, J.M.: Comparing fuzzy, probabilistic, and possibilistic partitions. *IEEE Trans. Fuzzy Syst.* **18**(5), 906–918 (2010). <https://doi.org/10.1109/TFUZZ.2010.2052258>
8. Hubert, L., Arabie, P.: Comparing partitions. *J. Classif.* 193–218 (1985)
9. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley Series in Telecommunications and Signal Processing. Wiley, New Jersey, USA (2006)
10. Kullback, S.: Information Theory and Statistics. Wiley Publication in Mathematical Statistics, New York (1959)
11. Hill, M.O.: Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**(2), 427–432 (1973)
12. Jost, L.: Entropy and diversity. *Oikos* **113**(2), 363–375 (2006). <https://doi.org/10.1111/j.2006.0030-1299.14714.x>
13. Gini, C.: Variabilità e mutabilità: Contributo allo studio delle distribuzioni e delle relazioni statistiche. Tipografia di P. Cuppini, Bologna (1912)
14. Simpson, E.H.: Measurement of diversity. *Nature* **163**(4148), 688 (1949)
15. Asuncion, A., Newman, D.J.: UCI machine learning repository (2007)

# Chapter 11

## Dimension Reduction Techniques in a Brain–Computer Interface Application



Federico Cozza, Paola Galdi, Angela Serra, Gabriele Pasqua,  
Luigi Pavone and Roberto Tagliaferri

**Abstract** Electroencephalography (EEG)-based Brain–computer interface (BCI) technology allows a user to control an external device without muscle intervention through recorded neural activity. Ongoing research on BCI systems includes applications in the medical field to assist subjects with impaired motor functionality (e.g., for the control of prosthetic devices). In this context, the accuracy and efficiency of a BCI system are of paramount importance. Comparing four different dimension reduction techniques in combination with linear and nonlinear classifiers, we show that integrating these methods in a BCI system results in a reduced model complexity without affecting overall accuracy.

---

F. Cozza · P. Galdi · A. Serra · R. Tagliaferri (✉)  
NeuRoNeLab, Department of Management and Innovation Systems,  
University of Salerno, 84084 Fisciano, SA, Italy  
e-mail: [rotag@unisa.it](mailto:rotag@unisa.it)

F. Cozza  
e-mail: [federicocozza@me.com](mailto:federicocozza@me.com)

G. Pasqua · L. Pavone  
Neuromed, Mediterranean Neurological Institute, 86077 Pozzilli, IS, Italy  
e-mail: [bioingegneria@neuromed.it](mailto:bioingegneria@neuromed.it)  
URL: <http://www.neuromed.it/>

L. Pavone  
e-mail: [bioingegneria@neuromed.it](mailto:bioingegneria@neuromed.it)

P. Galdi  
MRC Centre for Reproductive Health, University of Edinburgh, Edinburgh EH16 4TJ, UK  
e-mail: [paola.galdi@gmail.com](mailto:paola.galdi@gmail.com)

A. Serra  
Faculty of Medicine and Health Technology, Tampere University, 33200 Tampere, Finland  
e-mail: [angela.serra89@gmail.com](mailto:angela.serra89@gmail.com)

## 11.1 Introduction

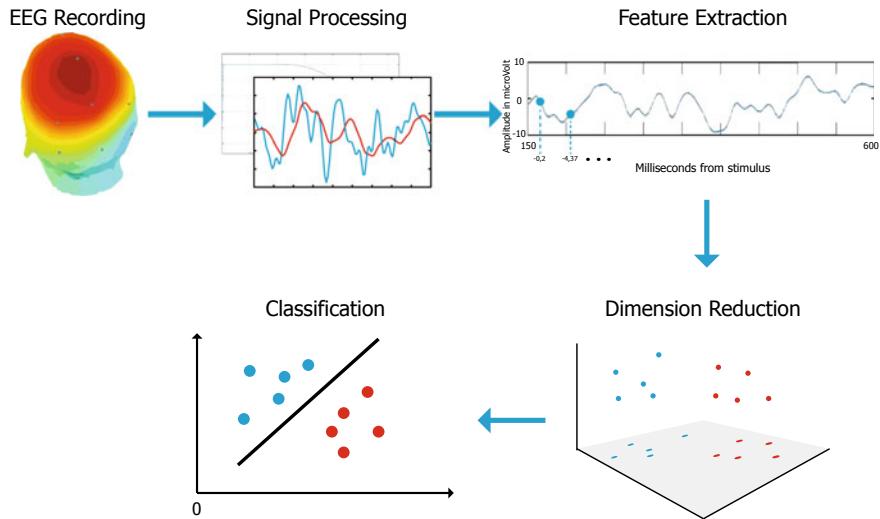
Brain–computer interface (BCI) technology is a set of tools that allow a user to control an external device using nothing else but his “mind,” with no muscle intervention. These systems are of the utmost importance if we consider all people suffering from motor neuron diseases, such as amyotrophic lateral sclerosis (ALS), stroke, and spinal cord injuries. As a consequence of these conditions, the brain and the body do not communicate anymore; however, neuronal activity is still present. Electroencephalography-based BCI systems use this activity, recorded by means of electroencephalography (EEG), to control an external device. User’s intent is inferred from characteristic features of EEG signals, called event-related potentials (ERPs) [1], which represent the brain response to specific cognitive or perceptive processes.

The high number of artifacts (i.e., eyeball movements, power lines, cardiac signal, muscle contraction) and the low amplitude of EEG signals, ranging from 1 to 100  $\mu$ V, make the design of a BCI system very challenging [2]. Great attention has to be paid to signal processing: EEG signal features slightly vary according to health condition [3], concentration level [4], and other factors [5], all of them contributing to inter-subject variability, thus making the analysis even more complicated. In order to deal with this variability, many current works follow a single-subject approach, where analyses are carried out separately for each patient. In this way, the resulting models are more accurate, but a calibration phase is required every time a different patient uses the BCI system.

The alternative is to adopt a multi-subject approach, where data of multiple patients are simultaneously analyzed, as if they came from a single subject. This leads to a ready-to-use system, with no calibration needed; the only disadvantage is a decrease in accuracy due to inter-subject variability. Most of BCI systems presented in literature show satisfying results (over 90% of accuracy [6–10]); however, in many cases, they resort to electrocorticography (ECoG), which results in much cleaner and noise-free signals, but requires invasive surgery. Furthermore, almost every work we analyzed presents a single-subject approach, sometimes combined with dimension reduction techniques (e.g., PCA in [11, 12]), to reduce the complexity of the system along with the amount of noise.

In this study, we compared for the first time four different dimension reduction techniques, which have been applied to EEG signals we recorded during P300 Speller tasks. PCA, autoencoders [13], and t-distributed stochastic neighbor embedding (tSNE) [14] were already used for BCI applications, but they have never been compared in a single study, while neighborhood component analysis (NCA) has never been used before for BCI applications, and it shows a very interesting “compression ratio” of model dimensionality [15].

After dimension reduction, we evaluated the performances of linear and nonlinear classifiers; this is the last building block of our BCI system, as shown in Fig. 11.1. Our comparison was made following a multi-subject approach with a larger sample of subjects compared with those used in literature.



**Fig. 11.1** System overview. Here we find typical components of a BCI system: the acquisition device, in our case, an EEG device; the signal processing system, to filter out as much noise as possible; feature extraction, in our case followed by dimension reduction techniques, to reduce the complexity of the system and to further remove noise from our data. Finally, we have the classifiers for predicting the user’s intent

The chapter is organized as follows: In Sect. 11.2, we describe our dataset, with the illustration of our signal processing pipeline plus a brief introduction to the dimension reduction techniques we used in our study; in Sect. 11.3, we present experimental results and a discussion about them; finally, in Sect. 11.4, we provide some concluding remarks.

## 11.2 Materials and Methods

### 11.2.1 Population Used in the Study

For our study, we acquired EEG signals from 25 healthy individuals, aged 19–34. None of them has been diagnosed with brain or motor disease or language problems. All subjects have never had any experience with BCI systems, and they agreed with the conditions of our experiment. Only the data from 24 subjects have been used. One subject was excluded because he was inattentive and moved excessively.

QUELFEZSGHEMBO(U)					
QU					
A	B	C	D	E	F
<b>G</b>	<b>H</b>	<b>I</b>	<b>J</b>	<b>K</b>	<b>L</b>
M	N	O	P	Q	R
S	T	U	V	W	X
Y	Z	1	2	3	4
5	6	7	8	9	_

**Fig. 11.2** P300 Speller matrix. In the upper part of the screen, there are two lines: The former is the phrase to be spelled; the latter contains characters predicted by the classifier. The subject is asked to focus his attention on the  $6 \times 6$  character matrix and specifically on the character indicated in round brackets to the right of the phrase to be spelled. Rows and columns of the matrix are randomly highlighted: When the row or the column contains the character in focus, a P300 ERP is recorded

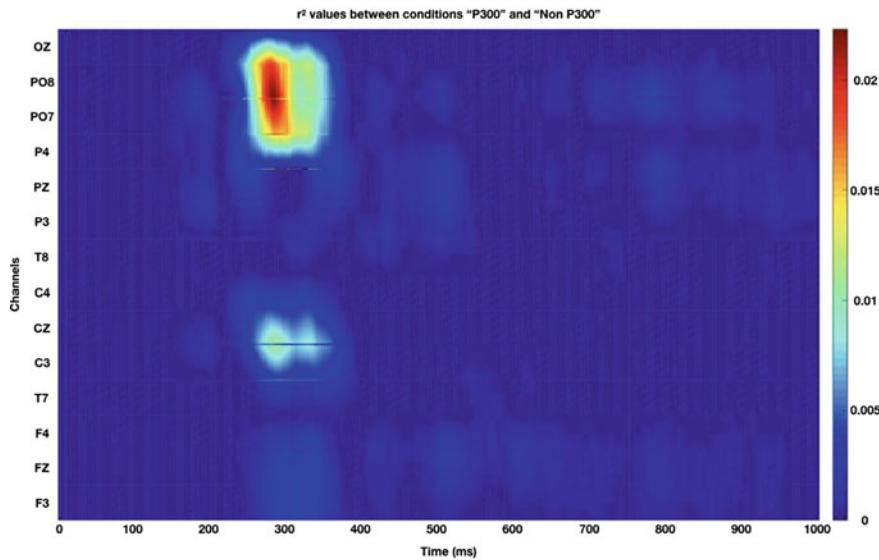
### 11.2.2 Experimental Design

Each subject participated in two identical P300 Speller sessions, for a total duration of 25 minutes (10 minutes per session plus a 5-minute break between the sessions). Participants had to spell the pangram *QUELFEZSGHEMBOCOPREDAVANTI*, that was divided into two blocks, *QUELFEZSGHEMBO* and *COPREDAVANTI*, with a 2-minute break between the blocks. We used the BCI2000 P300 Speller application [16] (Fig. 11.2) to implement our experiment and to acquire EEG data. This application is based on the P300 wave, an event-related potential (ERP) component (an electrical potential shift in EEG) recorded about 300ms after an expected stimulus is presented to the subject.

Subjects wore an EEG cap (g.GAMMAcap produced by g.tec Guger Technologies) with 14 active electrodes (F3, Fz, F4, T7, C3, Cz, C4, T8, P3, Pz, P4, PO7, PO8, and Oz) connected to a g.USBamp (256Hz sampling rate), and we considered A2 as the reference electrode [17]. Acquisitions took place at Mediterranean Neurological Institute “Neuromed,” which also provided the above-mentioned EEG instrumentation.

### 11.2.3 Signal Processing

To preserve meaningful information only, we selected the most discriminating channels among the 14 we used in our study, on the basis of the coefficient of determination  $r^2$  [18], which measures the total signal variance that is determined by the task condition (the presence of the P300 component, i.e., the presentation of the expected

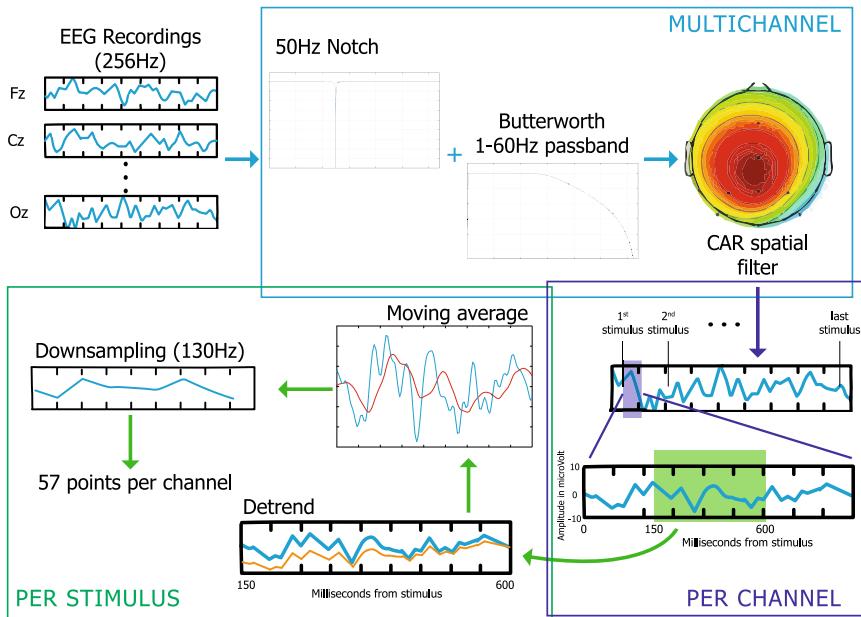


**Fig. 11.3**  $r^2$  values between P300 and non-P300 conditions. This is a sample plot for one of the 24 subjects we acquired. We see that the P300 component is mostly located in the parieto-occipital brain area (PO7, PO8, Oz), and its peak is approximately at 300ms after the stimulus

stimulus). We used BCI2000 *Offline Analysis Tool* to plot  $r^2$  values separately for each subject (a sample plot for one of the subjects is shown in Fig. 11.3), and we identified eight channels which best discriminated our two conditions (presence vs. absence of P300 component): Fz, Cz, P3, Pz, P4, PO7, PO8, and Oz. These are the most frequently used channels for P300 Speller applications [3, 12, 19].

All signals have been filtered by a 50 Hz notch filter [20] and an eighth-order Butterworth filter [21] between 1 and 60 Hz [10, 22–24]. Subsequently, we applied a common average reference (CAR) spatial filter, as it has been shown that it improves P300 Speller performances [25].

Then, we extracted epochs (trials) using stimuli markers. Each trial goes from 150 to 600 ms after the stimulus onset [26]; all further signal processing steps have been applied separately for each epoch. We removed linear trends from EEG signals using MATLAB *detrend* function, and we subsequently filtered the signals by using a moving average filter, where the order of the filter is the ceiling of the ratio between the sampling frequency (256 Hz) and the chosen decimation frequency (130 Hz) [16]. Downsampling is the last step of our signal processing chain (Fig. 11.4).

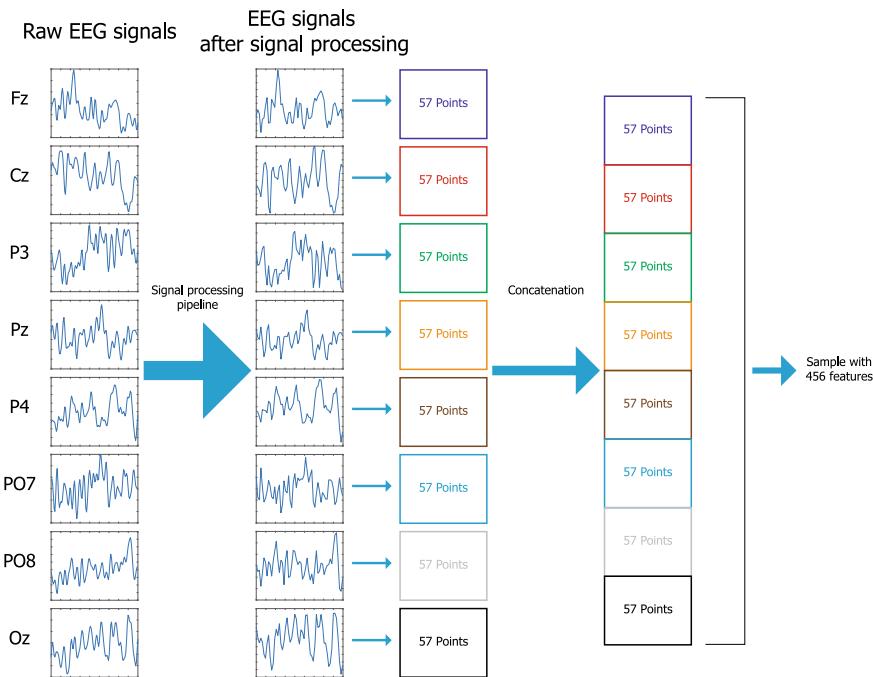


**Fig. 11.4** Signal processing pipeline. Raw EEG recordings of the eight selected channels are filtered by a notch filter and then by a 1–60 Hz eighth-order Butterworth filter. After the application of the CAR spatial filter, trials were extracted separately for each channel, by selecting the signal from 150 to 600 ms after each stimulus. Then, for each trial separately, we applied detrending, moving average filtering, and downsampling to 130 Hz. We obtained 57 points per stimulus, where these points describe the behavior of the signal of a channel after that stimulus. A data point in the dataset is relative to a single stimulus, and it is obtained through the concatenation of the 57 points derived by each of the eight channels

### 11.2.4 Feature Vector

As a result of our signal processing procedure, which was applied independently for each subject, we obtained 187200 non-P300 trials and 37440 P300 trials. The feature vector associated with a given stimulus consists of the concatenation of the samples from eight channels, relative to the epoch associated with the stimulus (Fig. 11.5). The total number of features is 456, which is the number of samples in a 450 ms window (57) multiplied by the number of channels (8).

We finally scaled our data using scikit-learn *RobustScaler*, and we removed outliers with scikit-learn *LocalOutlierFactor*. Because of computational limitations, we used only 10,000 points (5000 non-P300 and 5000 P300) for training and 10,000 for testing (5000 non-P300 and 5000 P300), sampled at random.



**Fig. 11.5** Construction of a feature vector relative to a single stimulus. Raw EEG signals are processed with the pipeline described in Fig. 11.4, resulting in 57 points per channel; samples from each channel are concatenated; the results are a vector of 456 features

### 11.2.5 Dimension Reduction

We applied dimension reduction techniques to reduce the number of features in order to reduce the complexity of the classification task and to improve the accuracy of the selected classifiers. We compared the performance, expressed in terms of accuracy, of four different methods: PCA, NCA, tSNE, and autoencoders.

PCA is one of the most used dimension reduction techniques. In BCI systems, especially P300 applications, PCA generally leads to an increase of accuracy [11, 27–29]. We applied PCA on the training data, and we evaluated accuracy with four different numbers of principal components: We selected the first  $k$  components explaining 80, 90, 95, and 99% of the data variance.

NCA is a supervised and easy-to-use dimension reduction technique, which has been never used in BCI applications. NCA aims at finding a linear transformation of input data, such that the average leave-one-out classification of a KNN classifier is maximized. We calculated the intrinsic dimension of training data to set the number of dimensions of NCA [15].

tSNE is mainly designed for high-dimensional data visualization, and to the best of our knowledge, it was used only once for P300 systems, and it led to an increase of

accuracy [30]. tSNE is based on the construction of a probability distribution in both the original and low-dimensional spaces, where similar objects have a high probability of being picked and Kullback–Leibler divergence between the two distributions is minimized.

In order to get a mapping to transform our test data, we used a parametric version of tSNE [31], that is based on restricted Boltzmann machines (RBMs) to generate a mapping between original and final spaces. We chose, for parametric tSNE, the same number of neurons we found for the autoencoders using cross-validation (CV).

*Autoencoders* are unsupervised neural networks, where input and output layers have the same number of neurons. This technique attempts to copy its input to its output; internal hidden layers describe a “code” to project the input into a lower-dimensional space. Shallow autoencoders are essentially equivalent to PCA transformations, especially when they are trained using weight decay regularization, but autoencoders with nonlinear encoder functions and nonlinear decoder functions can learn a more powerful nonlinear generalization of PCA [32]. In this chapter, we used ReLU and sigmoid functions, respectively, for encoding and decoding layers. Autoencoders, particularly stacked autoencoders [33], are nowadays very popular, due to the considerable reputation of deep learning. However, we used a shallow network, as we saw that the higher the number of hidden layers, the lower was the accuracy of the model. We chose the autoencoder configuration using cross-validation (CV).

### 11.3 Results and Discussion

In Table 11.1, we reported the results obtained both with and without dimension reduction. We used linear support vector machines (LSVM) [34] and linear discriminant analysis (LDA) [35] for classification, as they are the most used classifiers for P300-based BCI systems. We additionally used SVM with Gaussian kernels (RBF SVM) [36] and K-nearest neighbors (KNNs) [37], in order to evaluate also the performance of nonlinear classifiers. We selected the best combination of parameters for both classifiers and dimension reduction methods by using a fivefold cross-validation. CV mean accuracy and test accuracy have been reported in Table 11.1, along with standard deviation (in brackets).

Using the dimension reduction techniques we compared in this study, we dramatically reduced the complexity of the classification model, especially with NCA, where the number of features was reduced from 456 down to 5, with a model simplification of 99%!

Accuracy has not been penalized by these techniques, since it was very close to that we obtained considering all 456 features, and PCA actually led to a slightly better test accuracy, reaching a maximum value of 64.85%. Figure 11.6 shows ROC curves for our dimension reduction techniques: As we can see, we are above chance level in all cases, with PCA and autoencoder curves (especially for the PCA) which are similar to the one we obtained with no dimension reduction.

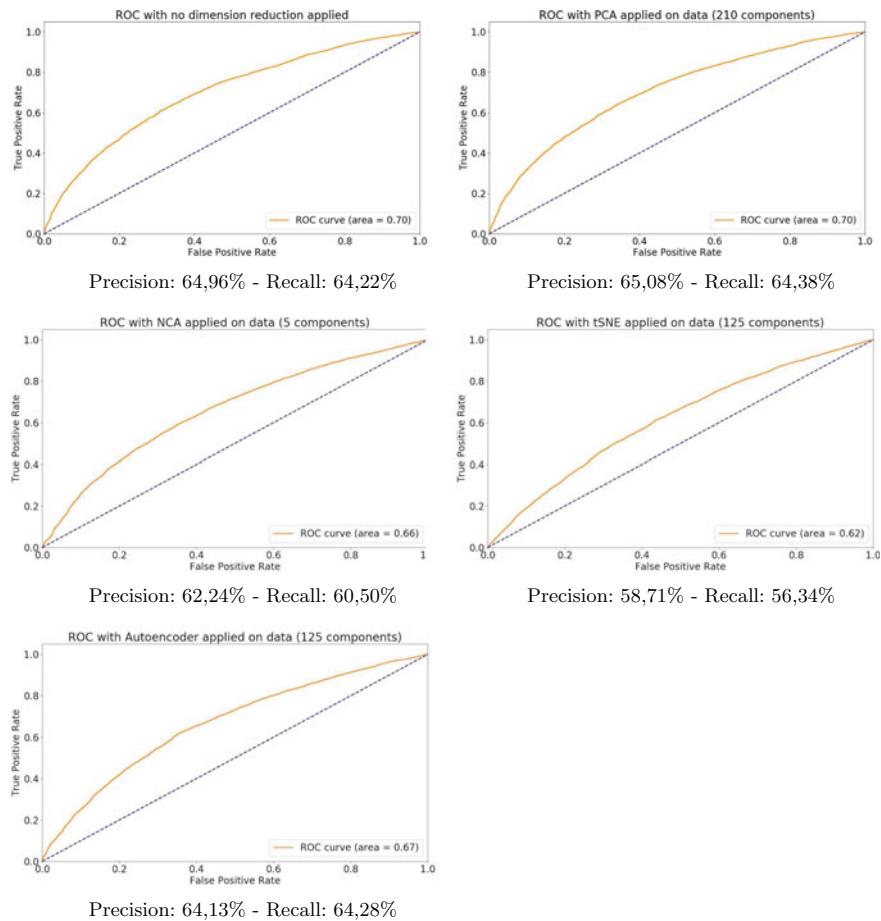
**Table 11.1** Fivefold cross-validation mean accuracy and test accuracy. These values are related to the best combination of parameters, both for classifiers and for dimension reduction techniques. For PCA, we indicate the number of features and we provide the percentage of variance explained by the selected components

Classifier	Dimension Reduction	Number of features	5-Fold CV	Test (%)
LSVM	—	456	62.40 (+/- 0.7)	64.14
RBF SVM	—	456	62.47 (+/- 0.8)	<b>64.79</b>
LDA	—	456	61.07 (+/- 0.7)	62.24
KNN	—	456	54.93 (+/- 1.4)	56.00
LSVM	PCA	147 (90% exp. var.)	62.53 (+/- 0.6)	64.17
RBF SVM	PCA	210 (95% exp. var.)	62.70 (+/- 0.9)	<b>64.85</b>
LDA	PCA	147 (90% exp. var.)	62.32 (+/- 0.4)	64.13
KNN	PCA	82 (80% exp. var.)	56.26 (+/- 1.1)	57.73
LSVM	NCA	5	60.30 (+/- 0.8)	61.44
RBF SVM	NCA	5	60.40 (+/- 0.6)	61.90
LDA	NCA	5	60.30 (+/- 0.6)	61.34
KNN	NCA	5	58.60 (+/- 0.7)	59.47
LSVM	TSNE	125	57.50 (+/- 0.7)	58.36
RBF SVM	TSNE	125	57.50 (+/- 0.8)	58.36
LDA	TSNE	125	56.70 (+/- 1.1)	58.16
KNN	TSNE	125	56.10 (+/- 0.5)	57.42
LSVM	AUTOENCODER	121	62.37 (+/- 0.3)	63.71
RBF SVM	AUTOENCODER	125	62.38 (+/- 0.3)	<b>64.10</b>
LDA	AUTOENCODER	121	62.30 (+/- 0.7)	63.96
KNN	AUTOENCODER	125	56.03 (+/- 1.0)	56.06

However, even if we obtained satisfactory results in terms of reduction of model complexity, accuracy is, in our opinion, still too low to build an effective online system. The performance of the system is probably negatively affected by artifact contamination, and there is a need for better noise reduction approaches.

## 11.4 Conclusions

The aim of this work was to evaluate the efficacy of four popular dimension reduction techniques in the context of a BCI system. We processed our EEG signals with well-established methods in this field: notch filter, Butterworth bandpass filter, CAR



**Fig. 11.6** ROC curves and precision and recall values for all the dimension reduction techniques we used in this study. We used RBF SVM to compute these results, as it resulted in being the best classifier for our analysis

spatial filter, and moving average filter. We then analyzed signals from 150 to 600 ms after each stimulus, as to fully include the P300 component. All the dimension reduction techniques we applied led to a dramatic reduction in the number of features (i.e., the complexity of our classification model), with a very small loss in accuracy (0.5–3%), compared to that we obtained with no dimension reduction (nearly 64%). Autoencoder and PCA performed better on our data, with SVM classifier (with both linear and RBF kernel) being the combination which globally led to a better test accuracy. Better results, for a multi-subject study, have been obtained in [33], where stacked autoencoders led to a 69.5% of accuracy. However, in that work training was made using fewer samples (90 per task condition) and data from just four subjects; furthermore, the quality of EEG acquisitions and the resulting amount of artifacts

has to be considered. Our conclusions are that further noise reduction methods have to be evaluated, in order to realize an effective online BCI. Our future research will be to combine promising approaches, as, for example, wavelet transform [38, 39], ICA [40, 41], and deep learning techniques.

## References

1. Sur, S., Sinha, V.K.: Event-related potential: an overview. *Ind. Psychiatry J.* **18**(1), 70 (2009)
2. Repovs, G.: Dealing with noise in EEG recording and data analysis. *Informatica Medica Slovenica* **15**(1), 18–25 (2010)
3. McCane, L.M., et al.: P300-based brain-computer interface (BCI) eventrelated potentials (ERPs): people with amyotrophic lateral sclerosis (ALS) versus age-matched controls. *Clin. Neurophysiol.* **126**(11), 2124–2131 (2015)
4. da Silva-Sauer, L., et al.: Concentration on performance with P300- based BCI systems: a matter of interface features. *Appl. Ergon.* **52**, 325–332 (2016)
5. Polich, J.: Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* **118**(10), 2128–2148 (2007)
6. Brunner, P., et al.: Rapid communication with a “P300” matrix speller using electrocorticographic signals (ECOG). *Front. Neurosci.* **5**, 5 (2011)
7. Hochberg, L.R., et al.: Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. In: *Nature* **485**(7398), 372 (2012)
8. Guy, V., et al.: Brain computer interface with the P300 speller: usability for disabled people with amyotrophic lateral sclerosis. *Ann. Phys. Rehabil. Med.* **61**(1), 5–11 (2018)
9. Wang, J., Liu, Y., Tang, J.: Fast robot arm control based on brain-computer interface. In: *Information Technology, Networking, Electronic and Automation Control Conference, IEEE*, pp. 571–575. IEEE (2016)
10. Speier, W., et al.: A comparison of stimulus types in online classification of the P300 speller using language models. *PloS one* **12**(4), e0175382 (2017)
11. Elsawy, A.S., et al.: A principal component analysis ensemble classifier for P300 speller applications. In: *2013 8th International Symposium on Image and Signal Processing and Analysis (ISPA)*, pp. 444–449. IEEE (2013)
12. Lotte, F., Guan, C.: An efficient P300-based brain-computer interface with minimal calibration time. In: *Assistive Machine Learning for People with Disabilities Symposium (NIPS’09 Symposium)* (2009)
13. Baldi, P.: Autoencoders, unsupervised learning, and deep architectures. In: *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, pp. 37–49 (2012)
14. van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2579**–2605 (2008)
15. Goldberger, J., et al.: Neighbourhood components analysis. *Adv. Neural Inf. Process. Syst.* 513–520 (2005)
16. Schalk, G., Mellinger, J.: *A Practical Guide to Brain-Computer Interfacing with BCI2000: General-Purpose Software for Brain-Computer Interface Research. Data Acquisition, Stimulus Presentation, and Brain Monitoring*. Springer (2010)
17. Siuly, S., Li, Y., Zhang, Y.: *EEG Signal Analysis and Classification*. Springer (2016)
18. Coefficient of Determination, Determination Coefficient, RSquared (2013). <https://www.bci2000.org/mediawiki/index.php/Glossary>. Visited on 06 Nov 2018
19. Krusinski, D.J., et al.: A comparison of classification techniques for the P300 Speller. *J. Neural Engi.* **3**(4), 299 (2006)
20. Yamada, T., Meng, E.: *Practical Guide for Clinical Neuro-Physiologic Testing: EEG*. Lippincott Williams & Wilkins (2012)

21. Coyle, D.: Brain-Computer Interfaces: Lab Experiments to Real-world Applications, vol. 228. Elsevier (2016)
22. Filters in the Electroencephalogram (2015). <https://clinicalgate.com/filters-in-the-electroencephalogram/>. Visited on 06 Nov 2018
23. Spataro, R., et al.: Reaching and grasping a glass of water by locked-In ALS patients through a BCI-controlled humanoid robot. *Front. Hum. Neurosci.* **11**, 68 (2017)
24. Rakotomamonjy, A., Guigue, V.: BCI competition III: dataset II-ensemble of SVMs for BCI P300 speller. *IEEE Trans. Biomed. Eng.* **55**(3), 1147–1154 (2008)
25. Alhaddad, M.J.: Common average reference (CAR) improves p300 speller. *Int. J. Eng. Technol.* **2**(3), 21 (2012)
26. Lugo, Z.R., et al.: A vibrotactile p300-based brain-computer interface for consciousness detection and communication. *Clin. EEG Neurosci.* **45**(1), 14–21 (2014)
27. Sharma, N.: Single-trial P300 Classification using PCA with LDA, QDA and Neural Networks. arXiv preprint [arXiv:1712.01977](https://arxiv.org/abs/1712.01977) (2017)
28. Selim, A.E., Wahed, M.A., Kadah, Y.M.: Electrode reduction using ICA and PCA in P300 visual speller brain-computer interface system. In: 2014 Middle East Conference on Biomedical Engineering (MECBME), pp. 357–360. IEEE (2014)
29. Kundu, S., Ari, S.: P300 Detection with brain-computer interface application using PCA and ensemble of weighted SVMs. *IETE J. Res.* 1–9 (2017)
30. Jadidi A.F., Zargar, B.S., Moradi, M.S.: Categorizing visual objects; using ERP components. In: 2016 23rd Iranian Conference on Biomedical Engineering and 2016 1st International Iranian Conference on Biomedical Engineering (ICBME), pp. 159–164. IEEE (2016)
31. Maaten, L.: Learning a parametric embedding by preserving local structure. *Artif. Intell. Stat.* 384–391 (2009)
32. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning MIT Press (2016). <http://www.deeplearningbook.org>
33. Vařeka, L., Mautner, P.: Stacked autoencoders for the P300 component detection. *Front. Neurosci.* **11**, 302 (2017)
34. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
35. McLachlan, G.: Discriminant Analysis and Statistical Pattern Recognition, vol. 544. Wiley (2004)
36. Shashua, A.: Introduction to machine learning: class notes 67577. arXiv preprint [arXiv:0904.3664](https://arxiv.org/abs/0904.3664) (2009)
37. Altman, N.S.: An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **46**(3), 175–185 (1992)
38. Choudhry, M.S., et al.: A survey on different discrete wavelet transforms and thresholding techniques for EEG denoising. In: 2016 International Conference on Computing, Communication and Automation (ICCCA), , pp. 1048–1053. IEEE (2016)
39. Krishnaveni, V., et al.: Automatic identification and removal of ocular artifacts from EEG using wavelet transform. *Measur. Sci. Rev.* **6**(4), 45–57 (2006)
40. Radünz, T., et al.: EEG artifact elimination by extraction of ICA-component features using image processing algorithms. *J. Neurosci. Methods* **243**, 84–93 (2015)
41. Winkler, I., Haufe, S., Tangermann, M.: Automatic classification of artifactual ICA-components for artifact removal in EEG signals. *Behav. Brain Functions* **7**(1), 30 (2011)

# Chapter 12

## Blind Source Separation Using Dictionary Learning in Wireless Sensor Network Scenario



Angelo Ciaramella , Davide Nardone and Antonino Staiano

**Abstract** The aim of this paper is to introduce a block-wise approach with adaptive dictionary learning for solving a determined blind source separation problem. A potential real-case scenario is illustrated in the context of a stationary wireless sensor network in which a set of sensor nodes transmits data to a multi-receiver node (sink). The method has been designed as a multi-step approach: the estimation of the mixing matrix, the separation of the sources by sparse coding and the source reconstruction. A sparse mixture from the original signals is used for estimating the mixing matrix, and later on, a sparse coding approach is used for separating the block-wise sources which are finally reconstructed by means of a dictionary. The proposed model is based on a block-wise approach which has the advantage of considerably improving the computational efficiency of the signal recovery process without particularly degrading the separation performance. Some experimental results are provided for comparing the computational and separation performances of the proposed system by varying the type of dictionary used, whether it is fixed or learned from the data.

### 12.1 Introduction

The human perception of acoustic mixtures, commonly referred to as the *cocktail party problem* [7], results from the vibration of the eardrum due to the overlapping of multiple signals that are emitted from different audio sources at the same time. Several techniques, including blind source separation (BSS), have been used for addressing this problem. A well-known technique used for BSS is the independent

---

A. Ciaramella · D. Nardone ( ) · A. Staiano  
Dept. of Science and Technology, University of Naples “Parthenope”,  
Centro Direzionale, Isola C4, 80143 Naples, Italy  
e-mail: [davide.nardone@studenti.uniparthenope.it](mailto:davide.nardone@studenti.uniparthenope.it)

A. Ciaramella  
e-mail: [angelo.ciaramella@uniparthenope.it](mailto:angelo.ciaramella@uniparthenope.it)

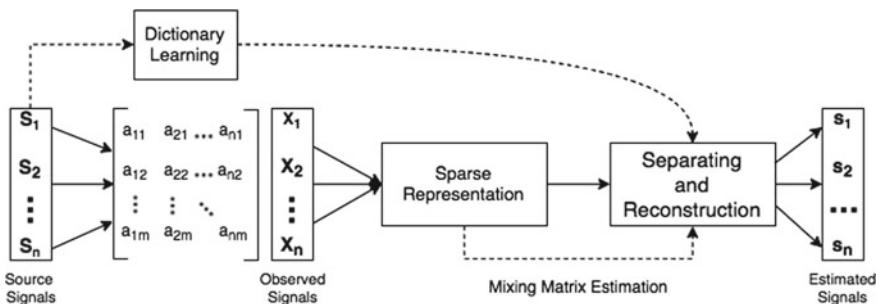
A. Staiano  
e-mail: [antonino.staiano@uniparthenope.it](mailto:antonino.staiano@uniparthenope.it)

component analysis (ICA) [7] which achieves its best performance on a determined or overdetermined case. In this work, a wireless sensor network (WSN) scenario has been hypothesised, where a set of sensor nodes (i.e. microphones) transmits data to a multi-receiver node, named sink in the WSNs jargon. A block-wise approach with adaptive dictionary learning for solving a determined blind source separation problem is introduced. To solve BSS, a *sparse representation of the signals* is adopted [3, 4], assuming that the sources are sparse or that they can be decomposed as a combination of sparse components. Based on this representation, the multi-step approach is used for signal recovering.

The paper is organised as follows. In Sect. 12.2, the proposed methodology and the used techniques are introduced. In Sect. 12.3, some experimental results are described and analysed. Finally, in Sect. 12.5, some conclusions and future actions are outlined.

## 12.2 Proposed Methodology

For accomplishing a determined blind source separation (DBSS) problem, firstly one has to estimate the mixing matrix  $\hat{A}$  and then the sources. In this work, the BSS model has been reformulated into a sparse signal recovery model with an *adaptive dictionary* learned from data. As depicted in Fig. 12.1, the methodology is made up of four different steps: (1) dictionary learning, (2) matrix estimation, (3) separation of sources and (4) their reconstruction. The proposed model is based on a block-wise approach, where once the dictionary is learned and the mixing matrix estimated, the sparse sources are recovered through a sparse coding approach. The components of the separate sources are then reshaped and linearly combined with the atoms of dictionary for the fully reconstruction of the original signals. The block-wise operation has the advantage of considerably improving the computational efficiency of the signal recovery process without particularly degrading the separation performance. Details of all phases are shown in the following sections.



**Fig. 12.1** Proposed system workflow for separating  $n$  signals from  $n$  mixtures

### 12.2.1 *Blind Source Separation*

Given  $m$  observations  $\{\mathbf{x}_1, \dots, \mathbf{x}_m\} \in \mathbb{R}^t$  arranged as the rows of the data matrix  $\mathbf{X}$ , we can express them as a linear combination of  $n$  given sources  $\{\mathbf{s}_1, \dots, \mathbf{s}_n\} \in \mathbb{R}^t$  and  $m$  stochastic processes  $\{\mathbf{a}_1, \dots, \mathbf{a}_m\} \in \mathbb{R}^n$ :

$$\forall i \in \{1, \dots, m\}, \quad x_i = \sum_{j=1}^n a_{ij} s_j. \quad (12.1)$$

The BSS techniques aim at recovering the original  $S = [\mathbf{s}_1^t, \dots, \mathbf{s}_n^t]^T$  taking advantage of some information contained in the way the signals are mixed in the observed data.

### 12.2.2 *Dictionary Learning*

The sparse decomposition of a signal, however, relies heavily on the degree of fit between the data and the dictionary, which leads to another important problem, namely the design of the  $\mathbf{D}$  dictionary. Based on some research, two main approaches are usually used: the *analytic approach* and the *learning-based approach*. In the first approach, a mathematical model is provided in advance for the dictionary (e.g. Fourier transform (FFT), the discrete cosine transform (DCT), the transformed wavelet, curvelets).

The second approach uses *machine learning* techniques for learning the dictionary from a set of data, so that its atoms can accurately represent the aspects of the signals used to build it.

The dictionary learning methodology uses a training set of signals  $s_i, i = 1, \dots, n$ , and it is equivalent to solve the following optimisation problem:

$$\begin{aligned} & \min_{\mathbf{D}, \{\lambda_i\}_{i=1}^n} \quad \sum_{i=1}^n \|\lambda_i\|_0 \\ & \text{subject to } \|\mathbf{s}_i - \mathbf{D}\lambda_i\|_2^2 \leq \varepsilon, \quad 1 \leq i \leq n \end{aligned} \quad (12.2)$$

This problem tries to jointly find the sparse representation of the signals and the dictionary. The role of penalty and constraints might also be reversed if we choose to constrain the sparsity and obtain the best fit for it, therefore solving the following optimisation problem.

The set of  $m$ -dimensional  $\mathbf{s} = (s_1, \dots, s_n)$  vectors could be concatenated by column-wise forming a matrix  $\mathbf{S} \in \mathbb{R}^{m \times n}$  and likewise, all the corresponding sparse representations in a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ . Therefore, the dictionary would satisfy the relation:

$$\mathbf{S} = \mathbf{D}\mathbf{A} \quad (12.3)$$

In this work, we used the *method of optimal directions* (MOD) [5] or *online dictionary learning* [8] as dictionary learning techniques.

The mixing model is appropriately rewritten in the following matrix form:

$$\mathbf{X} = \mathbf{AS}^T + \mathbf{N} \quad (12.4)$$

where  $\mathbf{A} \in \mathbb{R}^{k \times n}$ , with  $k \leq n$  is the unknown mixing matrix,  $\mathbf{X} \in \mathbb{R}^{k \times m}$  is the data matrix observed, whose the row vector  $\mathbf{x}_i$  is the  $i$ -th sensor signal and,  $\mathbf{N} \in \mathbb{R}^{k \times n}$  is the noise matrix containing noisy vectors.

### 12.2.3 Mixture Decomposition into Sparse Signals

The DBSS system we proposed needs to consider the data sources split into blocks/patches. In fact, after creating the mixture matrix  $\mathbf{X}$ , every single mixture  $\mathbf{x}_i$  is re-arranged as a matrix block. Each block is concatenated with the other so as to create a new mixture matrix  $\mathbf{X}_s$ . Since the *source separation* phase take into account sparse coding techniques, it is necessary decomposing each signal of the original mixture  $\mathbf{X}$  into sparse vectors  $\mathbf{x}_s$ , according to the following optimisation problem:

$$\begin{aligned} \min_{\mathbf{x}_s} \quad & \|\mathbf{x}_s\|_0 \\ \text{subject to} \quad & \|\mathbf{x}_i - (\mathbf{D}\mathbf{x}_s)^T\|_2^2 \leq \epsilon, 1 \leq i \leq k \end{aligned} \quad (12.5)$$

for each  $i = 1, \dots, n$ . In particular we have that

$$\mathbf{x}_i = \mathbf{A}_i \mathbf{S}^T = \mathbf{A}_i (\mathbf{D}\mathbf{\Lambda})^T \quad (12.6)$$

where  $\mathbf{A}_i$  is the  $i$ -th row of matrix  $\mathbf{A}$ . We stress that the constraint implies that

$$\|\mathbf{A}_i (\mathbf{D}\mathbf{\Lambda})^T - (\mathbf{D}\mathbf{x}_s)^T\|_2^2 = \|\mathbf{D}^T (\mathbf{A}_i \mathbf{\Lambda}^T - \mathbf{x}_s^T)\|_2^2. \quad (12.7)$$

To estimate the  $\mathbf{A}$ , a BSS mechanism will be adopted. In order to solve it, we relax the  $\ell_0$ -norm to  $\ell_1$ -norm. The  $\ell_1$ -norm minimisation problem is known as Basis Pursuit (BP) [2] which is a convex optimisation problem that can be recast as a linear programming problem. One of the several algorithms used for solving problems involving the  $\ell_1$ -norm is the *Orthogonal Matching Pursuit* (OMP) [9].

Every single mixture  $\mathbf{x}_s$  of the  $i$ -th mixture is re-arranged as a matrix block  $\mathbf{x}_s^i$ . Each block is concatenated with the other so as to create a new mixture matrix  $\mathbf{X}_s$ . Due to the compressing property exhibited by the sparse coding approaches, this process makes the signal recovering computationally less demanding, since the source separation problem is solved only for those vectors which are not full sparse (which have at least one nonzero component).

### 12.2.4 Estimate of the Mixing Matrix

For estimating the mixing matrix, we adopt the *Generalised Morphological Component Analysis (GMCA)* [1], that is a method for BSS exploiting both morphological and signal diversity (i.e. sparsity).

This method solves the following optimisation problem:

$$\{\hat{\mathbf{A}}, \hat{\mathbf{S}}\} = \arg \min_{\mathbf{A}, \mathbf{S}} \kappa \|\mathbf{X}_s - \mathbf{AS}\|_2^2 + \|\mathbf{S}\|_0 \quad (12.8)$$

### 12.2.5 Separating Sources Using Sparse Coding

Once the mixing matrix  $\hat{\mathbf{A}}$  is estimated, we formulate the DBSS as a sparse signal recovery problem by solving  $t$  times the following expression:

$$\mathbf{x}_s^i(t) = \hat{\mathbf{A}}\mathbf{s}(t) + \mathbf{n}(t), \quad 1 \leq t \leq T \quad (12.9)$$

where  $T$  is the number of nonzero column vectors computed during the *mixture decomposition into sparse signals* process. In solving this problem, we try to find the column vector  $\mathbf{s}(t)$  given the vector  $\mathbf{x}_s(t)$  for all the  $T$  components.

The above optimisation process for estimating  $\mathbf{s}(t)$  results in solving the following  $\ell_0$  optimisation problem:

$$\begin{aligned} \min_{\mathbf{s}(t)} \quad & \|\mathbf{s}(t)\|_0 \\ \text{subject to } \mathbf{x}_s^i(t) = \hat{\mathbf{A}}\mathbf{s}(t) \quad & \text{for } t = 1, \dots, T, \end{aligned} \quad (12.10)$$

where  $\|\mathbf{s}(T)\|_0$  is the  $\ell_0$ -norm measuring the sparsity of the signal  $\mathbf{s}$ . As for (Eq. 12.10), we relax the  $\ell_0$ -norm to  $\ell_1$ -norm, solved by means of OMP.

### 12.2.6 Source Reconstruction from Sparseness

Once the sources have been separated and the corresponding sparse representations  $\mathbf{S}_s = (s(1), \dots, s(T))$  are obtained, the reconstruction/estimation of the original sources  $\mathbf{S}_{est}$  is obtained by simply projecting each sparse vector  $\mathbf{s}(i)$  by using the dictionary  $\mathbf{D}$

$$\mathbf{s}_{est}(i) = D\mathbf{s}(i) \quad (12.11)$$

## 12.3 Experimental Results

A wireless sensor network (WSN) is an interconnection of autonomous sensors (e.g. microphones) that collect data from the surrounding environment and transmit it through the network in a primary location. In this kind of network, each node can be connected to several other nodes; therefore, it is possible to receive a mixture of signals to the receiver. The *sparse signal recovering* does not assume that the original signals are independent, so it is a good fit in WSNs where nodes can send highly correlated data. In order to transmit data through the network and then correctly deliver messages to the recipient (user), it is necessary that the  $i$ -th receiver node separates the sources from the mixtures it receives. The real case presented here is configured as a *BSS* problem that can be solved using the *sparse adaptive decomposition* system presented in this work.

### 12.3.1 Principles and Operating Simulation in a WSN

In a WSN configuration, the first phase (before data are sent) a learning phase takes place, where the samples acquired from each sensor nodes are used for learning a dictionary which is then exclusively sent to the other sensor nodes. This is a key phase because the dictionary plays a crucial role in the reconstruction process. Later on, each sensor node (belonging to a cluster) sends at the same time instant  $t$  a message containing information about the detected event. Since multiple nodes can send messages simultaneously, the receiving node (with multiple receive antennas) needs to separate the mixture before being able to forward it to another node.

In the second phase, the generic node decomposes the *mixture of signals* so as to obtain a sparse representation, obtained by solving several OMP problems. These sparse representations are then reshaped and concatenated for obtaining the matrix  $\mathbf{X}_s$ . In the third phase, based on the mixtures, a mixing matrix is estimated using the *GMCA* algorithm. In the fourth phase, the sparse representations of the mixtures are separated. In the context of WSN, we would like to transmit messages from multiple sensor nodes to a receiving node having multiple receiving antennas. For each time instant  $t$ , the receiving node tries to find a vector  $\mathbf{s}(t)$  given  $\mathbf{x}(t)$  (according to the Eq. 12.10). Finally, in the last phase, the obtained vectors are then expanded using the dictionary and defined as separate sources.

### 12.3.2 Datasets and Evaluation Metrics

The proposed method has been evaluated taking into account the source signals of the *underdetermined speech and music mixtures* related to the *Sixth Community-Based Signal Separation Evaluation Campaign*, SiSEC 2015 [10, 11]. The referring dataset

consists of three folders (div1, div2 and div3), each containing different audio files of male, female or musical instrument. Each speech signal has a duration of 10 s, sampled at 16 kHz.

To qualitatively evaluate our method, three performance criteria defined in the BSSEVAL [12] toolbox have been used. These criteria are: *signal-to-distortion ratio* (SDR), *source-to-interference ratio* (SIR) and *source-to-artefacts ratio* (SAR) [12].

According to [12], both SIR and SAR measure *local* performance. SIR mainly measures how well the algorithm performs for the suppression of interfering sources, while SAR measures the amount of artefact is within the separated (target) source. SDR is a global performance index, which might give better evaluation of the overall performance of the algorithms compared. For this reason, in the following experiments, we will focus more on the interpretation of the SDR results than to the results provided by the SIR and SAR.

### 12.3.3 Separation Results with Fixed Dictionary

In this section, we wanted to report the results of the proposed method by replacing the *the dictionary learning* step (illustrated in Fig. 12.1) with one of the following fixed dictionaries, DCT, Haar(L1) and {SIN, COS}. Alternatively, we can learn the dictionaries by using the MOD or ODL methods. From the results shown in Table 12.1, it is possible to observe that the *performance* achieved by using MOD or ODL is better than those obtained using fixed dictionaries.

### 12.3.4 Comparing Different Strategies for Learning Dictionary

Alternatively from the fixed dictionary, it is possible to learn the dictionary by using approaches such as MOD or ODL. Depending on the selected method, the number of parameters slightly changes:

**Table 12.1** Comparison of the separation performance among fixed (i.e. DCT, HWP (L1) and {SIN, COS}) and adaptive dictionary (MOD and ODL) approaches

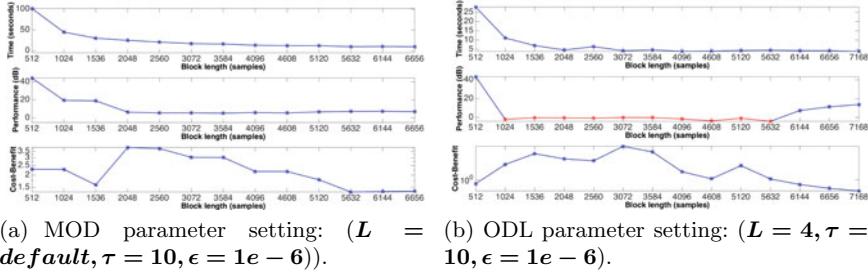
	MOD	ODL	DCT	HWP (L1)	{SIN, COS}
SDR	<b>45.4</b>	43	42.3	41.6	25.7
SIR	46.5	49.8	<b>50.2</b>	50	46
SAR	<b>47</b>	44.7	42.7	43.2	25.7
Time	<b>28</b>	40	50.5	61.4	66.8

- $\mathbf{K}$ : number of trained atoms;
- $\mathbf{P}$ : length of the  $i$ -th atom (dynamic parameter related to the *blocking factor*, see 12.2.3);
- $\tau$ : number of iterations for the method;
- $\epsilon$ : error threshold of  $\|\mathbf{X} - \mathbf{D}\lambda\|_2^2$ ;
- $\mathbf{L}$ : sparsity parameter;
- $\alpha$ : penalty parameter.

For the ODL approach, in addition to the parameters listed above, the parameter  $\eta$  relative to the size of *mini-batch* was also considered. From the results shown in Table 12.1, it is possible to observe that the SDR *separation performance* achieved by MOD or ODL dictionary learning approach is roughly similar to each other but is better than those obtained using fixed dictionaries. In MOD, we set  $\mathbf{K}$  to 470,  $\tau$  to 10,  $\epsilon$  to  $1e-6$  and  $\mathbf{L}$  and  $\alpha$ , respectively, to the default values set by the OMP algorithm. In ODL, the penalty parameter  $\alpha$  was set to 0.15. In particular, the *sparsity parameter*  $\mathbf{L}$  plays a fundamental role in estimating the mixing matrix  $\hat{\mathbf{A}}$  since it increases or decreases the sparsity index of the mixture  $\mathbf{X}_s$  (GMCA input). Regarding the MOD approach, the decision to set  $\mathbf{L}$  to its default value is not a purely random choice, in fact, although this increases the execution time of the method to  $\sim 100$ s, it makes the method more stable to the block changing factor (Fig. 12.2a). On the other hand, the heuristic choice of the parameter  $\mathbf{L}$  leads to good performance for certain blocks sizes but makes the method much more unstable for other dimensions (Fig. 12.2b). For example, comparing the two charts it is possible to note how the heuristic choice of the parameter  $\mathbf{L}$  leads to a strong instability (lot of distortions) when the block size varies (negative SDR values coloured red in Fig. 12.2b) while the use of the default value leads instead to a rather constant performance stability (SDR). Generally speaking, from the obtained results, it is clear that the performance of our method partly depends on some of the parameters listed above; therefore, the joint tuning of all parameters might improve the overall performance. However, we identified that the combination of the parameters listed above is those for which the optimal performances are obtained, for both the MOD and ODL approaches.

### 12.3.5 Effects of the block strategy on the system performance

In this section, we experimentally assess the effect of the block size on the computational and separation performances of the proposed system (using MOD) for both the configurations A and B (Fig. 12.2a, b). The *computational cost* relationship is shown in the upper sub-figure, whereas the *separation performance* (measured with the SDR metric) is shown in the central sub-figure. Every single result in Fig. 12.2 is averaged on 10 runs. For both configurations, it is possible to observe how the system becomes computationally more efficient as the block size increases (result



**Fig. 12.2** Effect of different block lengths on the computational efficiency and separation performance of the proposed method on two different configurations. The cost-benefit graph (i.e. computing time divided by the output SDR) is also shown for each configuration. Red SDR values indicate a poor separation, for which the distortion power is greater than the signal power

due to the greater compression of the sparse mixture  $\mathbf{X}_s$  as the block size increases), while the *separation performance* tends to have a rather constant trend (except for some cases, i.e. 512, 1024, 1536).

For example, when the block size is made of 512 samples, the system takes  $\sim 101 \sim 28$  s and  $\simeq 44$  dB (average SDR) as the computational and the separation performances, respectively, for both the configurations. By increasing the size of the block, the two configurations exhibit a significant difference both in separation performance and execution time. In fact, the *execution time* of the configuration A is on average three times faster than the configuration B, while the *separation performance* proves the best stability of the system for the configuration A. In varying the the block size, one tries to empirically find that block size for which the OMP algorithm converges efficiently and whose separation performance does not deteriorate. Based on the experiments conducted, the best block size is **512** for the configuration B, for which a  $\sim 44$  dB SDR and a running time of  $\sim 28$  s are achieved. Anyway, the running time used by OMP to converge varies according to different factors such as block size, the nature of the signal, the number of artefacts and the hardware used to perform the tests.

Finally, from the results obtained, we might conclude that the performance of the system intrinsically depends on the *sparsity and dimensionality* on the randomness of the  $i$ -th matrix mixture  $\mathbf{X}_s^i \in \mathbb{R}^{p \times np}$ , where  $p$  is the number of atoms in the dictionary and  $np$  is the number of blocks of the mixture. Therefore, since the *GMCA* algorithm works on a sparse mixture  $\mathbf{X}_s$ , the more its size decreases, the more the technique fails for estimating the mixing matrix.

## 12.4 Example of Separation of Four Female Speeches

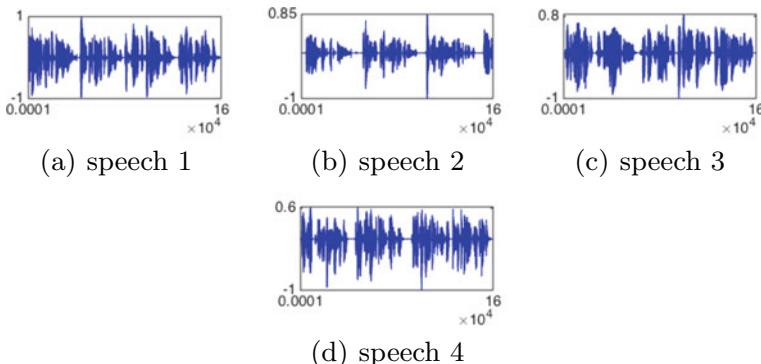
In this section we show the results obtained on a single run by using a matrix block size equal to 512 samples per column. At the beginning of the experiment, four instant mixtures were obtained by mixing four female speeches with the following defined matrix:

$$\mathbf{A} = \begin{pmatrix} 0.89553 & 0.83948 & 0.78655 & 0.43565 \\ 0.21061 & 0.41059 & 0.17296 & 0.77963 \\ 0.13818 & 0.30884 & 0.38806 & 0.3989 \\ 0.36683 & 0.17696 & 0.44814 & 0.20799 \end{pmatrix} \quad (12.12)$$

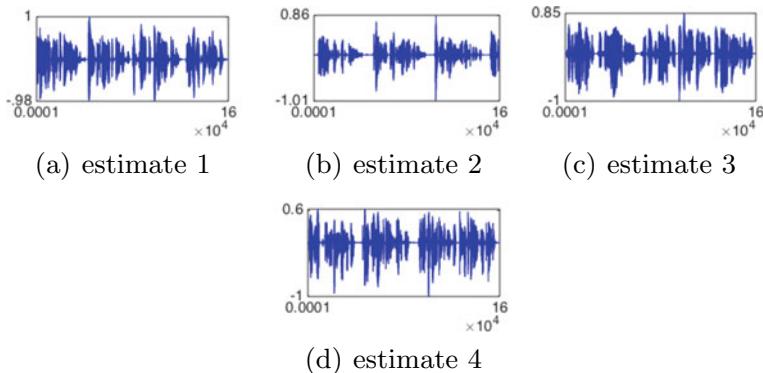
The estimated matrix  $\hat{\mathbf{A}}$  from the four sparse mixture  $\mathbf{X}_s$  using the GMCA algorithm is shown in 12.13. It can be seen that the estimated matrix  $\hat{\mathbf{A}}$  is quite similar to the original mixing matrix  $\mathbf{A}$  with the exception of the permutation ambiguity:

$$\hat{\mathbf{A}} = \begin{pmatrix} 0.89549 & 0.78654 & 0.83874 & 0.43719 \\ 0.21139 & 0.1774 & 0.41247 & 0.77854 \\ 0.13803 & 0.38833 & 0.30903 & 0.39838 \\ 0.36656 & 0.4462 & 0.17574 & 0.20984 \end{pmatrix} \quad (12.13)$$

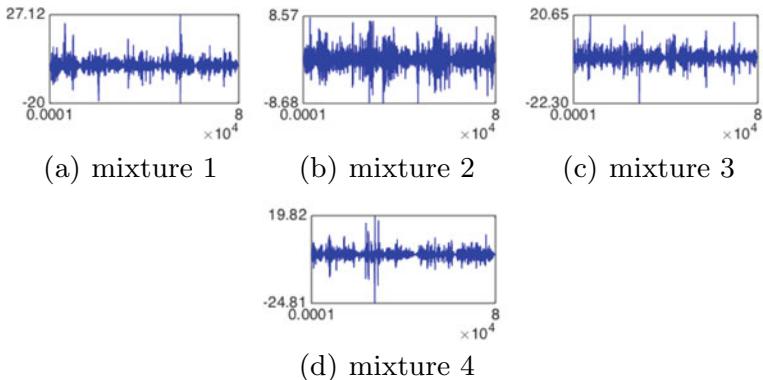
Applying the proposed approach, based on the estimated mixing matrix, it is possible to recover the four speeches by executing the remaining steps in the proposed system. For the mixtures shown in Fig. 12.5, the separation result is shown in Fig. 12.4, where the dictionary used was learned by using the MOD approach. Finally, it can be observed that the estimated sources in Fig. 12.4 are very similar to the original sources in Fig. 12.3.



**Fig. 12.3** Four original female speeches (a–d)



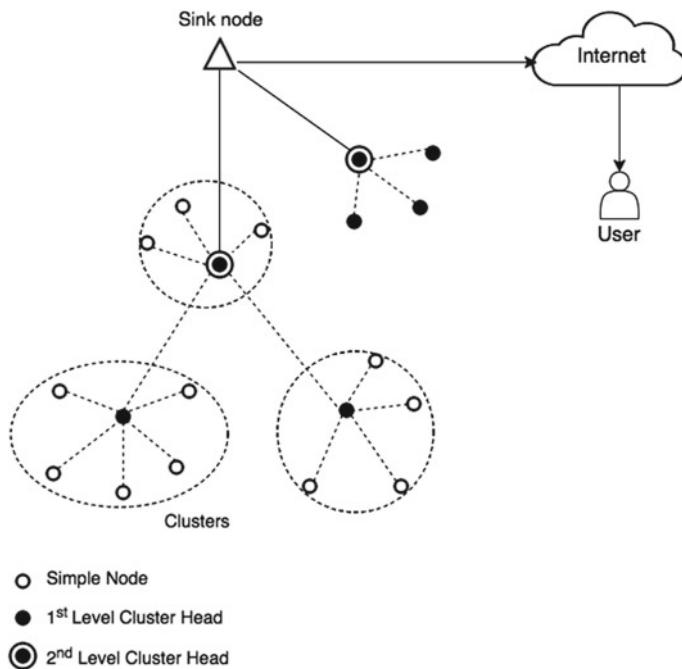
**Fig. 12.4** Four estimated female speeches



**Fig. 12.5** Four mixtures generated by the female speeches (a–d)

## 12.5 Conclusions

We proposed a system for determined blind source separation using a block-wise sparse coding approach with dictionary learning approach. Experimental results show good separation performance of the proposed system, when using dictionary learning methods. The proposed system provides a flexible structure for testing the performance of other dictionary learning algorithm as well as other signal recovery algorithms in source separation applications. Furthermore, the proposed system is also a worthy signal recovery tool in real contexts such as that of wireless sensor networks. In this context, a concrete usage scenario is in the *Low-Energy Adaptive Clustering Hierarchy (LEACH)* [6] MAC protocol (see Fig. 12.6). LEACH is a hierarchical protocol in which most sensor nodes transmit their information to sensor nodes called *cluster head (CH)* which aggregate and compress the data and then forward it to a *sink*, in order to lower energy consumption in WSN which is an imperative task. For underdetermined cases, better methods for estimating the mixing



**Fig. 12.6** Representation of a WSN using the LEACH protocol

matrix need to be used and larger number of atoms would be required in the dictionary. The work can be extended to design dictionaries according to the mixing matrix to ensure maximal separation. Higher accuracy methods for channel estimation may also be employed. Furthermore, since in many real-world applications of WSN, both the sensors and the sources might move, and the approach could be generalised in order to properly take into account the mobility of the main actors.

**Acknowledgements** The research was entirely developed when Davide Nardone was a master degree student in applied computer science at University of Naples Parthenope. This work was partially funded by the University of Naples Parthenope (*Sostegno alla ricerca individuale per il triennio 2016-2018* project).

## References

1. Bobin, J., Starck, J.-L., Fadili, J., Moudden, Y.: Sparsity and morphological diversity in blind source separation. *IEEE Trans. Image Process.* **16**(11), 2662–2674 (2007)
2. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. *SIAM Rev.* **43**(1), 129–159 (2001)
3. Ciaramella, A., Gianfico, M., Giunta, G.: Compressive sampling and adaptive dictionary learning for the packet loss recovery in audio multimedia streaming. *Multimedia Tools Appl.* **75**(24),

- 17375–17392 (2016)
- 4. Ciaramella, A., Giunta, G.: Packet loss recovery in audio multimedia streaming by using compressive sensing. *IET Commun.* **10**(4), 387–392 (2016)
  - 5. Engan, K., Skretting, K., Husøy, J.H.: Family of iterative ls-based dictionary learning algorithms, ils-dla, for sparse signal representation. *Digital Sig. Process.* **17**(1), 32–49 (2007)
  - 6. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H.: Energy-efficient communication protocol for wireless microsensor networks. In: Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, 2000, p. 10. IEEE (2000)
  - 7. Hyvärinen, A., Oja, E.: Independent component analysis: algorithms and applications. *Neural Netw.* **13**(4–5), 411–430 (2000)
  - 8. Mairal, J., Bach, F., Ponce, J., Sapiro, G.: Online dictionary learning for sparse coding. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 689–696. ACM (2009)
  - 9. Pati, Y.C., Rezaifar, R., Krishnaprasad, P.S.: Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In: 1993 Conference Record of the Twenty-Seventh Asilomar Conference on Signals, Systems and Computers, pp. 40–44. IEEE (1993)
  - 10. Vincent, E., Araki, S., Bofill, P.: The 2008 signal separation evaluation campaign: a community-based approach to large-scale evaluation. In: International Conference on Independent Component Analysis and Signal Separation, pp. 734–741. Springer (2009)
  - 11. Vincent, E., Araki, S., Theis, F., Nolte, G., Bofill, P., Sawada, H., Ozerov, A., Gowreesunker, V., Lutter, D., Duong, N.O.K.: The signal separation evaluation campaign (2007–2010): achievements and remaining challenges. *Sig. Process.* **92**(8), 1928–1936 (2012)
  - 12. Vincent, E., Gribonval, R., Févotte, C.: Performance measurement in blind audio source separation. *IEEE Trans. Audio Speech Lang. Process.* **14**(4), 1462–1469 (2006)

# Chapter 13

## A Comparison of Apache Spark Supervised Machine Learning Algorithms for DNA Splicing Site Prediction



Valerio Morfino · Salvatore Rampone and Emanuel Weitschek

**Abstract** Thanks to next-generation sequencing techniques, a very big amount of genomic data are available. Therefore, in the last years, biomedical databases are growing more and more. Analyzing this big amount of data with bioinformatics and big data techniques could lead to the discovery of new knowledge for the treatment of serious diseases. In this work, we deal with the splicing site prediction problem in DNA sequences by using supervised machine learning algorithms included in the MLlib library of Apache Spark, a fast and general engine for big data processing. We show the implementation details and the performance of those algorithms on two public available datasets adopting both local and cloud environments, emphasizing the importance of this last environment for its scalability and elasticity of use. We compare the performance of the algorithms with U-BRAIN, a general-purpose learning algorithm originally designed for the prediction of DNA splicing sites. Results show that, among the Spark algorithms, all have good prediction accuracy ( $>0.9$ )—that is comparable with the one of U-BRAIN—and much lower execution time. Therefore, we can state that Apache Spark machine learning algorithms are promising candidates for dealing with the DNA splicing site prediction problem.

### 13.1 Introduction

Thanks to next-generation sequencing (NGS), biological data have grown more and more and their availability in public repositories can lead to the discovery of new

---

V. Morfino · E. Weitschek

Department of Engineering, Uninettuno University, Rome, Italy  
e-mail: [valerio.morfino@ctcgroup.it](mailto:valerio.morfino@ctcgroup.it)

E. Weitschek  
e-mail: [emanuel.weitschek@uninettunouniversity.net](mailto:emanuel.weitschek@uninettunouniversity.net)

S. Rampone  
Department of Law, Economics, Management and Quantitative Methods (DEMM), Università degli Studi del Sannio, Benevento, Italy  
e-mail: [rampone@unisannio.it](mailto:rampone@unisannio.it)

knowledge hidden in the data. Big data technologies allow to compute analyses on large amounts of data. Especially machine learning techniques, both supervised and unsupervised, can be applied to a large variety of problems regarding genomics, genetics, and medical data. Predictive algorithms can be trained to recognize patterns in DNA sequences to identify promoters, enhancers, coding and noncoding regions, etc. [1]; gene expression data can be used to identify potentially disease biomarkers [1, 2].

In this work, we deal with the problem of finding splicing sites in DNA sequences by adopting a supervised machine learning approach. We analyze performance and accuracy of several general-purpose machine learning algorithms included in the MLLib library of Apache Spark [3], one of the most interesting and used technologies in the big data field, available with an open-source license and present in the cloud computing facilities of the main world players [4].

We use the BRAIN [5, 6] algorithm, specifically its uncertainty managing extension U-BRAIN [7, 8] and its HPC parallel implementation [9, 10] as a reference for comparisons. BRAIN (U-BRAIN) is a general-purpose learning algorithm originally designed for DNA analysis and extended to address data uncertainty. Given a set of binary instances that may have missing bits, it finds a Boolean formula in a disjunctive normal form (DNF), of roughly minimal complexity, which explains the data. The conjunctive terms of the formula are calculated in an iterative way identifying, from the given data, sets of conditions that must be satisfied by all the positive instances and violated by all the negative ones. These conditions allow the calculation of a probability distribution, driving the selection of the formula terms. The algorithm is characterized by a great versatility and effectiveness, highlighted in numerous applications [11–13], but the memory and the execution time required, respectively, of order  $O(n^3)$  and  $O(n^5)$ , appear unacceptable for huge datasets.

The paper is organized as follows: After a very brief biological background, we introduce Apache Spark and the MLLib Spark library for machine learning. Then, we describe the application of several machine learning algorithms of the MLLib library by using the Python programming language. In the last section, we show experimental results about computational time and accuracy parameters of the algorithms compared with U-BRAIN. In the “Conclusions and Future Works” section, we summarize the results, discuss the reasons that explain why the Spark implementation is much faster than the reference U-BRAIN implementation and the perspective to develop, in the near future, a Spark-based implementation of U-BRAIN.

### ***13.1.1 Brief Biological Background and the DNA Splicing Site Prediction Problem***

The expression of the genetic information stored in DNA involves the translation of a sequence of nucleotides in proteins. The flow is DNA → mRNA → protein. This concept is known as “The Central Dogma” [14]. A gene is a region of DNA

that controls a discrete hereditary characteristic. Most eukaryotic genes have their coding sequences, called exons, interrupted by noncoding sequences called introns. The role of the latter has been unknown for a long time; recent studies show that they have different functions both direct and indirect such as regulation of alternative splicing and positive regulation of gene expression [15]. The interruption points between coding and noncoding regions of a DNA sequence, exon–intron (EI) and intron–exon (IE) boundaries, are called donor and acceptor sites, respectively, and in general “splicing sites.” During the splicing process, introns are removed, meanwhile exons begin their path toward the encoding of proteins [14]. The DNA splicing site prediction problem deals with individuating those regions.

## 13.2 Methods and Description of the Experiments

### 13.2.1 *Apache Spark*

Apache Spark is a high-performance, general-purpose distributed computing system. It enables the process of large quantities of data, beyond what can fit on a single machine, with a high-level APIs, which are accessible in Java, Scala, Python, and R programming languages. It also supports a rich set of higher-level tools including Spark SQL for SQL and structured data processing, MLLib for machine learning, GraphX for graph processing, and Spark Streaming.

Spark allows users to write programs on a cluster computing system that can perform operations on very big amount of data in parallel. A large dataset is represented using a distributed data structure called RDD—Resilient, Distributed Dataset—which is stored in a distributed way in the executors (i.e., slave nodes). The objects that comprise RDDs are called partitions. They may (but not must) be computed on different nodes of a distributed system.

Spark evaluates RDDs lazily. Thus, RDD transformations are computed only when the final RDD data need to be computed. To speed up data access in case of repeated calculations, Spark can retain an RDD in the main memory for the whole application execution. RDDs are immutable: Transforming an RDD returns a new RDD, and the old one can be trashed by a garbage collector. The paradigms of lazy evaluation, in-memory storage, and immutability make Spark fault-tolerant, scalable, efficient, and easy to use [4]. In more detail, Spark is able to warrant resilience: When any node crashes in the middle of any operation, one other node has reference to the crashed one, thanks to a mechanism called lineage. In case of a crash, the cluster manager assigns the job to another node, which will operate on the particular partition of the RDD and will perform the operations that it has to execute without data loss [16].

To perform the experiments, we adopt the following algorithms from Apache Spark MLLib standard library: logistic regression (LR), decision tree (DT), random forest (RF), linear support vector machine (SVM), naïve Bayes (BAYES), and multilayer perceptron (MLPERC).

The algorithms have been chosen mainly because of their availability in MLlib, and also because some of them have been already compared with BRAIN (naïve Bayes) or have been already used with significant results in other bioinformatics classification analysis (decision tree, SVM, and random forest) [2].

We have collected training time, accuracy, and correlation (MCC) of each of them, and the values obtained have been compared with those related to the BRAIN algorithm. In particular, as a reference for accuracy and correlation we have used values from [5], for processing times the data available in [9], that describes the most recent parallel implementation of U-BRAIN.

### 13.2.2 Dataset Description

The test datasets used in [9] was the Irvine Primate splice-junction Dataset (IPDATA) [17], a subset of the Homo Sapiens Splice Sites Dataset (HS3D) [18] and a subset of the Catalogue of Somatic Mutations in Cancer (COSMIC) [19].

In this work, we use IPDATA and HS3D, and COSMIC gene p16 was discarded because of its low cardinality. The HS3D dataset was enriched to fit the training/test proportion of 70/30. IPDATA was used in the same way as [5, 9]. A brief description of the datasets is reported in Table 13.1.

IPDATA is a dataset of human splice sites, and it consists of 767 donor splice sites, 765 acceptor splice sites, and 1654 false splice sites. The authors of [9] considered 464 positive instances and 1536 negative, composed of sequences of 60 nucleotides (140 bits in input BRAIN encoding).

HS3D is a dataset of Homo Sapiens Exon, Intron and Splice sites extracted from GenBank Rel.123. It includes 2796 + 2880 donor and acceptor sites, windows of 140 nucleotides (560 bits in input BRAIN encoding) around a splice site, and 271,937 + 332,296 windows of false splice sites, selected by searching canonical GT-AG pairs in non-splicing positions. In the study, a subset of 161 donor sites and 2974 false ones was adopted.

**Table 13.1** Datasets used in the experiments

Dataset	#Nucleotides	Training instances (pos./neg.)	Test. instances (pos./neg.)	Total samples
IPDATA	60	464/1536	302/884	3186
HS3D_1	140	1960/2942	836/1307	7045
HS3D_2	140	1960/12571	836/5431	20,768

### 13.2.3 Dataset Encoding

All the algorithms of Apache Spark MLlib need numerical double precision values as input data format. So, when input data are categorical, in our case literal (A, C, G, T), we have to encode them. MLlib offers two standard methods for encoding: StringIndexer and OneHotEncoder [16]. StringIndexer encodes a string column of labels to a column of label indices, while one-hot encoding maps a column of label indices to a column of binary vectors, each containing all values “0” and a single value “1”. Both strategies create a different encoding for each feature column on which it is applied, because the index assignment is based on the frequency of each symbol in the rows of the feature. In order to have the same format of BRAIN, we prefer to write a custom encoder. To encode a possible unexpected character (therefore different from A C, G, T), we decide to adopt a sparse matrix with all “0” values. The encoding is given in Table 13.2.

Positive and negative instances were loaded from two different files. Training and test set creation were performed according to a random percentage split (70% training and 30% test) considering the percentage of positive and negative instances in the sets.

### 13.2.4 Execution Environment

The Python Spark script has been tested in two different environments:

- Local single node cluster (indicated below as Local 3-core)
  - CPUs: 3 COREs (Intel i7-6700 HQ 2.60 GHz)
  - RAM: 9.2 GB
  - Apache Spark 2.2.1, Scala 2.11
- Databricks Community Cloud [20] (indicated below as Databricks 1-core)
  - CPUs: 1 CORE
  - RAM: 6 GB
  - Apache Spark 2.2.1, Scala 2.11

**Table 13.2** Encoding of the dataset

Nucleotide	Encoded value stored as sparse matrix
A	{1,0,0,0}
C	{0,1,0,0}
G	{0,0,1,0}
T	{0,0,0,1}
Other values	{0,0,0,0}

### 13.3 Experimental Results

In Table 13.3, we report classification performance of the algorithms adopted in this study. For each algorithm, we show the following parameters [21, 22]:

Accuracy is defined as follows:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (13.1)$$

Error rate is defined as follows:

$$\text{ERR} = 1 - \text{ACC} \quad (13.2)$$

Correlation, also known as Matthews correlation coefficient (MCC), is defined as follows:

$$\text{MCC} = \frac{\text{TP} \times \text{TN} - \text{FP} \times \text{FN}}{\sqrt{(\text{TP} + \text{FP})(\text{TP} + \text{FN})(\text{TN} + \text{FP})(\text{TN} + \text{FN})}} \quad (13.3)$$

**Table 13.3** Classification performance of the adopted machine learning algorithms in the MLlib Spark Environment on the IPDATA, HS3D\_1 and HS3d\_2 datasets

Dataset	Algorithm	Accuracy	Error rate	Correlation
IPDATA	LR	0.948	0.052	0.865
IPDATA	DT	0.970	0.030	0.923
IPDATA	RF	0.965	0.035	0.906
IPDATA	SVM	0.960	0.040	0.894
IPDATA	BAYES	0.966	0.034	0.911
IPDATA	MLPERC	0.966	0.034	0.912
HS3D_1	LR	0.927	0.073	0.847
HS3D_1	DT	0.921	0.079	0.835
HS3D_1	RF	0.933	0.067	0.859
HS3D_1	SVM	0.935	0.065	0.864
HS3D_1	BAYES	0.861	0.139	0.706
HS3D_1	MLPERC	0.923	0.077	0.838
HS3D_2	LR	0.947	0.053	0.765
HS3D_2	DT	0.939	0.061	0.734
HS3D_2	RF	0.908	0.092	0.525
HS3D_2	SVM	0.949	0.051	0.776
HS3D_2	BAYES	0.902	0.098	0.614
HS3D_2	MLPERC	0.945	0.055	0.763

LR, DT, RF, SVM, MLPERC, and BAYES stand for linear regression, decision tree, random forest, linear support vector machine, multilayer perceptron, and naïve Bayes, respectively

For random forest, we used the following parameters:

- Number of trees: 100
- Max depth: 15

For multilayer perceptron, we used the following layers' configuration:

- For IPDATA: 240, 180, 50, 60, 2
- For HS3D: 560, 200, 150, 2

In both cases, the first layer is equal to the number of features (i.e., number of nucleotides multiplied by 4, since a 4-bit coding was used). The last layer is equal to the number of output classes, thus 2.

For all other classifiers, we used the default parameters as provided by MLlib, because the main aim of this work is to test the general features of MLlib library. In the case of random forest, we used different parameters, because the standard ones generated 0 occurrences for true positive (TP) for HS3D\_2 dataset, so a NaN value was generated for correlation value.

In Table 13.4, we show the training time of the algorithms.

**Table 13.4** Training time in seconds of the adopted machine learning algorithms in the MLlib Spark Environment on the IPDATA, HS3D\_1 and HS3D\_2 datasets

Dataset	Algorithm	Databricks 1-core	Local 3-core
IPDATA	LR	2.23	0.80
IPDATA	DT	1.48	0.66
IPDATA	RF	13.82	4.14
IPDATA	SVM	13.95	4.45
IPDATA	BAYES	0.75	0.16
IPDATA	MLPERC	49.39	9.87
HS3D_1	LR	6.68	1.56
HS3D_1	DT	3.83	1.37
HS3D_1	RF	43.20	14.15
HS3D_1	SVM	26.42	6.27
HS3D_1	BAYES	2.04	0.16
HS3D_1	MLPERC	91.73	44.31
HS3D_2	LR	6.20	1.53
HS3D_2	DT	5.32	2.51
HS3D_2	RF	67.02	25.40
HS3D_2	SVM	26.63	7.83
HS3D_2	BAYES	2.03	0.17
HS3D_2	MLPERC	157.37	156.76

LR, DT, RF, SVM, MLPERC, and BAYES stand for linear regression, decision tree, random forest, linear support vector machine, multilayer perceptron, and naïve Bayes, respectively

### 13.3.1 Comparison with the Performance of the U-BRAIN Algorithm

Classification metrics are the same used for the performance evaluation of U-BRAIN in [5]. For the IPDATA, BRAIN obtained the following performance results:

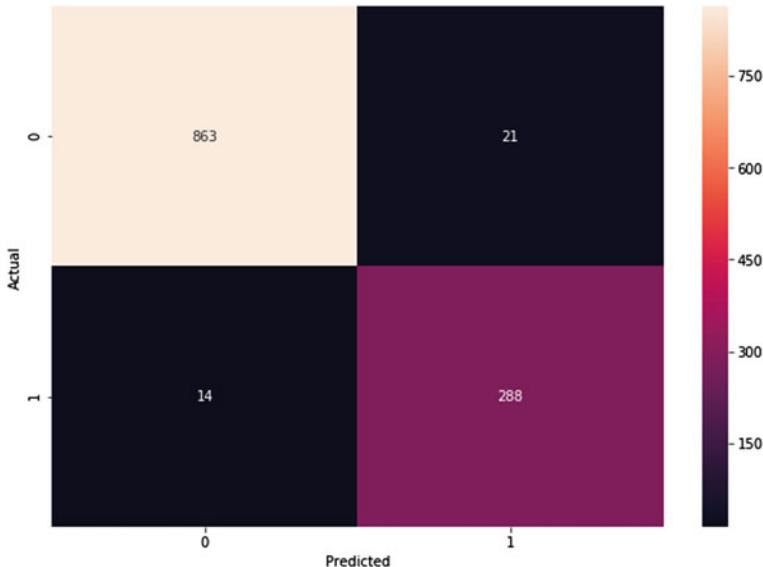
- Accuracy = 0.966
- Error rate = 0.034
- Correlation = 0.91

By optimizing the results on a restricted window of 20 nucleotides and by applying a neural network hybrid approach using a tenfold cross-validation approach, the final results described in [5] were as follows:

- Accuracy = 0.974
- Error rate = 0.026
- Correlation = 0.93

BRAIN also produced an explicit classification formula, likewise random forest and decision tree.

The best classification performance among the tested algorithms was found for the decision tree applied on IPDATA. In Fig. 13.1, we show the related confusion matrix. The total error was 35 (21 FP + 14 FN) with a correlation of 0.923.



**Fig. 13.1** Confusion matrix dataset: IPDATA; algorithm: decision tree

When analyzing the results on other datasets HS3D\_1 and HS3D\_2, we can confirm that classification performances decrease when the observation window increases, as already observed in [5].

Furthermore, there seems to be a negative relationship among the imbalance of positive and negative instances and classification performance, as we can see from the analysis of the same algorithm applied on the HS3D\_2 dataset, which has a greater number of negative instances and a window of 140 nucleotides.

Regarding training time performance, we refer to the parallel implementation of U-BRAIN [9]. On IPDATA, the training time using 3 CPUs is about 400 s. For the same dataset and the same three-core environment, the slowest of the algorithms of Spark MLlib (SVM) performs in 4.45 s. The fastest one (naïve Bayes) performs in 0.16 s. For HS3D, the training time of U-BRAIN using 3 CPUs is over 2000s when dealing with a dataset containing 161 positive instances and 2794 false instances. The slowest of the algorithms of MLlib (SVM) performs in 7.83 s. on a dataset containing 1960 positive instances and 12,571 negative ones. Comparing training time of the executed experiments (summarized in Table 13.4), we can observe a reduction of the processing time when increasing the number of cores, according to the scalability promised by Spark. A decrease of execution time can also be observed for U-BRAIN, although processing times are generally much higher.

## 13.4 Conclusions and Future Works

In this work, we have tested the supervised machine learning algorithms of Apache Spark MLlib library by applying them to splicing site recognition in DNA sequences. Accuracy and correlation were compared with the BRAIN algorithm, while processing times were compared with the U-BRAIN parallel implementation. The analysis of results shows that all the algorithms included in MLlib have performed much better than U-BRAIN regarding training time. At the same time, U-BRAIN confirmed better accuracy and correlation. In addition, U-BRAIN generates an explicit formula containing the inferred rules [5, 9]. These elements encourage the research for a more performing U-BRAIN implementation, suitable to work on large amounts of data.

In a future work, we plan to produce an implementation of U-BRAIN, which will be further optimized for parallel processing on cluster and cloud computing environments and will be based on Apache Spark. From the analysis of the mathematical formulation of U-BRAIN and of the parallel implementation choices, we can expect an important speedup by using such new implementation. In more depth, U-BRAIN shares data among the parallel tasks using a mechanism based on file systems, similar to MapReduce [9, 23]. Apache Spark will lead to possible speedups of 10X on disk and of 100X in memory [3, 24] when compared to MapReduce. Furthermore, the scheduling of the parallel tasks on the different processors in U-BRAIN is carried out in a static way without using a dynamic scheduler or a load balancer [9]. Spark can use different optimized schedulers, such as Yarn and Mesos [3, 24]. Another element that leaves hope for significant increases in performance is the “lazy evaluation”

mechanism of Spark: The execution plan is assessed after all the operations are executed. Therefore, Spark can optimize the whole execution considering all the tasks and the available resources, both memory and cores, in cooperation with the resource manager [4, 16]. A further aspect that can be improved, not concerning performance, but yet important for execution in a distributed environment, especially with a big amount of data, is that the current implementation of U-BRAIN does not support resiliency. So, if any of the processing nodes fails, the entire processing must be restarted without safeguarding the partial computed data. The Apache Spark RDDs, instead, automatically recover from the failures by using the “lineage” approach, an efficient way to reconstruct a lost partition based on the execution of the previously run transformations that have been logged on each RDD. One last aspect that we wish to mention is the possible opening toward the “streaming” technology offered by Spark. This technology allows to explore scenarios of great interest related to the use of the algorithm with data flows.

## References

1. Maxwell, W.L., Noble, W.S.: Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* **16**(6), 321 (2015)
2. Weitschek, E., Fiscon, G., Fustaino, V., Felici, G., Bertolazzi, P.: Clustering and classification techniques for gene expression profile pattern analysis. In: *Pattern Recognition in Computational Molecular Biology: Techniques and Approaches*, p. 347 (2015)
3. Apache Spark Home page. <http://spark.apache.org/>. Last accessed 10 April 2018
4. Zaharia, M., et al.: Apache spark: a unified engine for big data processing. *Commun. ACM* **59**(11), 56–65 (2016)
5. Rampone, S.: Recognition of splice junctions on DNA sequences by BRAIN learning algorithm. *Bioinformatics (Oxford, England)* **14**(8), 676–684 (1998)
6. Morfino, V., Rampone, S.: Metodi ed architetture per la creazione di applicazioni multicanale per la bioinformatica. In: Ceccarelli, M., Colantuoni, V., Graziano, G., Rampone, S. (eds.) *Bioinformatica. Sfide e prospettive*. Edizioni Franco Angeli (2007)
7. Rampone, S., Russo, C.: A fuzzified brain algorithm for learning DNF from incomplete data. *Electron. J. Appl. Statistical Anal. (EJASA)* **5**(2), 256–270 (2012)
8. Rampone, S.: An error tolerant software equipment for human DNA characterization. *IEEE Trans. Nucl. Sci.* **51**(5), 2018–2026 (2004)
9. D’Angelo, G., Rampone, S.: Towards a HPC-oriented parallel implementation of a learning algorithm for bioinformatics applications. *BMC Bioinform.* **15**(5), S2 (2014)
10. Aloisio, A., Izzo, V., Rampone, S.: FPGA implementation of a greedy algorithm for set covering, In: *14TH IEEE-NPSS Real Time Conference*, IEEE (2005)
11. D’Angelo, G., Palmieri, F., Ficco, M., Rampone, S.: An uncertainty-managing batch relevance-based approach to network anomaly detection. *Appl. Soft Comput. J.* **35**, 408–418 (2015)
12. D’Angelo, G., Rampone, S.: Diagnosis of aerospace structure defects by a HPC implemented soft computing algorithm. In: *IEEE Metrology for Aerospace (MetroAeroSpace)*, pp. 408–412. IEEE (2014)
13. D’Angelo, G., Rampone, S.: Feature extraction and soft computing methods for aerospace structure defect classification. *Meas. J. Int. Meas. Confederation* **85**, 192–209 (2016)
14. Kimmel, G., Farkash, A.: Lecturer Ron Shamir, “Algorithms for Molecular Biology”, Lecture 1: 25 Oct 2001, Fall Semester, Tel Aviv University (2001)
15. Jo, Bong-Seok, Choi, Sun Shim: Introns: the functional benefits of introns in genomes. *Genomics Informatics* **13**(4), 112–118 (2015)

16. Karau, H., Warren, R.: High Performance Spark: Best Practices for Scaling and Optimizing Apache Spark. O'Reilly Media, Inc. (2017)
17. Bache, K., Lichman, M: UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine, CA (2013). <http://archive.ics.uci.edu/ml>. Last accessed 10 April 2018
18. Pollastro, P., Rampone, S.: HS3D, a dataset of *Homo sapiens* splice regions, and its extraction procedure from a major public database. *Int. J. Mod. Phys. C* **13**(8), 1105–1117 (2003)
19. Forbes, S.A.: COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **39**(suppl 1), D945–D950 (2011)
20. Databricks Home page. <https://databricks.com/>. Last accessed 10 April 2018
21. Kennedy, J.: Encyclopedia of Machine Learning. Springer, US (2011)
22. Cestarelli, V., Fiscon, G., Felici, G., Bertolazzi, P., Weitschek, E.: CAMUR: Knowledge extraction from RNA-seq cancer data through equivalent classification rules. *Bioinformatics* **32**(5), 697–704 (2016)
23. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. *Commun. ACM* **51**(1), 107–113 (2008)
24. Celli, F., Cumbo, F., Weitschek, E.: Classification of large DNA methylation datasets for identifying cancer drivers. *Big Data Res.* **13**, 21–28 (2018)

# Chapter 14

## Recurrent ANNs for Failure Predictions on Large Datasets of Italian SMEs



Leonardo Nadali, Marco Corazza, Francesca Parpinel and Claudio Pizzi

**Abstract** The prediction of failure of a firm is a challenging topic in business research. In this paper, we consider a machine learning approach to detect the state of asset shortfall in the Italian small and medium-sized enterprises' context. More precisely, we use the recurrent neural networks to predict the insolvency of firms. The huge dataset we study allows us to overcome problems of distortions given by smaller sample sizes. The observed sample comes from AIDA database, and consider thirty variables replicated for five years. The main result is that recurrent neural networks outperform the multi-layer perceptron architecture used as benchmark. The obtained accuracy scores are in line with those found in the literature, and this suggests that the use of new techniques such as those tried out in this study could produce even better results.

### 14.1 Introduction

The purpose of this work is to understand whether a machine learning technique, known as recurrent neural network (RNN), is able to successfully predict the failures in a large dataset of Italian small- and medium-sized enterprises (SMEs) considering their account sheets. The European Commission defines SMEs as firms with a staff

---

L. Nadali  
Independent, Milano, Italy  
e-mail: [leonardo.nadali@outlook.com](mailto:leonardo.nadali@outlook.com)

M. Corazza · F. Parpinel (✉) · C. Pizzi  
Department of Economics, Ca' Foscari University of Venice,  
Cannaregio 873, 30121 Venezia, Italy  
e-mail: [parpinel@unive.it](mailto:parpinel@unive.it)

M. Corazza  
e-mail: [corazza@unive.it](mailto:corazza@unive.it)

C. Pizzi  
e-mail: [pizzic@unive.it](mailto:pizzic@unive.it)

headcount of less than 250 and a total turnover of less than 50 million of euros or a balance sheet of less than 43 million of euros.

It can be harder to model the risk of SMEs than that of larger firms, as their data on stock and bond prices are generally unavailable. We must stress that their most relevant variables are also those not captured in the accounts (such as the ability and social network of the management) [1]. So, if we want to predict the failure on the base only of the accounts sheets, the resulting models indirectly set an upper boundary to the importance of *out of the accounts* variables in determining the financial stability of a firm.

Remind that, among all sources of risk, credit is one that can have the largest impact on a corporate level. Being able to reduce it by a suitable tool could dramatically raise the profitability of a firm and liberate capitals for other purposes. With the approval of the Basel III regulations, this is one of the main topics. Furthermore, all the businesses and financial institutions whose business model depends strictly on the outcome of a loan (primarily banks but also insurance companies) have of course a high interest in predicting insolvencies. Both the ability to avoid bad borrowers and the losses that they create and the possibility to identify good customers can significantly improve the odds of these firms to outrank the competition and have a higher profitability. At last, a tool that can analyze the account sheets of a firm and assess its financial stability would be of huge help to that firm. By performing some self-analysis, firms can become aware of risky situations before the failure becomes unavoidable.

To summarize, this work explores the improvements with respect to the quality of the results that can be achieved by the use of RNNs in problems of definition and ranking credit risk firms. This is particularly important as, to the best of our knowledge, there are no accounts in the literature of the use of RNNs for the prediction of the failures of firms: even the studies that specifically focus on the use of artificial neural networks (ANNs) to solve this problem, in most cases, use the multi-layer perceptron (MLP) architecture and variations of it [8, 13].

The other topic faced in this work regards the overall levels of accuracy that this model can achieve on a large dataset of firms. It can be interesting to check whether the results of previous studies may be repeated on datasets of higher sizes: the largest datasets found in the literature contain only some thousands of firms [4]. The dataset we use considers hundreds of thousands of firms, avoiding problems of distortions given by smaller sample sizes.

## 14.2 RNNs Versus MLPs

In order to understand the improvements with respect to using RNN, we compare it with the MLP architecture, whose main component is the perceptron, called also as *artificial neuron* [17]. The perceptron is an operator that takes some inputs and through a particular nonlinear function (see further) outputs a result. A perceptron is described by the formula:

$$f(X) = \theta \left( \sum_{i=1} (w_i x_i + b) \right)$$

where  $x_i$  are the components of the input vector  $X$ ,  $w_i$  and  $b$  are parameters, respectively, called *weights*, and *bias*,  $\theta(\cdot)$  is the *activation function*. By stacking layers of perceptrons over the inputs and over each other, a structure called MLP can be obtained. The main underlying idea is that when a signal is passed from one layer to the next, the information contained in it is analyzed and summarized to a higher degree of abstraction, until a final result is obtained. The most popular technique used to calculate the parameters of the perceptrons in an MLP is called *backpropagation* [14], and has proven to be extremely effective in a wide variety of tasks. However, this model cannot be directly applied when the data are structured in a series format like, for instance, the financial state of a firm over some years. This means that the adjustment to achieve the final result may undermine the quality of the obtained predictions.

The RNNs were introduced to overcome these problems, [12], by allowing that the state of each neuron is determined both by the values they receive from the current input and by the state computed from previous input. This means that with the time series in the last step, the neurons are still able to look both at the present input and at the state coming from the preceding time steps, holding the whole information. This structure can produce good results in a wide variety of tasks and can be formalized with

$$h_t = \theta (W^{(hx)} x_t + W^{(hh)} h_{t-1} + b_h)$$

where  $h_t$  is the state of a layer of recurrent neurons at time  $t$ ,  $W^{(hx)}$  is the weight between the input  $x$  and the hidden layer  $h$ , and  $W^{(hh)}$  is the recurrent weight for the hidden layer  $h$  at adjacent time steps [12].

In the literature, we find some relevant applications of ANNs to failure prediction [8, 13]. Boritz and Kennedy [4], compare traditional techniques with ANNs. They use a dataset composed of 171 failed firms and 6153 *healthy* firms. Different machine learning approaches were applied including some MLP techniques, showing that MLPs provide better results than the other approaches, with Type I and Type II errors in the best network being, respectively, 20.82 and 14.99%.

The work by Altman and Sabato [1], focuses specifically on SMEs even without employing ANNs. The authors gathered data from the financial statements of 2010 US SMEs and used a logistic regression over the log transformation of five key financial ratios. They achieve an overall accuracy of 80.16%, with Type I and Type II errors, respectively, 11.76 and 27.92%. Brédart [3], uses an MLP to predict bankruptcies in a dataset of 3728 Belgian SMEs. Their MLP uses three input variables: an index of liquidity, one of solvency and one of profitability. They obtain a Type I accuracy of 81.50% and a Type II accuracy of 69.90%. Le and Viviani [11], use a dataset of the financial statements for the last 6 years of 3000 USA banks, from which 31 financial ratios were downloaded. They compare MLP technique with other two

**Table 14.1** Frequency distribution of *legal status*

Legal status	Frequency	Legal status	Frequency
Active	654,065	Dissolved (merger)	14,249
In liquidation	81,186	Active (default of payments)	4278
Bankruptcy	47,071	Dissolved (bankruptcy)	404
Dissolved	37,866	Dissolved (demerger)	251
Dissolved (liquidation)	32,376	Active (receivership)	212

statistical techniques ( $K$ -nearest neighbors and support vector machines), showing that 75.7% of firms predicted to be failed actually failed.

It is noteworthy to underline that the various clarifications listed above are not always comparable with each other and with those determined by us. Generally, they refer to different datasets: *training* or *validation* or *test* set in case of ANN-based analysis, the whole dataset in the case of regression-like analysis.

### 14.3 The Dataset

The observed sample comes from AIDA database, which collects financial, commercial, and anographic information of Italian companies. In particular, it holds the digitalized balance sheets of more than one million firms and their legal status that is whether they are still active, bankrupted, or else. Our approach requires five years of available accounts, in order to ensure the training of an RNN model, satisfying, in each year, the definition criteria of SMEs as provided by the European Commission.

With these requirements, the initial selected firms were  $N = 872,558$  with forty-five variables replicated for five years. As many missing values (NAs) were present, we deleted variables and firms with too many NAs. The remaining ones were replaced with the average value of each variable. After the cleaning and the arrangement, the dataset was composed by  $n = 722,160$  firms over thirty variables for five years.

To build the dependent variable of the model, which is a binary variable indicating 1 for a failed firm and 0 for a healthy firm, the possible values of *legal status* and *procedure/cessazione*<sup>1</sup> were considered. Table 14.1 shows the distribution of *legal status*, but the complete set of levels for *procedure/cessazione* contains 60 elements. Every firm whose value in the *legal status* variable is either *bankruptcy*, *dissolved (bankruptcy)*, *active (receivership)*, or *Active (default of payments)* is marked as failed, while the classification of the other not *Active* firms will be decided based on the value of *procedure/cessazione* which better specifies whether the legal status of the firm can be considered a failure or not. The values of *procedure/cessazione* are converted to indicate a failure or not. By this criterion, the number of failures in the dataset amounts to 137,938 firms, which constitutes 19.1% of the total.

---

<sup>1</sup>*procedure/cessazione* indicates that a pending procedures concerning the change of *legal status* for the considered firm is in place.

## 14.4 The Application

ANNs are here applied with two goals. First, the MLP architecture is used considering it as a benchmark and comparing the obtained results to those found in other works which use a similar model. Furthermore, it is also the basis on which to compute the improvements obtained by the use of RNNs. Initially, a set of MLP experiments are run using only the last year of accounts for each firm, in order to check the predictive power of the models. Then, a new set of experiments is done with the full dataset.

In the second phase, the RNN architecture is used. The input of the RNNs is always constituted of the accounts of five years collected for the firms, each year constituting one *step* of the RNN. As the results from the first set of RNN experiments show a large imbalance between the error on the failed firms and that on the healthy firms, the experiments are repeated also on a rebalanced dataset, created by randomly selecting a set of 137,938 non-failed firms (to match the number of failed ones) and discarding all the others, thus reducing the total sample size to 275,876 firms.

### 14.4.1 ANN Architectures

To define the ANNs, after the choice of the particular model to use (MLP or RNN), it is necessary to set an appropriate framework for it. In this work, the MLPs always use two hidden layers, in three different configurations (50 neurons in the first layer and 50 in the second one, 100 and 25, respectively, and finally 100 and 100), while the RNNs use one single layer of long short-term memory cells [7], in four different sizes: 20, 50, 100, or 200 neurons.

The initial location of weights and biases is chosen by extracting values from a standard normal distribution. This choices should help with speeding up of the training phase [10].

Here, we use as activation function the *rectified linear unit* as, in general, it provides the best results with respect to other activation functions when used for ANNs [5]. The output of all the implemented ANNs is always constituted by two neurons, one for each class (*failed* or *healthy*). The loss function is the cross-entropy function. In the RNN model, L2 regularization [15], is added to the loss function to avoid overfitting. The  $\beta$  value used for regularization is set to 0.01 in each model. Also, the *dropout* technique is used to reduce overfitting and improve the overall results [18], by randomly deactivating some neurons during training (both the MLPs and the RNNs use dropout with a keep value of 0.5). Furthermore, we chose the Adam optimizer [2], with an initial value of the learning rate set to 0.0001 for each MLP experiment and four different values for the RNN experiments (0.01, 0.001, 0.0001).

### 14.4.2 Performances

Each MLP experiment was run for a total of 2000 epochs; 50% of the dataset (randomly selected at the beginning of each session) was used for training the network. The remaining 50% of the dataset was splitted in two parts for forming the validation set and the test one; the test set may be used to correct some procedural biases [6].

The RNN experiments were run for a total of 200 epochs, instead of 2000 as for the MLP ANNs, because we want to check the robustness of the RNN model. Also in these experiments, 50% of the data were used to train the networks and the remaining was split into validation and test sets. All the code used in this study was written in Python 3. The code makes use of the TensorFlow library and particularly the possibility to train the ANNs on GPUs through the NVIDIA cuDNN toolkit. The machine used to run the experiments had a 8 cores 2.50GHz CPU, GeForce GTX 850 M GPU, 4GB of RAM, and 2GB of dedicated GPU memory; 2000 epochs took about two hours for the MLPs with the single-year input, and about five hours for the ones using the full input size. To iterate through 200 epochs of a RNN network, half to one hour was required, depending on the size of the hidden layer.

For the MLP experiment with single-year input (Phase 1), twelve training sessions were run: With every new training session, the dataset was randomly scrambled, and new validation and test sets were extracted. Three configuration of the hidden layers were replicated four times. The MLP experiment with the full-sized input (Phase 2) was repeated nine times, three times for each configuration of the hidden layers. The first set of RNN experiments, which used the *unbalanced* dataset (Phase 3), was repeated with two different sizes of the hidden layer (50 and 100 neurons), checking two learning rates for each size (0.01 and 0.001), twice for each learning rate, for a total of eight sessions. The second and most important set of RNN sessions (Phase 4) used a rebalanced dataset. This means that before every training session, the dataset was divided into two groups: the *failed* firms and the *healthy* firms. The *healthy* firms were sampled with the same size as the failed ones, and the two groups were then merged in a new *rebalanced* dataset that was then scrambled again to produce the training, validation, and test sets. Each combination of four hidden layer structure and three learning rate was trained five times. This means that a total of 60 experiments were carried out for Phase 4.

## 14.5 Results

Table 14.2 shows the results obtained by the MLP in Phase 1. In particular, column 7 reports the range of variation of the Cohen's kappa statistic ( $k$ ) and columns 8 and 9 provide the ranges of variation of the  $F_1$  score calculated over the datasets of the healthy firms ( $F_{1,h}$ ) and of the failed firms ( $F_{1,f}$ ), respectively.<sup>2</sup> The networks with

---

<sup>2</sup>Generally speaking, Cohen's kappa and  $F_1$  score are both known measures of classification accuracy. The former indicates substantial agreement between the considered model and the investigated phenomenon for values greater than 0.80 [9], closer the latter to the value 1 higher is the accuracy [16].

**Table 14.2** Phase 1: MLP, one year (values in %), sample size  $n = 722, 160$ 

Hidden layers structure	Run	Training accuracy (%)	Validation accuracy (%)	Type I (%)	Type II (%)	$k$	$F_{1,h}$	$F_{1,f}$
(50; 50)	$r_1$	83.06	83.24	74.20	3.20	0.298	0.903	0.374
	$r_2$	83.00	83.36	73.67	3.18	0.303	0.904	0.380
	$r_3$	82.99	83.34	73.54	3.24	0.304	0.904	0.381
	$r_4$	82.90	83.37	73.39	3.23	0.302	0.904	0.378
(100; 25)	$r_1$	83.27	83.40	73.74	3.11	0.305	0.904	0.380
	$r_2$	83.08	83.50	73.26	3.10	0.310	0.905	0.385
	$r_3$	83.17	83.48	73.24	3.13	0.310	0.905	0.386
	$r_4$	83.05	83.54	72.98	3.12	0.309	0.905	0.385
(100; 100)	$r_1$	83.16	83.43	73.00	3.25	0.307	0.904	0.384
	$r_2$	83.56	83.37	73.88	3.12	0.303	0.904	0.379
	$r_3$	83.36	83.53	72.12	3.34	0.315	0.908	0.392
	$r_4$	83.40	83.43	72.41	3.39	0.314	0.904	0.392

**Table 14.3** Phase 2: MLP, full-sized input (values in %), sample size  $n = 722, 160$ 

Hidden layers structure	Run	Training accuracy (%)	Validation accuracy (%)	Type I (%)	Type II (%)	$k$	$F_{1,h}$	$F_{1,f}$
(50; 50)	$r_1$	87.80	87.66	46.36	4.31	0.553	0.926	0.624
	$r_2$	87.60	87.35	47.59	4.40	0.540	0.924	0.613
	$r_3$	87.60	87.32	47.59	4.40	0.535	0.923	0.613
(100; 25)	$r_1$	88.10	87.50	47.02	4.35	0.546	0.925	0.617
	$r_2$	88.10	87.50	48.14	4.12	0.542	0.925	0.614
	$r_3$	88.10	87.49	47.72	4.20	0.545	0.925	0.617
(100; 100)	$r_1$	88.40	86.90	52.98	3.66	0.508	0.922	0.581
	$r_2$	88.50	87.88	46.64	3.97	0.558	0.928	0.627
	$r_3$	88.60	87.41	48.01	4.23	0.540	0.925	0.612

100 neurons on the first hidden layer slightly outperformed the others, both in terms of the mean accuracy and of the best overall result. The best validation accuracy achieved by this architecture is 83.54%, while in general, the validation accuracy in this phase is about 83.50%. The values of  $k$ ,  $F_{1,h}$ , and  $F_{1,f}$  confirm such results.

The results from Phase 2 are presented in Table 14.3. Such results show again that the best performing networks are the ones with the largest first hidden layer. The validation accuracies improve enough: The best one is 87.88%, and the general ones are at least 87.50%.

Let us observe that the MLP adopts a strategy that is very similar to *always guess not-failed*. As Table 14.2 highlights, there is an important problem in these results: The error rate on the healthy firms is very low as generally the MLPs predict everything as not-failed, while the error on the failed is generally greater than seven-tenths of the total.<sup>3</sup> Even though such a result is a sizeable improvement over the 100% error rate expected from that simple strategy, this is still a very poor performance, especially in the field of credit risk, where any single Type I error has an economic impact that is generally much larger than Type II errors. The values of  $k$ ,  $F_{1,h}$ , and  $F_{1,f}$  confirm this analysis. We deal with such a problem in the following phases.

Table 14.4 reports the best of the two results per each combination of hidden layer size and learning rate of Phase 3. These results show some improvement over the MLP model, mainly with respect to the Type I error, reduced by almost 40%. The general accuracy on the healthy firms is more or less the same, but now most of failed firms are classified correctly. This is not yet a satisfactory result to be used in any real-world decision making, but it highlights that just by applying the RNN model instead of the MLP one, without changing the organization of the data or the way it

---

<sup>3</sup>Type I error is the share of firms incorrectly classified among all the firms that failed, while Type II error is the number of healthy firms that were misclassified (based on the networks trained in Phase 1).

**Table 14.4** Phase 3: Accuracy for hidden layer size and learning rate (values in %), sample size  $n = 722, 160$ 

Neurons	Learning rate	Training accuracy (%)	Validation accuracy (%)	Type I (%)	Type II (%)	$k$	$F_{1,h}$	$F_{1,f}$
(50)	0.01	86.90	87.36	44.35	5.15	0.552	0.924	0.626
	0.001	87.60	87.54	44.49	4.90	0.556	0.925	0.629
(100)	0.01	86.80	87.39	47.26	4.43	0.541	0.925	0.614
	0.001	87.58	87.47	45.49	4.75	0.550	0.925	0.623

**Table 14.5** Phase 4: Accuracy for hidden layer size and learning rate (values in %), sample size  $n = 275, 876$ 

Neurons	Learning rate	Validation accuracy (%)	Type I (%)	Type II (%)	$k$	$F_{1,h}$	$F_{1,f}$
(20)	0.01	80.85	7.72	7.60	0.617	0.808	0.809
	0.001	81.54	7.41	7.36	0.631	0.815	0.816
	0.0001	81.04	7.73	7.44	0.621	0.810	0.810
(50)	0.01	80.69	8.75	6.70	0.614	0.811	0.802
	0.001	81.93	7.20	7.26	0.639	0.819	0.820
	0.0001	81.38	7.67	7.23	0.627	0.814	0.813
(100)	0.01	80.49	7.89	7.72	0.610	0.805	0.805
	0.001	82.02	6.88	7.50	0.640	0.819	0.822
	0.0001	81.45	7.39	7.45	0.629	0.814	0.815
(200)	0.01	80.18	8.08	7.78	0.604	0.802	0.801
	0.001	81.97	7.91	6.51	0.640	0.823	0.817
	0.0001	81.63	7.73	6.96	0.633	0.819	0.814

is presented to the network, the results are much better. We point out that the RNNs achieve such performances running for only 200 epochs.

The results from Phase 4 are presented in Table 14.5. The RNN with the rebalanced dataset sacrifices Type II and  $F_{1,f}$  accuracies. This is exactly the target that this experiment was aiming for: as the economic impact of a company that fails when it was granted credit is much larger than the impact of not granting credit to a company that would have repaid it, in a credit risk model, the preference is strong for a better accuracy on the former kind of mistakes. The overall accuracy in this case is lower than that of the MLP experiments, but this is misleading: The actual quality of the predictions has improved a lot, especially because these data prove that the RNN network has really built a model to interpret the accounts and does not rely on the simple strategy of always giving the same answer except for some easy cases. The values of  $k$  and  $F_{1,h}$  confirm this analysis.

These experiments also show that the ratio between Type I and Type II seems to depend strongly on the ratio between failed and healthy entries in the dataset, and this is very important as it means that it is possible to fine-tune the type of error based on the profile of the final user of the network predictions: If the user has a strong aversion toward the risk of granting credit, she/he can use an ANN trained on a high percentage of failed firms, while other users with different needs can adjust this percentage accordingly.

One more important result of this last set of experiments is a better understanding of dependence of the achieved accuracies on the parameters. Some trends can be identified: For every configuration of the hidden layer, the 0.01 learning rates performed much better, being the best in terms of general accuracy achieved on the overall dataset. The hidden layer configurations that seem to work better in almost every instance are the ones with 100 and 200 neurons. These results highlight the fact that a right balance must be found between simpler networks with a lower number of neurons that can be trained in an easier way but may not be large enough to model complex relations and larger networks that can accommodate complex ideas but are much harder to train.

## 14.6 Conclusions

The MLP architecture did not prove to be very effective, as it adopted a decision rule that was very similar to the trivial strategy of always guessing firms to be healthy. So, it was necessary to adapt the model and the dataset to the problem at hand, and this was done in two ways. First, a different and more complex learning model was used; secondly, the dataset was rebalanced through a random process. The first of these two methods implied an RNN. This model, which is a novelty in the field of ANNs for credit risk evaluation, proved to be effective and increased the general accuracy score significantly. The second method involved the reduction of the number of healthy firms in the dataset to eliminate class imbalance. The large number of data collected granted the possibility to operate this reduction without sacrificing the explanatory power of the model. This technique produced the results that were expected, making the ANN almost equally sensitive to Type I and Type II errors. The final general accuracy values can be considered satisfactory, especially in light of the strict limitations that were imposed on the dataset: The fact that the study focused on SMEs, the deletion of many features and data based on missing values, and other choices all contributed to impose conditions that likely made it much harder for the ANNs to make good predictions. Nevertheless, the final general accuracy scores are comparable with those found in literature, and this suggests that the use of new techniques like the ones tried out in this study could produce even better results in the future.

## References

1. Altman E.I., Sabato G.: Modelling Credit Risk for SMEs: Evidence from the U.S. Market. *Abacus* **43**(3) 332–357 (2007)
2. Ba J.L., Kingma D.P.: ADAM: a method for stochastic optimization, arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2015)
3. Brédart, X.: Bankruptcy prediction model using neural networks. *Account. Financ. Res.* **3**(2), 124–128 (2014)
4. Boritz, J.E., Kennedy, D.B.: Effectiveness of neural network types for prediction of business failure. *Expert. Syst. Appl.* **9**(4), 503–512 (1995)
5. Glorot, X., Bordes, A., Bengio, Y.: Deep sparse rectifier neural networks. In: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS) (2011)
6. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning, 2nd edn. Springer (2009)
7. Hochreiter, S., Schmidhuber, J.: Long short term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
8. Kumar, P.R., Ravi, V.: Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review. *Eur. J. Oper. Res.* **180**(1), 1–28 (2007)
9. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. *Biometrics* **33**(1), 159–174 (1977)
10. LeCun, Y., Bottou, L., Orr, G.B., Müller, K.-R.: Efficient BackProp. In: Montavon, G., Orr, G.B., Müller, K.-R. (eds.) *Neural networks: Tricks of the trade*, pp. 9–50. Springer (1998)
11. Le, H.H., Viviani, J.L.: Predicting bank failure: an improvement by implementing a machine-learning approach to classical financial ratios, *Research in International Business and Finance*, **44**(C), 16–25 (2018)
12. Lipton, Z.C., Berkowitz, J., Elkan, C.: A critical review of recurrent neural networks for sequence learning, arXiv preprint [arXiv:1506.00019](https://arxiv.org/abs/1506.00019) (2015)
13. Louzada, F., Ara, A., Fernandes, G.B.: Classification methods applied to credit scoring: systematic review and overall comparison. *Surv. Oper. Res. Manag. Sci.* **21**(2), 117–134 (2016)
14. Mitchell, T.M.: *Machine Learning*. McGraw-Hill (1997)
15. Ng A.Y.: Feature selection, L1 versus L2 regularization, and rotational invariance. In: *Proceedings of the 21st International conference on Machine Learning* (2004)
16. Powers, D.M.W.: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation. *J. Mach. Learn. Technol.* **2**(1), 37–63 (2011)
17. Rosenblatt F.: The Perceptron: a perceiving and recognizing automaton, Report 85-460-1, Cornell Aeronautical Laboratory Inc. (1957)
18. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent Neural Networks from overfitting. *J. Mach. Learn. Res.*, **15**(Jun), 1929–1958 (2014)

# Chapter 15

## Inverse Classification for Military Decision Support Systems



Pietro Russo and Massimo Panella

**Abstract** We propose in this paper a military application, which can be used in civil contexts as well, for solving inverse classification problems. Pattern recognition and decision support systems are typical tools through which inverse classification problems can be solved in order to achieve the desired goals. As standard classifiers do not work properly for inverse classification, which is an inherent ill-posed problem and therefore difficult to be inverted, we propose a new approach that exploits all the information associated with the decisions observed in the past. The experimental results prove the feasibility of the proposed algorithm, with errors lower than 10% with respect to standard classification models.

### 15.1 Introduction

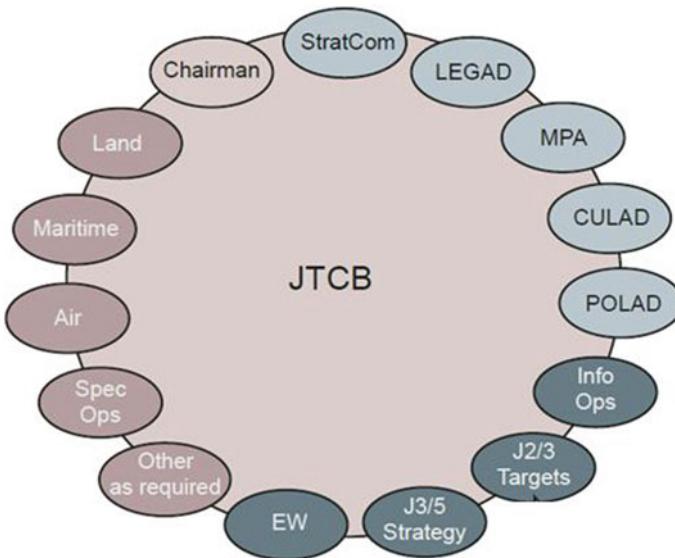
The aim of this paper is to propose a new algorithm for making a prediction about functions that a decision maker might use in order to take a correct decision in a military scenario [1, 2]. It implements a pattern recognition technique that, for instance, is widely adopted also in the medical world [3, 4], as well as in other civil scenarios. A typical military scenario where the proposed classification system can be applied is shown in Fig. 15.1.

Starting from the user name (identified in the system with a user\_id) and a specific situation (e.g., an enemy attack on friendly forces or the possibility to acquire a new flight route), the system is able to predict what functions the decision maker should use. The general structure of the application is shown in Fig. 15.2. There is a graphical user interface that interacts with the user taking his requirements (user and situation)

---

P. Russo  
Artillery Command of the Italian Army, 00066 Bracciano, Italy  
e-mail: [russop88@gmail.com](mailto:russop88@gmail.com)

M. Panella (✉)  
Department of Information Engineering, Electronics and Telecommunications (DIET), University of Rome La Sapienza, Via Eudossiana 18, 00184 Rome, Italy  
e-mail: [massimo.panella@uniroma1.it](mailto:massimo.panella@uniroma1.it)  
URL: <http://massimopanella.site.uniroma1.it>



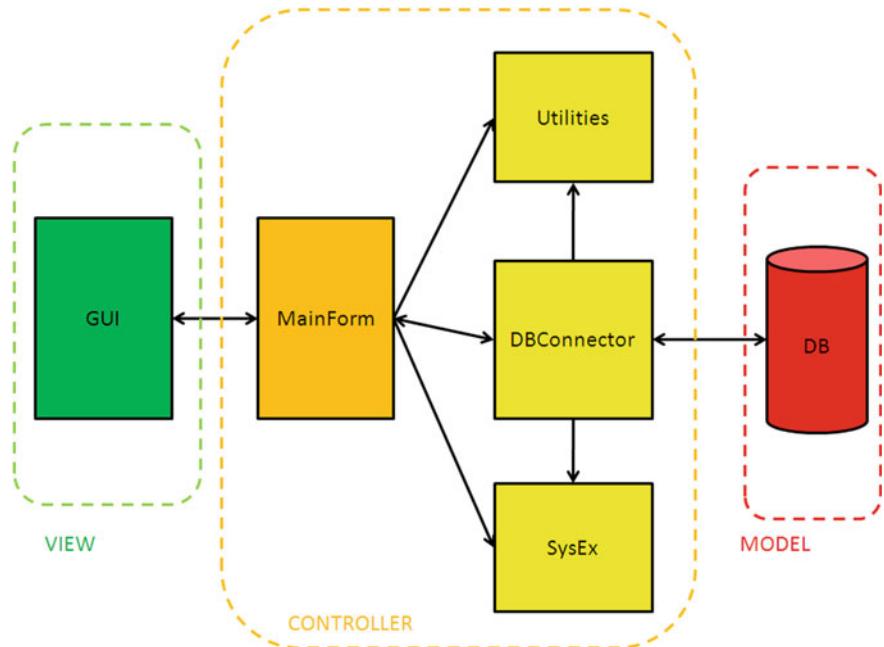
**Fig. 15.1** Joint targeting cycle composed by several functions that can be used by a military commander in order to decide whether to attack or not

passing them to the *mainForm*, the real core of the application. Inside this unit, there is the implementation of the algorithm that, through the use of *utilities* and *SysEx* units, makes the query for the *DbConnector* unit responsible for the connection with the database. Finally, the query results can be passed back to the user through a graphical user interface (GUI).

This contribution is organized as follows. In Sect. 15.2, the proposed algorithm and the dataset structure used to store the information will be presented. In Sect. 15.3, we will discuss how the algorithm has been tested and we will introduce the consequent experimental results. Finally, the conclusions about this work will be drawn in Sect. 15.4.

## 15.2 Proposed Methodology

Before starting with the description of the implemented algorithm, it is important to illustrate what is the easiest way of storing the considered data in a database (or, equivalently, a dataset) associated with the proposed application. The basic information needed by the algorithm is the name of the user, the working class and the functions activated or not. For this reason, we can define the user and the class with two different integer numbers (UserID and Class), and for each function we use a binary value defined in the following way:



**Fig. 15.2** General structure of the application

$$F_n = \begin{cases} 1, & \text{activated} \\ 0, & \text{not activated} \end{cases}$$

Finally, a parameter  $N_T$  will measure how many times that combination of activated and not activated functions has been chosen by the user. Consequently, the dataset will be composed by thousands of patterns similar to the one shown in Table 15.1.

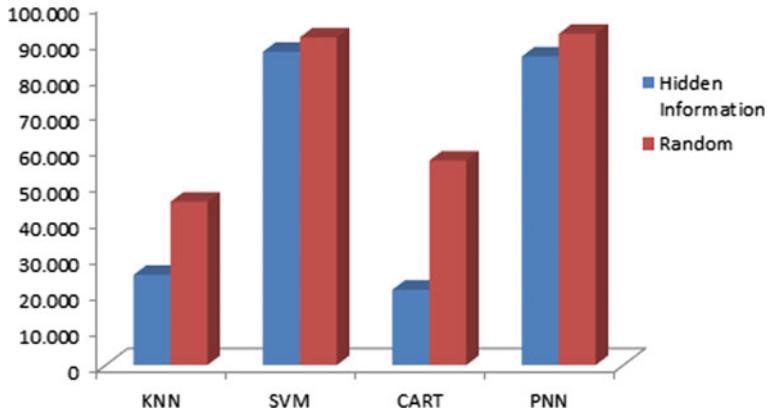
At this point, the problem is how to create different datasets that can simulate different scenarios. The datasets created to this end are:

- A random dataset where all the information is randomly inserted;
- A dataset with few hidden information on class and user (e.g., some users chose the same combination);
- A dataset with a larger quantity of hidden information.

A classification system is a tool that, starting from a data sample, is able to give it a class label [5–7]. In this case, starting from the UserID and Class labels, the clas-

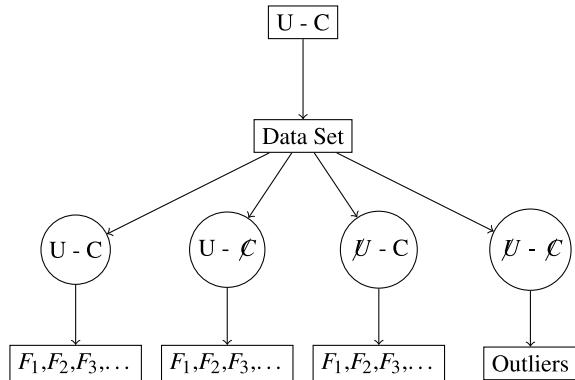
**Table 15.1** An example of raw entry in a dataset with ten functions contributing to classification

UserID	Class	$F_0$	$F_1$	...	$F_n$	$N_T$
15	3	0	1	...	1	32



**Fig. 15.3** Results of classification systems in case of hidden information on users and classes

**Fig. 15.4** Implemented decision tree



sification system should be able to predict the combination of the aforementioned functions. The considered classification models are: k-nearest neighbors (KNN), support vector machine (SVM), classification and regression tree (CART), probabilistic neural network (PNN). The error percentages in case of using such classification systems on a random dataset and on a dataset with hidden information are reported in Fig. 15.3.

It is evident that such systems cannot be used in a straightforward manner for this kind of problems because of the too high percentage of errors, only 22% in the best case. The reason for these errors is due to the ill-posed problem to be solved, which is an inverse classification problem where either the solution does not exist, the solution is not unique or the solution's behavior changes continuously with the initial conditions. Thus, four different paths can be identified, as described in Fig. 15.4: UserID and Class label (call) available in the dataset; UserID only; Class only; and both UserID and Class absent.

Nothing can be predicted when both user and class are absent; so, simply the most used combination of functions will be presented to the decision maker. In the other three cases, the purpose of application is to minimize the cost parameter  $R$ , that is:

$$R = \frac{N_T}{\sum_i C_i}, \quad (15.1)$$

where  $C_i$  is the cost of that combination. This is a hard combinatorial optimization problem, and there is not a probabilistic distribution of data. In addition, the prediction of user or class is quite hard as these two types of data are not correlated.

In order to overcome this problem, we propose the solution reported in Algorithm 1, which will be referred to in the following as ‘frequency-based inverse classifier’ (FBIC). In this algorithm, the adopted functions and parameters are:

- **UserPresenceVerification.** The input of this function is the inserted UserID, and its output is Boolean: True if the inserted user is present into the dataset; false otherwise.
- **ClassPresenceVerification.** The input of this function is the inserted class, and its output is Boolean: True if the inserted class is present into the dataset; false otherwise.
- **ClassForThatUserPresence.** This function checks if that particular user has expressed a choice in a particular class. It returns true if yes, false otherwise.
- **Record.** It is a vector of choices, where every choice is a structure composed similarly to Table 15.1, with the UserID, the Class label, the activation value for each function, the number of times  $N_T$  and the cost  $C_i$  as well.
- **FindRecordUserClass.** The inputs to this function are both user and situation, and it returns all the choices (saved into a record or, equivalently, a pattern) that a particular user did in that particular working operation (class).
- **FindRecordUser.** This function has a UserID at the input, and it returns all the choices (saved into a pattern) that he did.
- **FindRecordClass.** This function has a Class label at the input, and it returns all the choices (saved into a pattern) that all users did in that particular working operation (class).
- **FindMaxUsedRecord.** This function has a pattern of choices at the input, and it calculates the most frequent ones. The output of this function is a pattern of other choices.
- **CalculateMinRatio.** This function, which uses as input the output of the previous function, applies the formula expressed in (15.1); its output is a pattern of choices.
- **CalculateMinCost.** This function calculates the choices with the minimum cost into the input pattern; its output is another pattern.
- **CalculateMinLength.** This function calculates the choices with the minimum length (in terms of number of used functions) into the input pattern; its output is another pattern.

### 15.3 Validation and Results

The proposed system has been assessed using a leave-one-out method where the FBIC is applied once for each pattern of the dataset, using all other instances as a training set and using the selected instance as a single item test set. In addition, each item of the dataset has been considered with a different weight (because of field  $N_T$  that represents *how relevant* the combination is into the dataset). For this reason, each row has been used as test set based on the value of  $N_T$ . The error percentage  $P$  has been evaluated using the following formula:

$$P = \frac{E}{N_f \sum_{i=1}^{N_d} N_{Ti}} \quad (15.2)$$

---

#### Algorithm 1 (proposed FBIC algorithm)

---

```

1: Record, RecordMinRatio, RecordMinCost, RecordMinLength: Record;
2: FindUC, FindC, FindU : Boolean;
3: FindU = UserPresenceVerification(UserID);
4: FindC = ClassPresenceVerification(Class);
5: FindUC = ClassForThatUserPresence(UserID, Class);
6: if (FindU = true) and (FindUC = true) then
7:   Record=FindRecordUserClass(UserID, Class);
8: end if
9: if (FindU = true) and (FindUC = false) then
10:  Record=FindRecordUser(User);
11: end if
12: if (FindU = false) and (FindC = true) then
13:   Record=FindRecordClass(Class);
14: end if
15: if (FindU = false) and (FindC = false) then
16:   Record=FindMaxUsedRecord();
17: end if
18: RecordMinRatio = CalculateMinRatio(Record);
19: if RecordMinRatio.count > 1 then
20:   RecordMinCost = CalculateMinCost(RecordMinRatio);
21:   if RecordMinCost.count > 1 then
22:     RecordMinLength = CalculateMinLength(RecordMinCost);
23:   end if
24: end if
25: if RecordMinRatio.count = 1 then
26:   choice = RecordMinRatio[0];
27: end if
28: if RecordMinCost.count = 1 then
29:   choice = RecordMinCost[0];
30: end if
31: if RecordMinLength.count ≥ 1 then
32:   choice = RecordMinLength[0];
33: end if
34: return = choice;
```

---

where:

- $E$  is the total error sum. If the algorithm predicts the activation of a function instead of the deactivation (or vice versa), it is incremented by 1.
- $N_f$  is the number of functions.
- $N_d$  is the total number of dataset patterns.
- $N_{Ti}$  is the  $N_T$  value for the  $i$ th pattern,  $i = 1 \dots N_d$ .

We used 40 datasets, each one with 2200 patterns. The first 20 datasets are composed by randomly generated patterns. The other 20 datasets have a *hidden information* inside (e.g., same functions activated by the same users, a user who activates at least a particular function). This choice is motivated because, in a real operation scenario, every user usually makes more or less the same choices or, equivalently, the largest part of the users make the same choices.

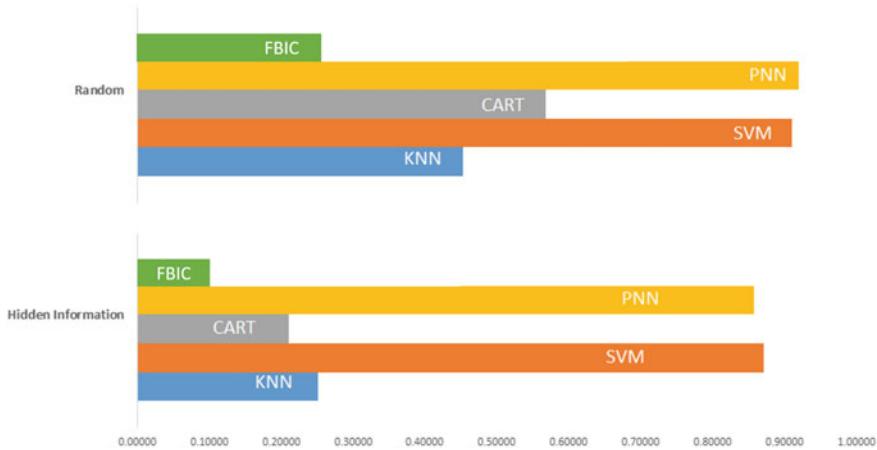
The numerical parameters to be set in advance for the standard classifiers have been determined using an inner threefold cross-validation on each training subset. For KNN, the value  $k$  of nearest neighbors varied in the range from  $k = 2$  to  $k = 10$  and we used the Euclidean distance, so the optimal choice was  $k = 3$ . CART does not depend upon any parameter to be fixed in advance. SVM adopted radial basis function (RBF) kernels for manifold adaptation, and each kernel used a Karush–Kuhn–Tucker (KKT) violation level equal to 0.05 [8]. PNN is applied with a spread of radial basis equal to 0.1.

The error percentages of classical classification systems are reported in Fig. 15.5 compared to the proposed FBIC algorithm. It is possible to appreciate two important results:

1. All algorithms work well when *hidden information* is present. This happens because there is much more information that can be used in order to predict more accurately any possible solution.
2. The error percentage of the FBIC algorithm in the best case (datasets with hidden information) is around 9%, which is about two times lower than the best error obtained by CART among standard classifiers.

The proposed FBIC algorithm has been also applied to a different scenario in order to test its behavior in more challenging situations. The adopted datasets are described in the following:

- *Random dataset*: All the dataset information is completely randomly generated, and it can be used as a performance baseline.
- *Hidden information dataset*: The information inside the dataset is stored considering hidden information; if the number of random patterns is more than 25%, this dataset is referred to as ‘Less Info DB’; otherwise, it is referred to as ‘More Info DB’.
- *Weighted error dataset*: Generally, the parameter  $E$  of (15.2) is incremented by 1 in case of error, independently of the algorithm’s decision to activate a function instead to deactivate it (or vice versa). In actual scenarios, this choice might not be useful because, for example, the wrong decision of activating a function could be



**Fig. 15.5** Comparison of error percentages for the considered standard classifiers and the proposed FBIC algorithm

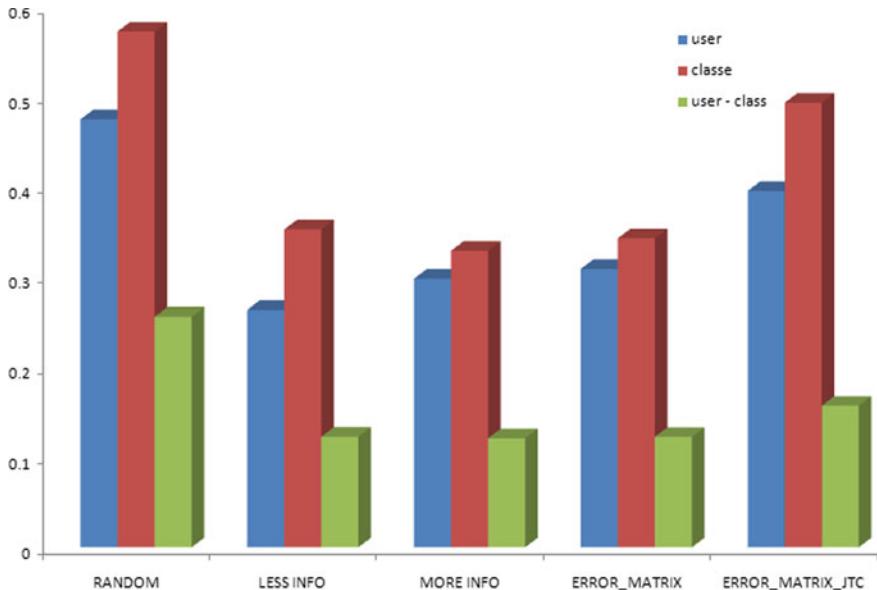
**Table 15.2** Adopted weight matrix for error computation

Function	Error (If activated)	Error (If unactivated)
$F_0$	1.3	0.7
$F_1$	1.0	1.0
$F_2$	0.0	2.0
$F_3$	0.8	1.2
$F_4$	1.7	0.3
$F_5$	0.9	1.1
$F_6$	2.0	0.0
$F_7$	0.4	1.6
$F_8$	1.0	1.0
$F_9$	1.5	0.5

more dangerous than a wrong deactivation. For this reason, we generated a matrix for which a suited weight is defined for every kind of error and thus the increment of  $E$  in (15.2) takes into account the chosen value of such weights as shown in Table 15.2.

- *JTC dataset*: In a military scenario, like the joint targeting cycle (JTC), it is much more dangerous if a function is unactivated instead of being activated than the vice versa. For this reason, the number of errors  $E$  is not incremented if the function is activated instead of unactivated, while it is incremented by 2 in the opposite case.

We created 20 datasets for all of the above-mentioned types. Every dataset is composed by 2200 patterns and the results of the FBIC algorithm, which were obtained



**Fig. 15.6** Error percentages of FBIC algorithm for the different considered datasets

also in this case by using a leave-one-out method, are presented in Fig. 15.6. Also in this case, the FBIC performance is encouraging, as it was able to achieve about 90% of accuracy when UserID and Class label are present, and such performances are much better than in the case of a purely random dataset, for which we were expecting a worse performance as actually we experienced.

## 15.4 Conclusion

A military application for strategic defense, also applicable to civil contexts, is proposed in this paper with the aim to predict the different functions that a decision support system should deal with. The goal is not only to improve performances but also to achieve accuracy when a novel solution is proposed to the user. In this regard, we firstly considered a classical classification problem applied to two typical datasets: the first one with random information and the second one with hidden information that represents the typical user behavior (such as the activation of a particular function). In these applications, the proposed FBIC algorithm is able to reduce the error percentage about two times than classical classification systems. Moreover, we tested the FBIC algorithm on datasets associated with typical operation scenarios, even considering the particular case of the joint targeting cycle. Also in this case, the results are encouraging with the lowest error percentage when the algorithm can find both user information and operation class label into the dataset. We finally out-

line that the proposed approach can be efficiently used in the analysis of datasets containing either binary or continuous data.

## References

1. U.S. Army, Joint Staff: Joint Targeting. Joint Publication 3-60, January 31 (2013). [https://www.jjustsecurity.org/wp-content/uploads/2015/06/Joint\\_Chiefs-Joint\\_Targeting\\_20130131.pdf](https://www.jjustsecurity.org/wp-content/uploads/2015/06/Joint_Chiefs-Joint_Targeting_20130131.pdf)
2. NATO, NSA: Allied Joint Doctrine for Information Operations AJP-3.10. NATO European High Quarter, Brussels (2012). <https://info.publicintelligence.net/NATO-IO.pdf>
3. Meyer, B.A.: Pattern Recognition for Medical Imaging. Elsevier Academic Press, London (2003)
4. Corr, P.: Pattern Recognition in diagnostic imaging. World Health Organization, Geneve (2001)
5. Rizzi, A., Buccino, M., Panella, M., Uncini, A.: Genre Classification of Compressed Audio Data. In: Proc. of IEEE Workshop on Multimedia Signal Processing (MMSP 2008), pp. 654–659 (2008)
6. Scardapane, S., Fierimonte, R., Wang, D., Panella, M., Uncini, A.: Distributed Music Classification Using Random Vector Functional-Link Nets. In: Proc. of International Joint Conference on Neural Networks (IJCNN 2015), pp. 272–279 (2015)
7. Altilio, R., Paoloni, M., Panella, M.: Selection of clinical features for pattern recognition applied to gait analysis. Medical and Biological Engineering and Computing **55**(4), 685–695 (2017)
8. Friedman, J., Hastie, T., Tibshirani, R.: The elements of statistical learning, 2nd edn. Springer, New York (2009)

# Chapter 16

## Simultaneous Learning of Fuzzy Sets



Luca Cermenati, Dario Malchiodi and Anna Maria Zanaboni

**Abstract** We extend a procedure based on support vector clustering and devoted to inferring the membership function of a fuzzy set to the case of a universe of discourse over which several fuzzy sets are defined. The extended approach learns simultaneously these sets without requiring as previous knowledge either their number or labels approximating membership values. This data-driven approach is completed via expert knowledge incorporation in the form of predefined shapes for the membership functions. The procedure is successfully tested on a benchmark.

### 16.1 Introduction

Fuzzy sets constitute a sort of backbone for all fuzzy constructs, such as fuzzy models, fuzzy classifiers, and fuzzy reasoning schemes. Therefore, the quality of the former directly impacts on the performance and readability of such constructs. The design of fuzzy sets is a crucial problem in both the theory and the practice of fuzzy methodologies, and indeed there is a broad spectrum of approaches aiming at building fuzzy sets. On one side, fuzzy sets are designed exploiting human knowledge through a mix of different interpretations [9], expert-driven approaches [14], predefined shapes for membership functions [15], and specific degranulation processes [5, 17]. However, the availability of experts in the modeled domain might be a critical aspect, and in any case this kind of estimation has been shown to suffer from incompleteness, inconsistencies, or bias linked to the perception of specific concepts captured by humans [16, 18]. For these reasons, on the other extreme of the spectrum of methodologies there are data-driven approaches, relying only on experimental evidence (see, for instance, [1, 10, 19]). Several strategies actually position themselves between the two extremes, combining them in a hybrid fashion [2, 4, 8, 11].

---

L. Cermenati · D. Malchiodi (✉) · A. M. Zanaboni  
Dipartimento di Informatica, Università degli Studi di Milano, Milan, Italy  
e-mail: [malchiodi@di.unimi.it](mailto:malchiodi@di.unimi.it)

A. M. Zanaboni  
e-mail: [zanaboni@di.unimi.it](mailto:zanaboni@di.unimi.it)

In this work, we propose a technique mixing data-driven and expert-driven approach. The former is used to infer the number of fuzzy sets in a given domain and their approximate localization, while the latter is used to define *a priori* the family shape of such sets. The starting point is a procedure exploiting a modified support vector clustering approach [12] in order to learn the membership function of a single fuzzy set, starting from examples of objects labeled with their membership value (see also [6] for a similar approach based on modified regression). In this paper, such approach is extended in two significant ways: On the one hand, the need of labels representing membership values of the observed objects is dropped, and on the other the inference process now concerns *several* fuzzy sets simultaneously. The number and location of such sets are found through application of the original version of the support vector clustering algorithm [7], in order to label objects via approximate membership values; the labels are subsequently used in order to separately learn each fuzzy set.

The paper is structured as follows: Sect. 16.2 briefly describes the technique used for inferring the membership function of a single fuzzy set on the basis of a sample of objects in the universe of discourse, each one labeled with its membership degree to such set. Section 16.3 exploits the above-mentioned technique in order to simultaneously learn several fuzzy sets, starting from a set of unlabeled objects. Section 16.4 describes a preliminary experimental campaign. Some concluding remarks end the paper.

## 16.2 Inferring the Membership Function to a Fuzzy Set

In this section, we briefly recall the procedure used in order to learn the membership function to a fuzzy set starting from a labeled sample  $\{(x_1, \mu_1), \dots, (x_m, \mu_m)\}$ , where for each  $i$  the value  $x_i$  denotes an object in a space  $X$  and the label  $\mu_i$  is the membership grade of  $x_i$  to a fixed, yet unknown, fuzzy set  $A$ . Readers interested in further details may refer to the original paper [12].

The main component of the learning procedure is a modified version of the support vector clustering algorithm proposed in [7], enhanced in order to deal with labels  $\mu_1, \dots, \mu_m$ . Namely, objects are transformed through a nonlinear mapping  $\Phi$  onto a space within which a sphere  $S$  is found such that:

- The higher the  $\mu_i$ , the closer  $x_i$  is to the border of  $S$  (and when  $\mu_i = 1$  the object belongs to  $S$ ).
- Vice versa, as  $\mu_i$  gets smaller the corresponding object lies farther from  $S$ .
- The radius of  $S$  is constrained to be as small as possible.

More precisely, denoting by  $a$  and  $R$  the center and the radius of  $S$ , respectively, this amounts to considering the problem

$$\min R^2 + C \sum (\xi_i + \tau_i) \quad (16.1)$$

$$\mu_i \|\Phi(x_i) - a\|^2 \leq \mu_i R^2 + \xi_i, \quad (16.2)$$

$$(1 - \mu_i) \|\Phi(x_i) - a\|^2 \geq (1 - \mu_i) R^2 - \tau_i, \quad (16.3)$$

$$\xi_i \geq 0, \tau_i \geq 0, \quad (16.4)$$

where  $\xi_i$  and  $\tau_i$  denote slack variables allowing the management of possible outliers and  $C > 0$  is a hyperparameter defining a trade-off between the two components of the objective function in (16.1). As usual with support vector methods, the solution can be found considering the dual version of (16.1–16.4), which reads

$$\max \sum_{i=1}^m \epsilon_i k(x_i, x_i) - \sum_{i,j=1}^m \epsilon_i \epsilon_j k(x_i, x_j) \quad (16.5)$$

$$\sum_{i=1}^m \epsilon_i = 1, \quad (16.6)$$

$$-C(1 - \mu_i) \leq \epsilon_i \leq C\mu_i, \quad (16.7)$$

where  $\epsilon_i = \alpha_i \mu_i - \beta_i(1 - \mu_i)$  for each  $i = 1, \dots, m$  (being  $\alpha_i$  and  $\beta_i$  the Lagrangian multipliers associated with the constraints (16.2) and (16.3), respectively) and  $k(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ ; that is,  $k$  is the *kernel function* associated with the mapping  $\Phi$ . Thus, the considered objects do not need to be numerical vectors: The only requirement is the existence of a similarity measure  $k$  between them. For instance, [13] applies this technique to the problem of detecting a set of reliable axioms starting from a set of OWL formulas.

The experiments shown in Sect. 16.4 make use of the *Gaussian kernel*

$$k(x_i, x_j) = \exp \left( -\frac{\|x_i - x_j\|^2}{2\sigma^2} \right),$$

although other choices are possible. Here,  $\sigma > 0$  is a second hyperparameter to be tuned when performing experiments.

Once the optimal values  $\epsilon_1^*, \dots, \epsilon_m^*$  of (16.5–16.7) have been computed, it is easy to show that

$$R^2(x) = k(x, x) - 2 \sum_{i=1}^m \epsilon_i^* k(x, x_i) + \sum_{i,j=1}^m \epsilon_i^* \epsilon_j^* k(x_i, x_j) \quad (16.8)$$

amounts to the squared distance between  $a$  and the image through  $\Phi$  of a generic object  $x$ . Moreover, given any  $k$  such that  $-C(1 - \mu_k) < \epsilon_k^* < C\mu_k$ , the quantity  $R^{2,*} = R^2(x_k)$  equals the squared radius of  $S$ . Thus, it is easy to take a further step and induce an approximation  $\hat{\mu}_A$  of the membership function of  $A$  as follows: having fixed a suitable *fuzzifier* (that is, a nonincreasing function  $f : \mathbb{R}^+ \mapsto [0, 1]$  turning the distance of the image of a generic point from  $S$  into a membership value), and

given a generic object  $x^N$ , let  $\hat{\mu}_A(x^N) = f(R^2(x^N) - R^{2,*})$ . A simple choice for  $f$  is that of a piecewise linear function equal to 1 when the image of an object lies within  $S$  and equal to 0 when such image is farther from  $S$  than the farthest observed distance  $x_{\max}$ , and decreases linearly between these extremes in the remaining cases:

$$f(x) = \begin{cases} 1 & \text{if } x < 0, \\ 1 - \frac{x}{x_{\max}} & \text{if } 0 \leq x \leq x_{\max}, \\ 0 & \text{otherwise.} \end{cases} \quad (16.9)$$

More complex choices for  $f$ , such as a special exponential decaying function linked with the quantiles of the observed distances of  $x_i$ s from the border of  $S$ , may be considered (see [13] for further details).

### 16.3 Simultaneously Inferring Several Membership Functions

As a general case, the method outlined in the previous section requires as input a set of objects each labeled with a degree of membership, thus a  $[0, 1]$ -valued number. However, the procedure can be run even when the information about membership degrees is not available, yet each object is labeled with a  $\{0, 1\}$  value denoting its crisp membership to a (classical) set.

In this section, we address the more general case in which even this weaker form of information is missing; that is, the only available data is the set  $T = \{x_1, \dots, x_m\}$  of objects, in the idea that *several* fuzzy sets are defined on the universe of discourse  $X$ . As a first stage, the original version of the support vector clustering can be applied in order to detect a sort of *core* for each fuzzy set. This step can be described as a simplified version of the procedure described in Sect. 16.2: Now  $S$  identifies with the smallest sphere containing most of the images of objects, and an analogous procedure allows to compute:

- A function  $R_{\text{cluster}}^2$  mapping a generic object to the squared distance of its image through  $\Phi$  from the center of  $S$ , and
- The squared radius  $R_{\text{cluster}}^{2,*}$  of  $S$ .

Now, let  $x_a$  and  $x_b$  denote two objects in  $X$  belonging to different clusters, and consider the segment joining them. It can be shown that the trajectory described by the images through  $\Phi$  of all points laying on this segment is not fully contained in  $S$  [7]. This fact can be easily checked considering suitable discretizations of the segments joining all possible pairs of objects. As a result, the set  $T$  can be partitioned in  $c$  subsets, namely  $T = \cup_{k=1}^c T_k$  and  $T_i \cap T_j = \emptyset$  for each  $i \neq j$ . These subsets can be interpreted as an initial approximation of the localization for  $c$  fuzzy sets  $A_1, \dots, A_c$ . Thus, for each  $k \in \{1, \dots, c\}$ , objects in  $T_k$  and  $T \setminus T_k$  can be assigned a membership equal to 1 and 0, respectively. The next step consists in applying the

---

**Algorithm 1** Procedure for simultaneous learning of memberships to several fuzzy sets, using crisp intermediate membership values

---

Input:

- A set  $T = \{x_1, \dots, x_m\}$  of objects;
- Two hyperparameters  $C, \sigma > 0$ ;
- A fuzzifier  $f : X \mapsto [0, 1]$ .

1. Apply support vector clustering to  $T$  using  $C$  and  $\sigma$  as hyperparameters, obtaining a partition  $T_1, \dots, T_c$  of  $T$ .
2. For each  $k = 1, \dots, c$ :

- (a) For each  $i = 1, \dots, m$  set

$$\mu_i = \begin{cases} 1 & \text{if } x_i \in T_k, \\ 0 & \text{otherwise.} \end{cases}$$

- (b) Starting from  $T$  and  $\{\mu_1, \dots, \mu_m\}$ , infer  $\hat{\mu}_{A_k}$  using the procedure of Sect. 16.2, the fuzzifier  $f$ , and the hyperparameters  $C$  and  $\sigma$ .

Output:  $\hat{\mu}_{A_1}, \dots, \hat{\mu}_{A_c}$ .

---

procedure of Sect. 16.2 in order to obtain an approximation  $\hat{\mu}_{A_k}$  of the membership function  $\mu_{A_k}$ . Algorithm 1 formalizes this procedure. It is worth noting that the number  $c$  of obtained fuzzy sets is not fixed *a priori*, albeit it is influenced from the choice of hyperparameters, notably  $C$ . This means that any preexisting clue about the number of sets can in principle be used in order to restrict the variability of hyperparameters during the model selection phase.

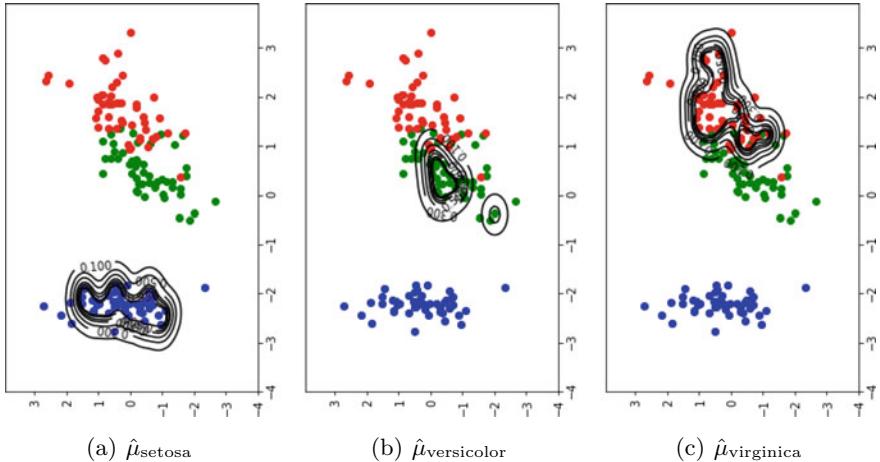
A variant of the proposed technique considers a different way of computing the intermediate values  $\mu_1, \dots, \mu_c$  in step 2a of Algorithm 1. Indeed, a better approximation might be found in terms of:

- The squared distance  $R_{\text{cluster}}^2(x_i)$  of the images of each  $x_i$  from the center of the sphere learnt during the support vector clustering phase, and
- The squared radius  $R_{\text{cluster}}^{2,*}$  of the same sphere.

More precisely, having fixed the fuzzifier  $f$ ,  $f(R_{\text{cluster}}^2(x_i) - R_{\text{cluster}}^{2,*})$  can be used as a guess for the membership value of  $x_i$ , as illustrated in Algorithm 2.

## 16.4 Experiments

We used as benchmark the Iris dataset, consisting of 150 observations of iris plants in terms of length and width of their petal and sepal. The observations are organized in the three classes *Setosa*, *Virginica*, and *Versicolor*, with only the first class being linearly separable from the remaining ones. As a first experiment, for the sake of visualization we extracted the first two principal components from the observations and we iterated for ten times the holdout scheme described below.



**Fig. 16.1** Contour plots of the functions  $\hat{\mu}_{\text{setosa}}$ ,  $\hat{\mu}_{\text{versicolor}}$ , and  $\hat{\mu}_{\text{virginica}}$  learnt during one of the ten holdout iterations of the first experiment. Blue, green, and red bullets, respectively, denote observations from the setosa, versicolor, and virginica classes, after the first two principal components have been extracted from the original data

- After having randomly shuffled all data, we partitioned the benchmark into three sets devoted to training, model selection, and model validation (retaining in each one the 80, 10, and 10% of available data, respectively).
- For each choice of  $C$  and  $\sigma$  in a grid, we applied Algorithm 1 and obtained three membership functions which we called  $\hat{\mu}_{\text{setosa}}$ ,  $\hat{\mu}_{\text{virginica}}$ , and  $\hat{\mu}_{\text{versicolor}}$ , and we used them to compute the accuracy in classification of data in the validation set (namely, each item was assigned to the class whose corresponding membership function attained the maximum value). It is worth noting that, being the number  $c$  of fuzzy sets learnt by the algorithm, the former could be different from the expected value corresponding to the three classes in the dataset. We dropped all singleton clusters after the initial phase and sorted the remaining ones w.r.t. their size. The three biggest resulting clusters were subsequently associated each with the most represented class.<sup>1</sup>
- The choice of hyperparameters maximizing the above-mentioned accuracy was selected in order to retrain a model, now merging training and validation sets, and the result was scored in terms of accuracy on the test set.

The average accuracy on the ten holdout iterations was 0.8, with a standard deviation of 0.11. Figure 16.1 shows the contour plots of the inferred membership functions for the three classes in the benchmark in one of the iterations.

<sup>1</sup>Note that even with this careful setting, there is no guarantee that the three clusters will get associated injectively with the three available classes. We simply re-executed the iterations in which these cases occurred.

---

**Algorithm 2** Procedure for simultaneous learning of memberships to several fuzzy sets, using fuzzy intermediate membership values

---

Input:

- A set  $T = \{x_1, \dots, x_m\}$  of objects;
- Two hyperparameters  $C, \sigma > 0$ ;
- A fuzzifier  $f : X \mapsto [0, 1]$ .

1. Apply support vector clustering to  $T$  using  $C$  and  $\sigma$  as hyperparameters, obtaining:

- A partition  $T_1, \dots, T_c$  of  $T$ ;
- A mapping  $R_{\text{cluster}}^2 : X \mapsto \mathbb{R}^+$ ;
- A value  $R_{\text{cluster}}^{2,*}$ .

2. For each  $k = 1, \dots, c$ :

- (a) For each  $i = 1, \dots, m$  set  $\mu_i = f(R_{\text{cluster}}^2(x_i) - R_{\text{cluster}}^{2,*})$
- (b) Starting from  $T$  and  $\{\mu_1, \dots, \mu_m\}$ , infer  $\hat{\mu}_{A_k}$  using the procedure of Sect. 16.2, the fuzzifier  $f$ , and the hyperparameters  $C$  and  $\sigma$ .

Output:  $\hat{\mu}_{A_1}, \dots, \hat{\mu}_{A_c}$ .

---

We repeated the experiment considering three and four principal components, obtaining the results shown in Table 16.1. The accuracy rates are sufficiently high to state that the method succeeds in rebuilding the information about the three classes although such information has been hidden to the learning procedure. The table also shows the performance of an analogous procedure based on the fuzzy C-means algorithm as base learner. The results, slightly in favor of the proposed methodology, could be improved using a more refined model validation scheme and/or trying different shapes, notably nonlinear ones, for the fuzzifier. We are currently testing Algorithm 2 on the same benchmark.

**Table 16.1** Results of ten holdout procedures of the simultaneous fuzzy set learning procedure on the Iris dataset. Each row shows average and standard deviation (columns Avg. and Stdev., respectively) of train and test error, in function of the number of principal components extracted from the original sample (#PC)

No. of principal compo- nents #PC	SVC				FCM			
	Train error		Test error		Train error		Test error	
	Avg.	Stdev.	Avg.	Stdev.	Avg.	Stdev.	Avg.	Stdev.
2	0.82	0.05	0.8	0.11	0.83	0.01	0.80	0.05
3	0.87	0.03	0.83	0.07	0.83	0.02	0.84	0.07
4	0.85	0.05	0.89	0.09	0.83	0.01	0.84	0.08

## 16.5 Conclusions

The design of fuzzy sets is an essential component in the search of successful fuzzy models. We considered how to extend an existing learning algorithm for the membership function of a fuzzy set on the basis of objects labeled with the corresponding membership grades. This algorithm was enhanced in order to simultaneously learn several fuzzy sets defined in the considered universe of discourse. The number of sets can in principle be induced directly from data, and the latter do not need to be labeled with any information concerning the membership grades w.r.t. the models to be learnt. We preliminarily tested the proposed approach on the Iris dataset, showing how the three existing clusters can be discovered without using the class information recorded in the benchmark, but only with little post-processing, as mentioned. Besides a more refined experimental campaign, the technique can be further refined analyzing how preexisting information about the number of fuzzy sets is related to a proper choice of the hyperparameters of the learning algorithm, and by testing the effect of using different nonlinear fuzzifiers. An analysis of the theoretical properties of this approach, for instance exploiting game-based results [3], as well as its extension to the field of type-2 fuzzy sets, can also be envisaged.

**Acknowledgements** The authors would like to thank Angelo Ciaramella and Antonino Staiano for the fruitful discussion concerning the organization of the experimental phase.

## References

1. Afify, A.: FuzzyRULES-II: a new approach to fuzzy rule induction from numerical data. *Front. Artif. Intell. Appl.* **281**, 91–100 (2016)
2. Apolloni, B., Bassis, S., Gaito, S., Malchiodi, D.: Bootstrapping complex functions. *Nonlinear Anal.: Hybrid Syst.* **2**(2), 648–664 (2008)
3. Apolloni, B., Bassis, S., Gaito, S., Malchiodi, D., Zoppis, I.: Controlling the losing probability in a monotone game. *Inf. Sci.* **176**(10), 1395–1416 (2006)
4. Apolloni, B., Bassis, S., Malchiodi, D., Pedrycz, W.: Interpolating support information granules. Proceeding of the 16th International Conference on Artificial Neural Networks - Part II. ICANN'06, pp. 270–281. Springer-Verlag, Berlin, Heidelberg (2006)
5. Apolloni, B., Iannizzi, D., Malchiodi, D., Pedrycz, W.: Granular regression. In: Neural nets: 16th italian workshop on neural nets, WIRN2005 and international workshop on natural and artificial immune systems, NAIS 2005, vol. 3931 LNCS, pp. 147–156. Springer (2006)
6. Apolloni, B., Malchiodi, D., Valerio, L.: Relevance regression learning with support vector machines. *Nonlinear Anal., Theory, Methods Appl.* **73**(9), 2855–2864 (2010)
7. Ben-Hur, A., Horn, D., Siegelmann, H.T., Vapnik, V.: Support vector clustering. *J. Mach. Learn. Res.* **2**(Dec), 125–137 (2001)
8. Dubois, D., Hájek, P., Prade, H.: Knowledge-driven versus data-driven logics. *J. Log., Lang. Inf.* **9**(1), 65–89 (2000)
9. Dubois, D., Prade, H.: The three semantics of fuzzy sets. *Fuzzy Sets Syst.* **90**, 141–150 (1997)
10. Hong, T.P., Lee, C.Y.: Induction of fuzzy rules and membership functions from training examples. *Fuzzy Sets Syst.* **84**(1), 33–47 (1996)
11. Lo Presti, M., Poluzzi, R., Zanaboni, A.M.: Synthesis of fuzzy controllers through neural networks. *Fuzzy Sets Syst.* **71**, 47–70 (1995)

12. Malchiodi, D., Pedrycz, W.: Learning membership functions for fuzzy sets through modified support vector clustering. In: Masulli, F., Pasi, G., Yager R.R. (eds.) *Fuzzy Logic and Applications - 10th International Workshop, WILF 2013, Genoa, Italy, November 19–22, 2013. Proceedings*, vol. LNCS 8256, pp. 52–59. Springer (2013)
13. Malchiodi, D., Tettamanzi, A.G.B.: Predicting the possibilistic score of OWL axioms through modified support vector clustering. In: SAC 2018: Symposium on Applied Computing, April 9–13, 2018, Pau, France. ACM, New York, NY, USA (2018). <https://doi.org/10.1145/3167132.3167345>
14. Mottaghi-Kashtiban, M., Khoei, A., Hadidi, K.: Optimization of rational-powered membership functions using extended kalman filter. *Fuzzy Sets Syst.* **159**(23), 3232–3244 (2008)
15. Nguyen, H., Walker, E.: *A First Course in Fuzzy Logic*. CRC Press, Boca Raton, Chapman Hall (1999)
16. de Oliveira, J.V.: Semantic constraints for membership function optimization. *IEEE Trans. Syst., Man, Cybern.-Part A: Syst.Humans* **29**(1), 128–138 (1999)
17. Pedrycz, W.: *Granul. Comput.: Anal. Des. Intell. Syst.* CRC Press/Francis Taylor, Boca Raton (2013)
18. Setnes, M., Babuska, R., Verbruggen, H.B.: Transparent fuzzy modelling. *Int. J. Hum.-Comput. Stud.* **49**(2), 159–179 (1998)
19. Wu, J.N., Cheung, K.C.: An efficient algorithm for inducing fuzzy rules from numerical data. In: *Proceedings of the Eleventh International FLAIRS Conference*, pp. 221–224. AAAI (1998)

# Chapter 17

## Trees in the Real Field



Alessandro Betti and Marco Gori

**Abstract** This paper proposes an algebraic view of trees which opens the doors to an alternative computational scheme with respect to classic algorithms. In particular, it is shown that this view is very well-suited for machine learning and computational linguistics.

### 17.1 Introduction

In the last few years, models of deep learning have been successfully applied to computational linguistics. Among others, the translation problem has benefited significantly from simple approaches based on recurrent neural networks. In particular, because of the classic problem of capturing long-term dependencies [1], LSTM [3] architectures have been mostly used which can better deal with this classic problem.

In this paper, we go beyond this approach and assume to characterize linguistic production by means of generative trees by relying on the principle that the complexity of the problem of long-term dependencies is dramatically reduced because of the exponential growth of nodes of the trees with respect to their height. In general, the relations between trees and their corresponding linear encoding is not easy to grasp. For example, when restricting to binary trees, it can be proven that we need a pair of traversals to fully characterize a given tree, one of which may be the symmetric one [4]. However, whenever a sequence presents a certain degree of regularity, the ambition arises to establish a bijection with a corresponding tree (e.g., the parsing tree).

---

A. Betti  
University of Florence, Florence, Italy  
e-mail: [alessandro.betti@unifi.it](mailto:alessandro.betti@unifi.it)

A. Betti · M. Gori (✉)  
SAILab, University of Siena, Siena, Italy  
e-mail: [marco@diism.unisi.it](mailto:marco@diism.unisi.it)  
URL: <http://sailab.diism.unisi.it>

While encoding mechanisms are quite straightforward to design every time that it is possible to assign to each sequence a tree-like structure since it is sufficient to propagate the information (e.g., with a linear scheme) through the nodes up to the root of the tree [2]; it is much harder to come up with a decoding scheme that generates the translated sequence. Here, we prove that we can construct a decoding scheme that naturally extend those used nowadays in recurrent neural nets that can be potentially very interesting in computational linguistics.

## 17.2 Uniform Real-Valued Tree Representations

A binary tree is recursively defined as

$$T = \begin{cases} T_\emptyset & \text{basis} \\ (L, y, R) & \text{induction} \end{cases} \quad (17.1)$$

where  $T_\emptyset$  is the empty tree,  $y \in \Sigma$  is the labeled root, which takes on values from the alphabet  $\Sigma$ ,  $L$  (Left), and  $R$  (Right) are trees. We assume that we are given a *coding function*  $\ell : \Sigma \rightarrow \mathcal{Y} \subset \mathbb{R}^p$ , so as the nodes of the tree are related to an associated point<sup>1</sup> of  $\mathcal{Y}$ . Now, let us consider the pair

$$\begin{aligned} \text{i. } T &:= (L, x, R) \\ \text{ii. } \gamma : \mathcal{X} \subset \mathbb{R}^n &\rightarrow \mathcal{Y}, \quad \gamma(x) := Cx \end{aligned} \quad (17.2)$$

which consists of the triple  $(L, x, R)$  and of the *linear labeling function*  $\gamma$ , which returns points, that will be related to the labels of  $T$ . In the triple, we have  $x \in \mathbb{R}^n$ ,  $L, R \in \mathbb{R}^{n \times n}$ , while  $C \in \mathbb{R}^{n \times p}$ . Basically, we introduce a computational scheme on the *embedding space*  $\mathcal{X}$ . If  $x = 0$ , then we assume that  $(L, 0, R) \sim (0, 0, 0) := T_\emptyset$ . We want to explore the relations between the tree definition (17.1) and the related real-valued representation given by Eq. (17.2). To this end, we start noticing that the void tree  $T_\emptyset$  can be associated with  $T_\emptyset$ . The idea is that we can specify a tree  $T$  once the triple  $(L, x, R)$  and  $C$  are given. Beginning from  $Cx = \text{root}(T)$ , we process the children of the root by applying  $L$  and  $R$  to  $x$ , so that  $CRx$  is the right child and  $CLx$  is the left child. Then, the left child of the left child of the root is obtained as  $CLRx$ , and the right child of the left child of the root as  $CRLx$ , and so on and so forth, until we find, for each branch of the tree, a node  $l \in \mathbb{R}^n$  for which  $Cl = Rl = 0$ . This will be the leaf of that particular path, and we will say that the children of the leaves are buds; more generally every null node will be denoted as a bud.

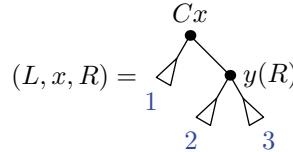
We say that  $(L, x, R)$  is an  $n$ -dimensional *real representation* of the obtained tree  $T$ .

---

<sup>1</sup>In the following, we will often regard the elements of  $T$  as elements of  $\mathbb{R}^p$ , without mentioning function  $\ell$  explicitly.

In order to get an insight into this construction, let us consider the following examples.

*Example 1* The first non-trivial example is the tree that consists of the root only. In our representation, this tree is obtained by picking up any two matrices  $L$  and  $R$ , such that  $x$  in their kernel, that is  $Lx = Rx = 0$ . The simplest next example is given by



The decoding equations that define this tree are

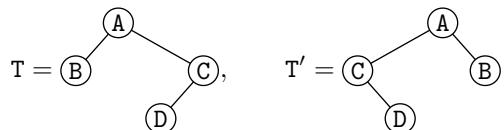
$$\begin{cases} Cx = \text{root}(T); \\ CRx = y(R), \end{cases} \quad \begin{cases} CLx = 0, & \text{bud 1;} \\ CLRx = 0, & \text{bud 2;} \\ CR^2x = 0, & \text{bud 3,} \end{cases}$$

They are conveniently separated into the “node conditions” and “bud conditions.” In order to be even more explicit consider the case  $C = I, x = (1, 0)'$  and  $y(R) = (0, 1)'$ , then it is easy to check that

$$\begin{array}{c} (1) \\ \diagdown \quad \diagup \\ \bullet \quad \bullet \\ \diagup \quad \diagdown \\ (0) \quad (0) \\ \text{1} \quad \text{2} \quad \text{3} \end{array} = \left( \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \right). \quad (17.3)$$

We can easily see that in this special case, this representation is unique in  $\mathbb{R}^2$ .

As soon as we think about the next example with two nodes , a symmetry property of the decoding scheme becomes evident. Given  $T$ , let us define the symmetric left-right  $T'$  as the tree that one obtains from  $T$  by recursively exchanging the left with the right subtrees. For example,



are related by the defined symmetry operation. Clearly, for these trees we can state an immediate property on their representation.

**Proposition 1** Let  $T$  and  $T'$  be related by left-right symmetry and let  $(L, x, R)$  be a real representation of  $T$ . Then,  $(R, x, L)$  is the representation of  $T'$ .

*Proof.* Straightforward.

□

This result immediately shows us when looking at the tree given by (17.3) that we have

$$\begin{array}{c} (1) \\ \bullet \\ (0) \quad \bullet \\ (1) \quad \bullet \\ \diagdown \quad \diagup \\ 3 \quad 2 \end{array} = \left( \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right).$$

*Example 2* In this case, we show the role of the embedding space  $\mathcal{X} \subset \mathbb{R}^n$ . In particular, we will see that the decoding might not be solvable at certain dimensions and that there could be also infinite solutions. Let us consider the following tree with the associated *decoding equations*

$$(L, x, R) = \begin{array}{c} Cx \\ \bullet \\ y(L) \quad \bullet \\ 1 \quad y(RL) \\ \diagdown \quad \diagup \\ 2 \quad 3 \end{array}, \quad \begin{cases} Cx = \text{root}(T); \\ CLx = y(L); \\ CRLx = y(RL), \end{cases} \quad \begin{cases} CL^2x = 0, & \text{bud 1;} \\ CLRLx = 0, & \text{bud 2;} \\ CR^2Lx = 0, & \text{bud 3;} \\ CRx = 0, & \text{bud 4.} \end{cases}$$

We consider two different cases  $n = 2$  and  $n = 3$ .

– **Case  $n = 2$ .** Let us consider  $n = 2$  and assume  $C = I$ . In addition, let us assume that the nodes of  $T$  are coded by

$$\text{root}(T) = (1, 0)', \quad y(L) = (0, 1)', \quad y(RL) = (1, 1)'.$$

From  $CL^2x = 0$  and from  $CLx = y(L)$ , we get  $L(Lx) = 0$  that is  $Ly(L) = 0$ . This yields a constraint on the structure of  $L$ ; we have

$$\begin{pmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \rightarrow L = \begin{pmatrix} l_{11} & 0 \\ l_{21} & 0 \end{pmatrix}.$$

Likewise from  $CRLx = y(RL)$ , we get

$$\begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rightarrow R = \begin{pmatrix} r_{11} & 1 \\ r_{21} & 1 \end{pmatrix}$$

From  $CRLx = 0$ , we get

$$l_{11}(r_{11}l_{11} + l_{21}) = 0 \quad (17.4)$$

$$l_{21}(r_{21}l_{11} + l_{21}) = 0 \quad (17.5)$$

Now, let  $x = (x_1, x_2)'$ . From  $Lx = y(L)$ , we get  $l_{11}x_1 = 0$  and  $l_{21}x_1 = 1$ . Then,  $l_{11} = 0$ , which, in turn, satisfies (17.5). Then, from (17.5) we get  $l_{21} = 0$ . Then, we end up into an impossible satisfaction of  $l_{21}x_1 = 1$ .

- **Case  $n = 3$ .** Let us consider  $n = 3$  and still assume  $C = I$ . In addition, let us assume that the nodes of T are coded by

$$\text{root}(T) = (1, 0, 0)', \quad y(L) = (0, 1, 0)', \quad y(RL) = (0, 0, 1)'.$$

From  $Cx = \text{root}(T)$ , we get  $x = (1, 0, 0)$ . From  $CL^2x = 0$  and from  $CLx = y(L)$ , we get  $L(Lx) = 0$  that is  $Ly(L) = 0$ . This yields a constraint on the structure of  $L$ ; we have

$$\begin{pmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \rightarrow L = \begin{pmatrix} l_{11} & 0 & l_{13} \\ l_{21} & 0 & l_{23} \\ l_{31} & 0 & l_{33} \end{pmatrix}.$$

From  $Lx = y(L)$ , we get

$$\begin{pmatrix} l_{11} & 0 & l_{13} \\ l_{21} & 0 & l_{23} \\ l_{31} & 0 & l_{33} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

that is  $l_{21} = 1$  and  $l_{11} = l_{31} = 0$ . Likewise from  $CRLx = y(RL)$ , we get

$$\begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \rightarrow R = \begin{pmatrix} r_{11} & 0 & r_{13} \\ r_{21} & 0 & r_{23} \\ r_{31} & 1 & r_{33} \end{pmatrix}.$$

From  $CLRLx = 0$ , we get

$$\begin{pmatrix} 0 & 0 & l_{13} \\ 1 & 0 & l_{23} \\ 0 & 0 & l_{33} \end{pmatrix} \cdot \begin{pmatrix} r_{11} & 0 & r_{13} \\ r_{21} & 0 & r_{23} \\ r_{31} & 1 & r_{33} \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & l_{13} \\ 1 & 0 & l_{23} \\ 0 & 0 & l_{33} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Hence,

$$\begin{pmatrix} 0 & 0 & l_{13} \\ 1 & 0 & l_{23} \\ 0 & 0 & l_{33} \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

which is satisfied if  $l_{13} = l_{23} = l_{33} = 0$ .

From  $CR^2Lx = 0$ , we get

$$\begin{pmatrix} r_{11} & 0 & r_{13} \\ r_{21} & 0 & r_{23} \\ r_{31} & 1 & r_{33} \end{pmatrix} \cdot \begin{pmatrix} r_{11} & 0 & r_{13} \\ r_{21} & 0 & r_{23} \\ r_{31} & 1 & r_{33} \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \rightarrow \begin{pmatrix} r_{11} & 0 & r_{13} \\ r_{21} & 0 & r_{23} \\ r_{31} & 1 & r_{33} \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Finally, from  $Rx = 0$  we need  $r_{11} = 0$ . Then, we conclude that  $L$  and  $R$  are solutions whenever they have the structure

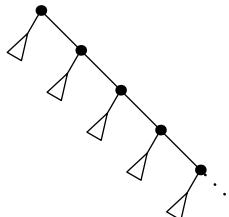
$$L = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad R = \begin{pmatrix} 0 & 0 & 0 \\ r_{21} & 0 & 0 \\ r_{31} & 1 & 0 \end{pmatrix}.$$

Notice that in this case we discover infinite solutions. In addition, it is worth mentioning that this solution originates from the required labeling, since it immediately requires to choose  $x = (1, 0, 0)$ . This makes it possible to satisfy the matrix monomial equations without requiring strong nilpotent conditions on the matrices. In addition, in this case, there is no solution for any  $x$ , since, otherwise, we need to require  $R = 0$ . As a consequence, the other labeling conditions would not be met. If we assume to keep a representation based on the above matrices  $L, R$ , then a different choice of  $x$  may lead to a completely different tree. For example, we can easily see that the choices  $x = (0, 1, 0)', (0, 0, 1)'$  yield infinite trees.

Interestingly, the generation of infinite trees is not an exception, but quite a common property of the introduced generative scheme.

Let us consider a simple example that clearly shows the possible explosion of the introduced generation scheme. Let us consider a tree whose elements are two-dimensional vectors, and consider a two-dimensional representation; in addition, for the sake of simplicity, let us assume that  $C = I$  and  $\text{root}(T) = (1, 0)'$ . Then let us assume that  $R$  is a  $\pi$  rotation and  $L$  is a projection onto the  $y$  axis:

$$x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad L = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad R = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.$$



An infinite tree with flipping labels is generated that is shown in the above figure.

As shown in the previous examples, we are interested in solving equations involving monomials of matrices. Let us focus on the algebraic side and consider the following example.

*Example 3* Let us consider the monomial equation

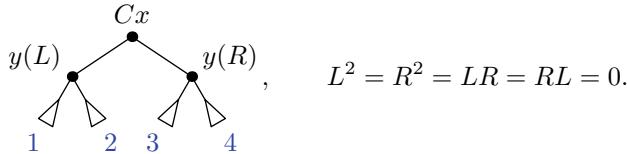
$$LR = 0. \quad (17.6)$$

What are the non-null matrices  $L$  and  $R$  which satisfy this equation? Clearly, equations like  $L^2 = 0$  and  $R^2 = 0$  define nilpotent matrices of order 2. Equation (17.6) can be regarded as a sort of generalization of the notion of nilpotent matrix to the case in which the property involves two matrices.

This problem has generally infinite solutions. Any pair of matrices  $L, R$  such that the image space of  $R$  is in the kernel of  $L$  is a solution. The pair  $L = \begin{pmatrix} -2 & 1 \\ 2 & -1 \end{pmatrix}$  and  $R = \begin{pmatrix} 1 & -2 \\ 2 & -4 \end{pmatrix}$  is an example. The image space of  $R$  is in the kernel of  $L$ . Of course, matrix  $R$  must be singular, otherwise its image space would invade the whole  $\mathbb{R}^2$  and  $\text{Ker}(A) = \{0\}$ , which would require matrix  $L = 0$ .

As discussed in example 2, in general, we need the satisfaction of monomial equations that also involve  $x \in \mathbb{R}^n$ .

*Example 4* Suppose we are given  $T = (x, L, R)$  where  $L = \begin{pmatrix} 2 & -2 \\ 2 & -2 \end{pmatrix}$  and  $R = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}$ . We can promptly see that  $L^2 = R^2 = 0$ , and  $[L, R] = LR - RL = 0$ . The last one comes out in any case in which  $R = \alpha L$ , with  $\alpha \in \mathbb{R}$  (here  $\alpha = 1/2$ ). We can immediately conclude that any pair  $(L, R)$ , where  $L^2 = 0$  and  $R = \alpha L$  correspond with a balanced tree composed of three nodes.



Notice that in order to define the formal correspondence with this non-void balanced tree, we need to restrict to the condition  $x \notin \text{Ker } L$ . On the opposite, if we choose  $x = \beta(1, 1)'$  with  $\beta \in \mathbb{R} \setminus \{0\}$ , then the triple represents a tree composed of the root only. If  $x = 0$ , then the triple degenerates to one of the infinite representations of the void tree.

Now, let us consider the problem of mapping the above tree in the representation  $(L, x, R)$ . We need to match the labels  $\text{root}(T)$ ,  $y(L)$  and  $y(R)$ . Hence, we must impose:

$$Cx = \text{root}(T), \quad CLx = y(L), \quad CRx = y(R).$$

Since  $R = \alpha L$ , we have  $y(R) = \alpha CLx = \alpha y(L)$ . This clearly indicates that while the representation  $(L, x, \alpha L)$  is a balanced tree, there is a strong restriction on the label that it can produce.

**Paths and monomial correspondence.** The discussion on the representation of trees in the real field given in the previous examples enlightens on a nice connection between paths and monomials. In order to decode a certain node, we generally need to associate nodes with monomials like

$$L, \quad R, \quad L^2, \quad LR, \quad RL, \quad R^2, \quad L^3, \quad L^2R, \quad RL^2, \quad R^3, \quad LRL, \quad RLR, \dots$$

composed with the two variables  $L$  and  $R$ . This kind of monomials turn out to be just another way of expressing a path in a tree. The above monomial is of degree 3, but we are interested in monomials of any order, which can be represented by the language generated with symbols  $L$  and  $R$ . For instance, the sequence

$$LRLLRLLLRLRLRLR = (LR) \cdot (L^2) \cdot (R^1) \cdot (L^2)^2 \cdot (R^2) \cdot (LR)^3$$

is a way of constructing a monomial with  $L$  and  $R$  that could also be regarded as an element of the language generated by  $S_1 = R$ ,  $S_2 = L^2$ ,  $S_3 = LR$ . These monomials can be described as follows. Let  $\ell_\nu$  and  $r_\nu$  be the integer vectors that count the repetitions of  $L$  and  $R$  in the sequence, respectively. In the above sequence, we have

$$\begin{aligned}\ell^\nu &= (1, 2, 4, 1, 1, 1) \\ r^\nu &= (1, 1, 2, 1, 1, 1).\end{aligned}$$

This notation makes it possible to express the sequence as

$$\pi^\nu = LRLLRLLLRLRLRLR := L^{(1,2,4,1,1,1)} R^{(1,1,2,1,1,1)} = L^{\ell^\nu} R^{r^\nu},$$

where we assume that the above path characterizes node  $\nu$ . Consistently with what we have done so far, we will indicate the label on the node  $\nu$  with the notation  $y(\pi^\nu) \in \mathcal{Y}$ . Here, the notations  $L^{\ell^\nu} R^{r^\nu}$  reminds us of a generalized notion of matrix power for the matrices  $L$  and  $R$ . The notation used for  $\pi^\nu$  reminds the characterization of the node  $\nu$ , while the generic arc of the path  $\pi^\nu$  is simply an element  $\pi_\kappa^\nu$  of vector  $\pi^\nu$ . Moreover, we also use the notation  $|\ell^\nu| = \sum_\kappa \ell_\kappa^\nu$  and  $|r^\nu| = \sum_\kappa r_\kappa^\nu$ . Clearly  $|\pi^\nu| = |\ell^\nu| + |r^\nu|$ .

Example 4 gives an insight into draw the following general conclusion

**Proposition 2** *Let  $\alpha \in \mathbb{R}$  and  $R = \alpha L$  be. Moreover, let us assume that  $h \in \mathbb{N}$ ,  $h \geq 1$  is the first integer such  $L^h = 0$ . If  $x \notin \text{Ker } L^{h-1}$ , then the decoding of the triple  $T = (L, x, R)$  is a balanced tree  $T$  with height  $h$ .*

*Proof.* The proof can be given straightforwardly by induction on  $h$ .

□

The possible generation of infinite trees raises the question on which conditions we need to impose in order to gain the guarantee that a given representation yields finiteness. In addition to the condition stated in Proposition 2, in the next section we will present another class of representations which gives rise to the finite tree. The following proposition states a general property on the generation of “vanishing trees.”

**Proposition 3** Given  $T = (L, x, R)$ , let us assume that  $\|L\| < 1$ ,  $\|R\| < 1$ . Then, if  $\nu$  is a leaf of path  $\pi^\nu$

$$\lim_{|\pi| \rightarrow \infty} L^{\ell^\nu} R^{r^\nu} x = 0.$$

*Proof.* We have

$$y(\pi^\nu) = CL^{\ell^\nu} R^{r^\nu} x.$$

When taking the norm on both sides

$$\|y(\pi^\nu)\| \leq \|C\| \cdot \|L^{\ell^\nu} R^{r^\nu}\| \cdot \|x\|$$

Now, let  $\theta < 1$  be an upper bound of  $\|L\|$  and  $\|R\|$ . Then

$$\|y_\nu\| \leq \|C\| \cdot \|x\| \cdot \theta^{|\pi^\nu|}.$$

Finally, the proof follows when computing  $\lim_{|\pi| \rightarrow \infty}$ .

□

We are now ready to formulate the decoding problem in its general form.

**Decoding Problem.** Given the tree  $T$  with  $m$  nodes, we consider the equations

$$\begin{cases} CL^{\ell^\nu} R^{r^\nu} x = y(\pi^\nu) & \text{for all nodes } \nu; \\ CL^{\ell^\beta} R^{r^\beta} x = 0 & \text{for all buds } \beta, \end{cases}$$

which refer to the nodes and to the buds, respectively (remember that a binary tree with  $m$  nodes has  $m + 1$  buds). When using the vectorial form, we can rewrite this conditions in the form  $Mx = y$  where

$$M := \begin{pmatrix} CL^{\ell^1} R^{r^1} \\ CL^{\ell^2} R^{r^2} \\ \vdots \\ CL^{\ell^m} R^{r^m} \\ CL^{\ell^{m+1}} R^{r^{m+1}} \\ \vdots \\ CL^{\ell^{2m+1}} R^{r^{2m+1}} \end{pmatrix} \quad \text{and} \quad y := \begin{pmatrix} y(\pi^1) \\ y(\pi^2) \\ \vdots \\ y(\pi^m) \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

**Definition 1** The representation  $(L, x, R)$  of  $T$  is completely reachable if and only if  $\text{rank } M = \min\{n, p \cdot (2m + 1)\}$ .

**Proposition 4** Let us consider any completely reachable pair  $(L, R)$  of  $\mathbb{T}$ . If  $n \geq p \cdot (2m + 1)$ , then the decoding problem of  $\mathbb{T}$  admits the solution

$$x = M^+y,$$

where  $M^+$  is Penrose pseudo-inverse of  $M$ .

### 17.3 Non-commutative Left-Right Matrices

As we have already seen,  $T$  can yield an infinite tree. Here is another example.

*Example 5* Let us consider the triple  $T = (L, x, R)$  where  $L = \begin{pmatrix} 2 & -1 \\ 4 & -2 \end{pmatrix}$  and  $R = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}$ . We can promptly see that  $L^2 = 0$  and  $R^2 = 0$ , but  $[L, R] \neq 0$ . In particular,

$$[L, R] = \begin{pmatrix} 2 & -1 \\ 4 & -2 \end{pmatrix} \cdot \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} - \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 2 & -1 \\ 4 & -2 \end{pmatrix} = \begin{pmatrix} 3 & -2 \\ 4 & -3 \end{pmatrix}$$

We can easily check that the recursive propagation yields an infinite tree.

No matter whether a finite or an infinite tree is generated, a uniform representation  $(T, \gamma)$  is especially interesting whenever  $[L, R] \neq 0$ . In the opposite case, as already seen, the representation is dramatically limited. The following example suggests to consider a nice class of uniform non-commutative representations. The following example shows a representation  $(L, x, R)$  which yields finite trees.

*Example 6* Let us consider the triple  $T = (L, x, R)$  where

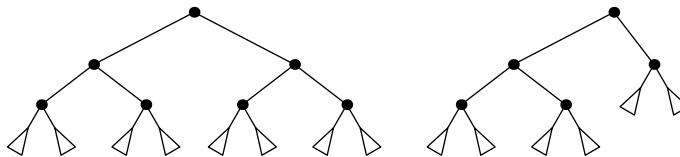
$$L = \begin{pmatrix} 0 & 0 & 0 \\ b_l & 0 & 0 \\ a_l & c_l & 0 \end{pmatrix} \quad R = \begin{pmatrix} 0 & 0 & 0 \\ b_r & 0 & 0 \\ a_r & c_r & 0 \end{pmatrix}$$

Let  $a_l, b_l, c_l, a_r, b_r, c_r$  be non-null reals and associate any non-null real with symbol  $\odot$ . Then, we have

$$\begin{aligned} |\pi^\nu| = 2 \rightarrow \pi^\nu &= \begin{pmatrix} 0 & 0 & 0 \\ \odot & 0 & 0 \\ \odot & \odot & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 0 \\ \odot & 0 & 0 \\ \odot & \odot & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \odot & 0 & 0 \end{pmatrix} \\ |\pi^\nu| = 3 \rightarrow \pi^\nu &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ \odot & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & 0 \\ \odot & 0 & 0 \\ \odot & \odot & 0 \end{pmatrix} = 0 \end{aligned}$$

This corresponds with the balanced tree in Fig. 17.1.

Now, we can exploit the non-commutativity  $[L, R] \neq 0$  to generate other trees with missing nodes. We easily see that if  $b_r = 0$  then  $RL = R^2 = 0$  (see Fig. 17.1).



**Fig. 17.1** Balanced tree on the left in the case of non-null coefficients. If  $b_r = 0$  then the asymmetry yields the unbalanced tree on the right

## 17.4 Conclusions

The encoding-decoding scheme presented in this paper opens the doors to new learning algorithms that seem to be adequate in computational linguistics. A different path that may be followed is one of restricting to commuting matrices where different matrices are used for any layer.

**Acknowledgements** We thank Ilaria Cardinali for insightful discussions.

## References

1. Bengio, Y., Frasconi, P., Simard, P.: Learning long-term dependencies with gradient descent is difficult. *IEEE Trans. Neural Netw.* **5**(2), 157–166 (1994). Special Issue on Dynamic Recurrent Neural Networks
2. Frasconi, P., Gori, M., Sperduti, A.: A general framework for adaptive processing of data structures. *IEEE Trans. Neural Netw.* **9**(5), 768–786 (1998)
3. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
4. Knuth, D.E.: *The Art of Computer Programming*, Volume 1: Fundamental Algorithm, 3rd edn. Addison-Wesley (1997)

# Chapter 18

## Graded Possibilistic Meta Clustering



Alessio Ferone and Antonio Maratea

**Abstract** Meta clustering starts from different clusterings of the same data and aims to group them, reducing the complexity of the choice of the best partitioning and the number of alternatives to compare. Starting from a collection of single feature clusterings, a graded possibilistic medoid meta clustering algorithm is proposed in this paper, exploiting the soft transition from probabilistic to possibilistic memberships in a way that produces more compact and separated clusters with respect to other medoid-based algorithms. The performance of the algorithm has been evaluated on six publicly available data sets over three medoid-based competitors, yielding promising results.

### 18.1 Introduction

In parallel with the increasing amount of unlabeled data produced every day, clustering algorithms become more important to make these data manageable and interpretable. Clustering algorithms group objects according to a similarity measure, without *a priori* information on their true membership and produce a limited number of *centroids* that synthesize the whole data set, representing one of its possible partitions. The first problem is computational complexity: given a data set  $X$  with  $|X| = n$ , the number of possible partitions is  $2^n$ ; hence, the exhaustive search for the optimal partition is unfeasible; the second is how to validate the obtained partition in the absence of a ground truth. These problems make clustering ill-defined [20] and application dependent [12]. Different partitions can be obtained employing different algorithms or employing the same algorithm with different parameters or initialization [5]: each algorithm produces an optimal partition with respect to its specific

---

A. Ferone (✉) · A. Maratea  
Department of Science and Technologies,  
University of Naples “Parthenope”, 80143 Naples, Italy  
e-mail: [alessio.ferone@uniparthenope.it](mailto:alessio.ferone@uniparthenope.it)

A. Maratea  
e-mail: [antonio.maratea@uniparthenope.it](mailto:antonio.maratea@uniparthenope.it)

criterium, but when such a criterium cannot be specified in advance, different results are equally plausible.

*Meta clustering* (namely a clustering of clusterings) aims to obtain a grouping of partitions starting from a number of different clusterings of the same data [1]. To this purpose, given different clusterings (partitions) of the same data, a similarity matrix among clusters (not objects) is computed, then the partitions are clustered according to their similarity with a relational clustering algorithm, as they were simple objects. A *meta cluster* is a collection of similar partitions that are grouped and can be analyzed together.

In the following, the use of a graded probabilistic approach [6] to meta clustering is proposed, in order to address important real-life data challenges, such as overlapping clusters, outliers contamination and uncertainty in memberships. To the best of our knowledge, this is the first attempt to tackle the meta clustering problem by means of a relational medoid-based clustering. The proposed approach has been compared with three leading medoid-based algorithms, namely Hard  $c$ -medoids, Fuzzy  $c$ -medoids and Rough  $c$ -medoids.

## 18.2 Meta Clustering

To test the meta clustering algorithms presented in this paper, first a collection of baseline clusterings of the same data, then a similarity measure for each pair of clusterings and finally a grouping criterium are required.

### 18.2.1 Baseline Clusterings

One-dimensional fuzzy  $c$ -means clustering [3, 18] has been used to generate the baseline clusterings that is the collection of different clusterings of the same data to be clustered by the meta clustering algorithm. For each data set with  $d$  features,  $d$  different clusterings are obtained (each data set has one different clustering with respect to the values of each feature). Analyzing separately, the one-dimensional clusterings can be useful to better understand the structure of the features and to assess whether or not it make sense to cluster with respect to a specific feature. Each baseline clustering is then used as an input for the meta clustering process.

### 18.2.2 Clusterings Similarity

A distance or similarity measure [21] for each pair of clusterings is required to evaluate the concordance of the partitions. The straightforward way to measure the similarity between two clusterings consists in counting the pairs of objects assigned

to the same cluster in both clusterings that is the Rand Index [19]. Let  $X$  be a finite set with  $|X| = n$ . A clustering  $\mathcal{C} = \{C_1, \dots, C_n\}$  is a set such that  $C_i \neq \emptyset \forall i$ ,  $C_i \cap C_j = \emptyset \forall i, j$  and  $\bigcup C_i = X$ . The set of all clusterings of  $X$  is the power set of  $X$  ( $\mathcal{P}(X)$ ). Let  $\mathcal{C}$  and  $\mathcal{C}' \in \mathcal{P}(X)$  two clusterings of  $X$ . The contingency matrix  $M = m_{ij}$  of  $\mathcal{C}$  and  $\mathcal{C}'$  is a  $|\mathcal{C}| \times |\mathcal{C}'|$  matrix where  $m_{ij} = |C_i \cap C'_j|$ .

Each pair of objects of  $X$  can be classified in one of the following four sets:

$$S_{11} = \{\text{pairs that are in the same cluster under } \mathcal{C} \text{ and } \mathcal{C}'\};$$

$$S_{00} = \{\text{pairs that are in different clusters under } \mathcal{C} \text{ and } \mathcal{C}'\};$$

$$S_{10} = \{\text{pairs that are in the same cluster under } \mathcal{C} \text{ but in different ones under } \mathcal{C}'\};$$

$$S_{01} = \{\text{pairs that are in different clusters under } \mathcal{C} \text{ but in the same under } \mathcal{C}'\}.$$

The Rand Index [19] is:

$$\mathcal{R}(\mathcal{C}, \mathcal{C}') = \frac{2(n_{11} + n_{00})}{n(n - 1)} \quad (18.1)$$

where  $n_{ab} = |S_{ab}|$  with  $a, b \in \{0, 1\}$ .  $\mathcal{R}$  ranges from 0, when no pairs belong to the same cluster, to 1 when all the pairs belong to the same cluster (i.e., the clusterings are identical). Although the value of  $\mathcal{R}$  depends on both the number of clusters and the number of elements, in [16] authors showed that it highly depends upon the former. Moreover, in [8] authors show that, in case of independent clusterings, as the number of clusters increases, the Rand Index converges to 1. In [10], Hubert and Arabie proposed a modified version of the Rand Index, called Adjusted Rand Index, that is the normalized difference of the Rand Index and its expected value under the null hypothesis. The Adjusted Rand Index is defined as follows:

$$\mathcal{R}_{\text{adj}}(\mathcal{C}, \mathcal{C}') = \frac{\sum_{i=1}^k \sum_{j=1}^l \binom{m_{ij}}{2} - t_3}{\frac{1}{2}(t_1 + t_2) + t_3} \quad (18.2)$$

where

$$t_1 = \sum_{i=1}^k \binom{|C_i|}{2}, t_2 = \sum_{j=1}^l \binom{|C'_j|}{2}, t_3 = \frac{2t_1 t_2}{n(n - 1)} \quad (18.3)$$

Also this index ranges from zero to one. In the following, the Adjusted Rand Index has been used as the measure of similarity between clusterings.

### 18.2.3 Clustering Partitions by Relational Clustering

When only the pairwise similarities between observations are available, the so-called *relational clustering* is required. It has two advantages: it does not require the preliminary computation of pairwise similarities among all data, resulting faster, and it can be used with data having different domains and sample spaces, as long as their similarities are available. Data are clustered completely ignoring their attributes and

sample space, once their similarity matrix is known. Partitive algorithms are usually preferred to hierarchical ones due to their lower computational complexity. In [6], a partitive relational clustering has been tested on biological data. Let  $X^{n \times q}$  be the data matrix where  $n$  is the number of observations and  $q$  the number of variables; let  $\mathbf{x}_j$  be the row vector corresponding to a single observation and  $c$  be the number of clusters. In relational clustering  $X^{n \times q}$  is unknown and only the similarity matrix  $R^{n \times n}$  is available. Hathaway and Bezdek [9] removed the prototypes from the objective function of fuzzy  $c$ -means, obtaining the *Relational Fuzzy c-Means* (RFCM). Its objective function follows:

$$\min_M \sum_{i=1}^c \frac{\sum_{j=1}^n \sum_{k=1}^n \mu_{ij}^f \mu_{ik}^f r_{jk}}{2 \sum_{t=1}^n \mu_{it}^f} \quad (18.4)$$

where  $M^{n \times c}$  is the matrix of all memberships,  $\mu_{ij}$  is the membership of  $\mathbf{x}_j$  to cluster  $i$ ,  $f$  is the fuzzifier and  $r_{jk}$  is the euclidean distance.

The *medoid* of a cluster is defined as the data sample which is closest to its center. Compared to the mean, the medoid is an actual sample; whereas, the mean in general is a value that does not correspond to any of the sampled data, and that can be even far from them. In the following, three relational medoid-based clustering algorithms are presented.

### 18.2.3.1 The Hard $c$ -Medoids

The Hard  $c$ -Medoids (HCMdd from now on) [11] is essentially a  $k$ -means algorithm modified in terms of medoids. First  $c$  samples are chosen randomly, then the samples are assigned to one of the  $c$  clusters based on the maximum value of the similarity measure  $d(\mathbf{x}_j, \mathbf{c}_i)$  between the sample  $\mathbf{x}_j$  and the medoid  $c_i$ . After the assignment of all the samples to the various clusters, the new medoids are calculated as follows:

$$\mathbf{c}_i = \mathbf{x}_z \quad \forall i \in [1, \dots, c] \quad (18.5)$$

where  $z$  is given by the following formula:

$$\operatorname{argmin}_z \left( \sum_{\mathbf{x}_k \in \beta_i} d(\mathbf{x}_z, \mathbf{x}_k) \right) \quad (18.6)$$

### 18.2.4 Fuzzy $c$ -medoids

Fuzzy  $c$ -medoids (FCMdd from now on) can be seen as a fuzzification of the HCMdd [13], or as a modification of the FCM in terms of medoids. It has the following objective function:

$$J = \sum_{j=1}^n \sum_{i=1}^c (\mu_{ij})^f \{d(\mathbf{x}_j, \mathbf{c}_i)\} \quad (18.7)$$

where  $1 \leq f < \infty$  is the fuzzifier, and  $\mu_{ij} \in [0, 1]$  is the fuzzy membership of the object  $\mathbf{x}_j$  in cluster  $\beta_i$ .

### 18.2.5 Rough $c$ -Medoids

In the Rough  $c$ -Medoids (RCMdd from now on) [17], each cluster  $\beta_i$  is considered a rough set, so it is necessary to define its lower and upper approximations.  $\langle \beta_i, \overline{\beta_i} \rangle$  for every set  $\beta_i \subset U$ , where  $U$  is the universe. Upper and lower approximations are required to follow some of the basic rough set properties, such as:

1. an object  $\mathbf{x}_j$  can be part of at most one lower approximation;
2.  $\mathbf{x}_j \in \underline{\beta_i} \Rightarrow \mathbf{x}_j \in \overline{\beta_i}$
3. an object  $\mathbf{x}_j$  that is not a part of any lower approximation  $\Rightarrow \mathbf{x}_j$  belongs to two or more upper approximations.

An insight on what is to be considered a medoid in this case is necessary. For each cluster, the medoid is calculated first computing the pairwise distance among all samples in the lower approximation, then computing the pairwise distance among all samples in the boundary. Samples in the lower approximations should weight more than the samples in the boundary, so a weighted average with weights  $w$  and  $\tilde{w} = 1 - w$ , with  $w > \tilde{w}$ , is used to find the most central sample. The medoid for RCMdd is given by:

$$\mathbf{c}_i = \mathbf{x}_z \quad (18.8)$$

where  $z$  is given by the following formula:

$$\operatorname{argmin}_z \begin{cases} w \times A + \tilde{w} \times B & \text{if } \underline{\beta_i} \neq \emptyset \text{ and } \overline{\beta_i} \neq \emptyset \\ A & \text{if } \underline{\beta_i} \neq \emptyset \text{ and } \overline{\beta_i} = \emptyset \\ B & \text{if } \underline{\beta_i} = \emptyset \text{ and } \overline{\beta_i} \neq \emptyset \end{cases} \quad (18.9)$$

where

$$\mathcal{A} = \sum_{\mathbf{x}_k \in \underline{\beta_i}} d(\mathbf{x}_k, \mathbf{x}_z); \quad \mathcal{B} = \sum_{\mathbf{x}_k \in BND(\beta_i)} d(\mathbf{x}_k, \mathbf{x}_z); \quad (18.10)$$

$w$  and  $\tilde{w}$  represent the relative importance of lower approximation and boundary region.

### 18.2.6 Graded Possibilistic $c$ -medoids

In the  $c$ -means family, centroids are obtained through formula (18.11), independently from the chosen distance or membership function:

$$\mathbf{c}_i = \frac{\sum_{j=1}^n \mu_{ij} \mathbf{x}_j}{\sum_{j=1}^n \mu_{ij}} \quad \forall i \in [1, \dots, c] \quad (18.11)$$

where  $n$  is the number of observations,  $\mu_{ij}$  is the membership value of observation  $\mathbf{x}_j$  to cluster  $i$ , and  $c$  is the number of clusters.

The core observation of GPC is that many clustering algorithms can be obtained changing the constraint on the sum of membership for all  $\mathbf{x}_j$  that is the value of  $\zeta_i$  in formula (18.12).

$$\zeta_j = \sum_{i=1}^f \mu_{ij} \quad \forall j \in [1, \dots, n] \quad (18.12)$$

Changing the constraint, the two extreme cases are standard fuzzy clustering, obtained when the sum (18.12) is exactly one, and possibilistic clustering, obtained when the sum (18.12) can be greater than one.

Graded possibilistic  $c$ -medoids (GPCMdd) is a modification of the graded possibilistic  $c$ -means in terms of medoids. An insight on what is to be considered a medoid also in this case is necessary. For each cluster, the medoids are computed similarly to FCMdd, with the difference that the memberships can become possibilistic.

$$\mathbf{c}_i = \mathbf{x}_z \quad (18.13)$$

where  $z$  is given by:

$$\operatorname{argmin}_z \sum_{\mathbf{x}_k \in \beta_i} (u_{ik})^f d(\mathbf{x}_k, \mathbf{x}_z); \quad 1 \leq j \leq n \quad (18.14)$$

With respect to FCMdd, GPCMdd allows the soft transition from probabilistic to possibilistic membership functions; hence, it is able to represent effectively loosely related data and it is more robust to outliers and noise.

## 18.3 Experiments

In the absence of ground truth, clusters can only be evaluated in terms of internal validity measures that consider their shape, compactness and separation. In order to evaluate the performance of the proposed relational medoid-based soft clustering algorithms for meta clustering, GPCMdd has been compared with HCMdd, FCMdd

and RCMdd on six real world data sets, using three metrics: the Dunn index  $D$ , the Davies-Bouldin index  $DB$  and the Compactness Index  $CI$ .

### 18.3.1 Data

The six publicly available [14] data sets reported in Table 18.1 have been considered. They come from different domains and have from nine to 60 features to be used for baseline clusterings.

### 18.3.2 Performance Measures

A well-known internal criterium is the Dunn index  $D$  [4] that compares the minimum inter-cluster distance and the maximum intra-cluster distance (it has been used for example in [15]). Let  $d_m$  and  $d_M$  be the minimum inter-cluster distance and the maximum intra-cluster distance, respectively, then:

$$D = \frac{d_m}{d_M} \quad (18.15)$$

When  $D$  is large, clusters are well separated.

Another well-known internal criterium, similar to  $D$ , is the Davies-Bouldin index  $DB$  index [2] that computes the average ratio between the within cluster distances and the between cluster distances.

$$DB = \frac{1}{c} \sum_{h=1}^c \max_{h \neq k} \frac{\Delta(C_h) + \Delta(C_k)}{\delta(\mathbf{c}_h, \mathbf{c}_k)} \quad \forall h, k \in [1, \dots, c] \quad (18.16)$$

where  $c$  is the number of clusters,  $C_h$  is the cluster  $h$ ,  $\Delta(C_h)$  is the diameter of cluster  $C_h$  and  $\delta(\mathbf{c}_h, \mathbf{c}_k)$  is the distance between centroids  $\mathbf{c}_h$  and  $\mathbf{c}_k$ .

When  $DB$  is small, clusters are well separated.

**Table 18.1** Datasets

Data set	#Feature	#Objects
Glass	9	214
Olitos	25	120
Sonar	60	208
Water2	39	390
Wdbc	9	683
Wine	13	178

An internal criterium suitable for meta clustering [1] is the compactness index  $CI$  that computes the average pairwise distance between points in the same cluster.

$$CI = \frac{\sum_{i=1}^c N_i \frac{\sum_{j=1}^{N_i-1} \sum_{k=j+1}^{N_i} d_{jk}}{N_i(N_i-1)/2}}{N} \quad (18.17)$$

where  $c$  is the number of clusters;  $N_i = |C_i|$  is the cardinality of the  $i$ -th cluster;  $d_{jk}$  is the distance between samples  $j$  and  $k$ , and  $N = \sum_{i=1}^c N_i$ .

When  $CI$  is small, clusters are more compact.

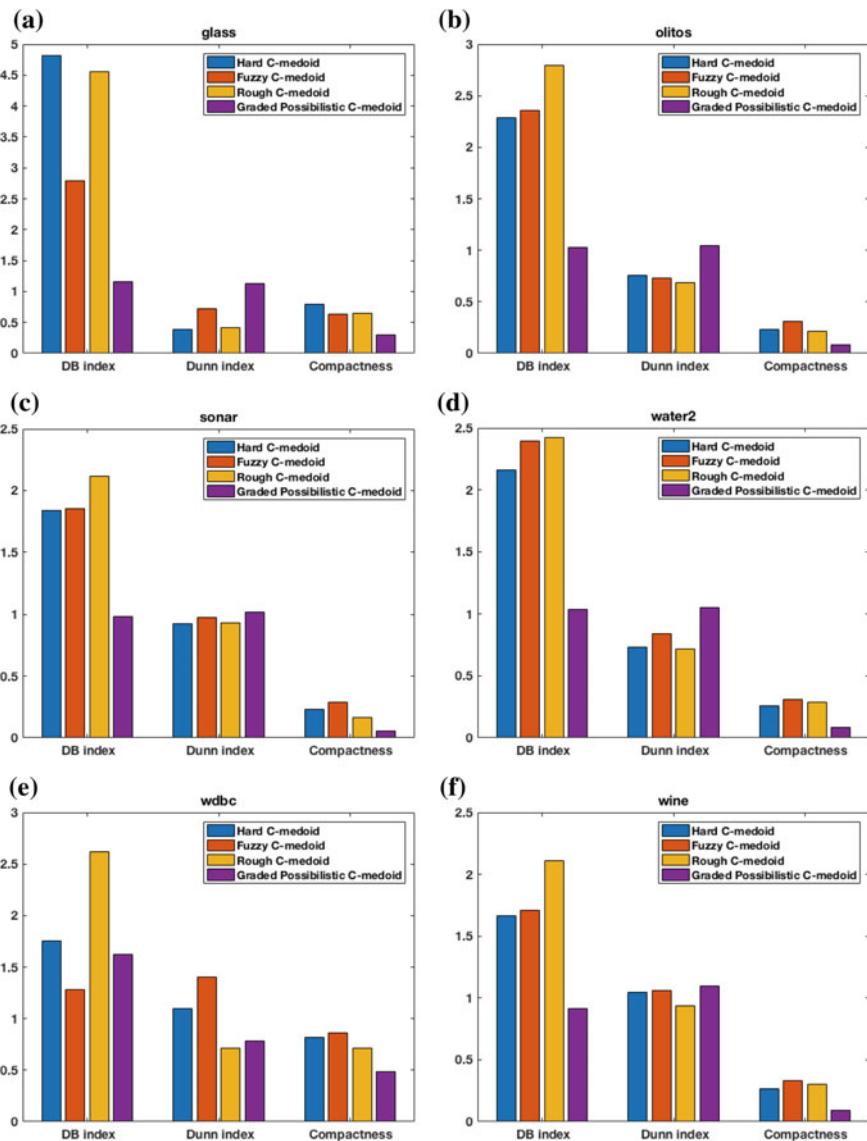
### 18.3.3 Results and Discussion

Figure 18.1 shows the Dunn, DB and CI indices for the graded possibilistic  $c$ -medoid, Hard  $c$ -medoid, Fuzzy  $c$ -medoid and Rough  $c$ -medoid algorithms on the tested data sets. Results are averaged over 100 runs.

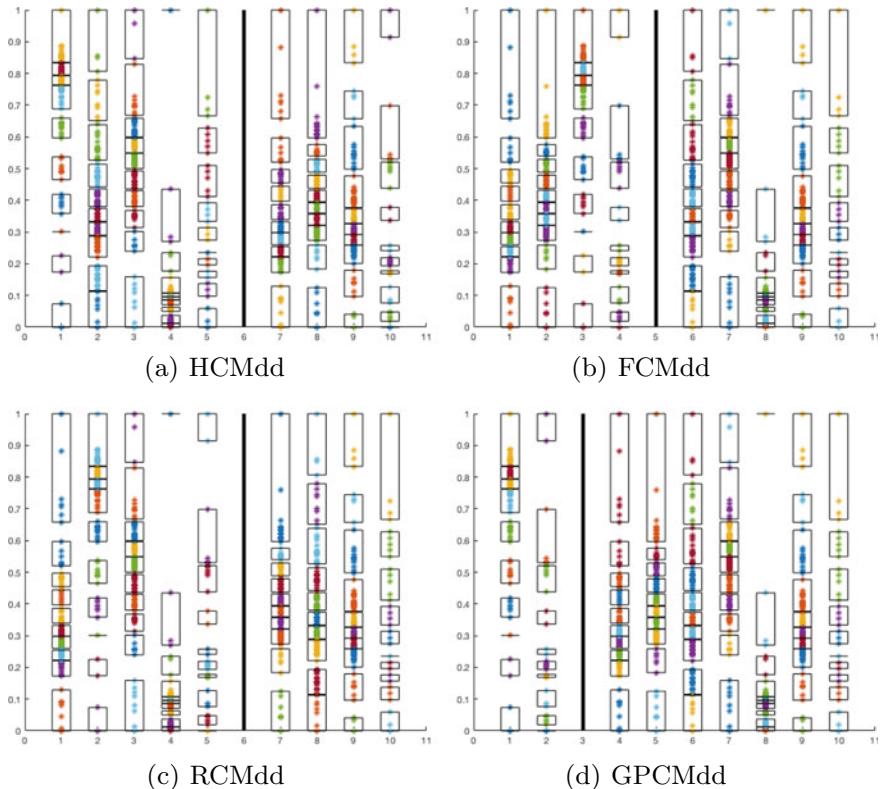
Overall, GPCMdd always performs better than the other algorithms, producing both compact and well-separated clusters. This is not surprising given that possibilistic memberships are particularly useful in real-life applications to better handle noisy patterns, outliers contamination, uncertainty in memberships and overlapping partitions.

Nevertheless, in the considered application, the values of  $DB$  and  $D$  suggest that clusters generally tend to be close to each other. In this situation, it is crucial the compactness of clusters that can mitigate their relative closeness. Indeed, considering only the compactness index, GPCMdd performs way better than the other algorithms yielding very compact clusters. This is due to the way in which the mixed probabilistic and possibilistic memberships are handled by the algorithm that gradually pushes them toward the final values. In particular, the latter result is quite interesting in the case of similar clusterings that would be naturally grouped in similar partitions characterized by low inter-cluster variability.

In order to visualize these results, Fig. 18.2a–d shows an example of meta clustering on data set Glass, where each bar represents the clustering of a single feature and the black vertical line separates different meta clusters (two in this example). It can be noted how GPCMdd (Fig. 18.2d) is able to group visually similar clusterings (with a similar structure) compared to the other algorithms that yields less compact meta clusters. For instance, consider clusterings 2, 3 and 5 in Fig. 18.2a, clusterings 1 and 2 in Fig. 18.2b and clusterings 1 and 3 in Fig. 18.2c that are separated from other similar clusterings while are put in the same cluster by GPCMdd. The bar 8 in GPCMdd represents an outlier. The results suggest that medoid-based clustering algorithms are effective for meta clustering applications and that graded possibilistic  $c$ -medoid helps in identifying compact clusters even in case of similar partitions.



**Fig. 18.1** DB, Dunn and compactness indexes for data set **a** Glass, **b** Olitos, **c** Sonar, **d** Water2, **e** WDBC, **f** Wine



**Fig. 18.2** Example of meta clustering on data set glass: **a** HCMdd, **b** FCMdd, **c** RCMdd and **d** GPCMdd

## 18.4 Conclusions

The intrinsic ambiguity of the clustering problem produces many plausible solutions for the same data set. Starting from a collection of single feature clusterings, a graded possibilistic medoid meta clustering has been proposed in this paper, grouping most similar clusterings in a way that produces more compact and separated clusters with respect to other medoid-based algorithms. Even though fuzzy c-medoid presents the best overall performance, rough c-medoid is able to produce more compact clusters in the case of similar clusterings characterized by low inter-cluster variability. Ongoing work is in testing fuzzy–rough hybrid approaches [7]. The hybrid notion of fuzzy–rough sets allows to exploit, at the same time, properties like vagueness and coarseness, realizing a more expressive model.

**Acknowledgements** Authors would like to acknowledge the financial support for this research through “Bando di sostegno alla ricerca individuale per il triennio 2015–2017 – Annualità 2017” granted by University of Naples “Parthenope”.

## References

1. Caruana, R., Elhawary, M., Nguyen, N., Smith, C.: Meta clustering. In: Proceedings of the Sixth International Conference on Data Mining. ICDM '06, IEEE Computer Society (2006) 107–118
2. Davies, D.L., Bouldin, D.W.: A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence* **1**(2), 224–227 (1979)
3. Dharan, S., Nair, A.S.: Bioclustering of gene expression data using reactive greedy randomized adaptive search procedure. *BMC Bioinformatics* **10**(Suppl 1), S27 (2009). Jan
4. Dunn, J.: Well-separated clusters and optimal fuzzy partitions. *Journal of Cybernetics* **4**(1), 95–104 (1974)
5. Ferone, A., Galletti, A., Maratea, A.: Variable width rough-fuzzy c-means. In: 2017 13th International Conference on Signal-Image Technology Internet-Based Systems (SITIS). (Dec 2017) 458–464
6. Ferone, A., Maratea, A.: Decoy clustering through graded possibilistic c-medoids. In: 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). (July 2017) 1–6
7. Ferone, A., Petrosino, A.: Feature selection through composition of rough-fuzzy sets. In: International Workshop on Fuzzy Logic and Applications, Springer (2016) 116–125
8. Fowlkes, E.B., Mallows, C.L.: A method for comparing two hierarchical clusterings. *Journal of the American Statistical Association* **78**(383), 553–569 (1983)
9. Hathaway, R.J., Davenport, J.W., Bezdek, J.C.: Relational duals of the c-means clustering algorithms. *Pattern Recognition* **22**(2), 205–212 (1989)
10. Hubert, L., Arabie, P.: Comparing partitions. *Journal of Classification* **2**(1), 193–218 (1985). Dec
11. Kaufman, L., Rousseeuw, P.J.: Finding groups in data : an introduction to cluster analysis. Wiley series in probability and mathematical statistics. Wiley, New York (1990) A Wiley-Interscience publication
12. Kleinberg, J.M.: An impossibility theorem for clustering. In: Becker, S., Thrun, S., Obermayer, K. (eds.) Advances in Neural Information Processing Systems 15, pp. 446–453. MIT Press (2002). <http://papers.nips.cc/paper/2340-an-impossibility-theorem-forclustering.pdf>
13. Krishnapuram, R., Joshi, A., Nasraoui, O., Yi, L.: Low-complexity fuzzy relational clustering algorithms for web mining. *IEEE Transactions on Fuzzy Systems* **9**(4), 595–607 (2001)
14. Lichman, M.: UCI machine learning repository (2013)
15. Maji, P., Pal, S.K.: Rough set based generalized fuzzy-means algorithm and quantitative indices. *Trans. Sys. Man Cyber. Part B* **37**(6), 1529–1540 (2007)
16. Morey, L., Agresti, A.: The measurement of classification agreement: An adjustment to the rand statistic for chance agreement. " **44** (03 1984) 33–37
17. Peters, G., Lampart, M.: A partitive rough clustering algorithm. In: Greco, S., Hata, Y., Hirano, S., Inuiguchi, M., Miyamoto, S., Nguyen, H.S., Słowiński, R. (eds.) Rough Sets and Current Trends in Computing, pp. 657–666. Springer, Berlin Heidelberg, Berlin, Heidelberg (2006)
18. Pontes, B., Girddez, R., Aguilar-Ruiz, J.S.: Bioclustering on expression data: A review. *Journal of Biomedical Informatics* **57**(Supplement C) (2015) 163 – 180
19. Rand, W.: Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association* **66**(336), 846–850 (1971)
20. Slonim, N., Tishby, N.: Agglomerative information bottleneck. In: Solla, S.A., Leen, T.K., Müller, K. (eds.) Advances in Neural Information Processing Systems 12, pp. 617–623. MIT Press (1999). <http://papers.nips.cc/paper/1651-agglomerative-informationbottleneck.pdf>
21. Wagner, S., Wagner, D.: Comparing clusterings- an overview (2007)

# Chapter 19

## Probing a Deep Neural Network



**Francesco A. N. Palmieri, Mario Baldi, Amedeo Buonanno,  
Giovanni Di Gennaro and Francesco Ospedale**

**Abstract** We report a number of experiments on a deep convolutional network in order to gain a better understanding of the transformations that emerge from learning at the various layers. We analyze the backward flow and the reconstructed images, using an adaptive masking approach in which pooling and nonlinearities at the various layers are represented by data-dependent binary masks. We focus on the field of view of specific neurons, also using random parameters, in order to understand the nature of the information that flows through the activation’s “holes” that emerge in the multi-layer structure when an image is presented at the input. We show how the peculiarity of the multi-layer structure is not so much in the learned parameters, but in the patterns of connectivity that are partly imposed and then learned. Furthermore, a deep network appears to focus more on statistics, such as gradient-like transformations, rather than on filters matched to image patterns. Our probes seem to explain why classical image processing algorithms, such as the famous SIFT, have provided robust, although limited, solutions to image recognition tasks.

---

F. A. N. Palmieri (✉) · M. Baldi · A. Buonanno · G. Di Gennaro · F. Ospedale  
Dipartimento di Ingegneria Industriale e dell’Informazione, Università della Campania  
“Luigi Vanvitelli” (ex SUN), via Roma 29, 81031 Aversa (CE), Italy  
e-mail: [francesco.palmieri@unicampania.it](mailto:francesco.palmieri@unicampania.it)

M. Baldi  
e-mail: [mario.baldi@studenti.unicampania.it](mailto:mario.baldi@studenti.unicampania.it)

A. Buonanno  
e-mail: [amedeo.buonanno@unicampania.it](mailto:amedeo.buonanno@unicampania.it)

G. Di Gennaro  
e-mail: [giovanni.digennaro@unicampania.it](mailto:giovanni.digennaro@unicampania.it)

F. Ospedale  
e-mail: [francesco.ospedale@studenti.unicampania.it](mailto:francesco.ospedale@studenti.unicampania.it)

## 19.1 Introduction

Deep convolutional networks (DCN) have recently shown outstanding performances in pattern recognition on images and signals. The results are attributed mainly to the multi-layer convolutional structure of the network and to a number of “tricks” in the learning algorithms (dropout, unsupervised initialization, etc.) [1]. However, many open questions remain on the reasons why some structures perform so well in supervised learning and how the choice of parameters (such as number of layers, connectivity, stride in the convolutions, dimensionality of the feature spaces, pooling dimensions, type of nonlinearity, etc.) can significantly affect the recognition performances. In fact, most of the successful results have been reached after many lengthy and computationally expensive trials on configurations and learning algorithms. This reveals that there is still a fundamental lack of tools for understanding the basic functioning of the network, even if probes into the network activities [2], and partial analyses [3, 4], have been presented in the literature.

Although more complicated architectures have been proposed, such as LSTM [5] or Wavenets [6], that use computational units formed by decision nodes combined with linear units, we limit ourselves in this work to a structure with standard linear combiners and rectifiers (RELTUs), leaving the study of other paradigms to future works.

In the architecture analyzed in this paper, the peculiar role is played by the RELTUs that provide an unfolding of the input space with the activations, being the essence of how the information is stored at the various layers. We demonstrate that the representation power of a multi-layer network is not so much in the specific parameters, but mostly in the configurations of activations that emerge at the various levels. It has been demonstrated that the hyperplane arrangements at the various stages correspond to a number of linear regions that grow exponentially with the number of layers [7, 8]. This is in contrast to single-layer networks that, even though they have universal approximation capabilities [9], would require a very large number of units.

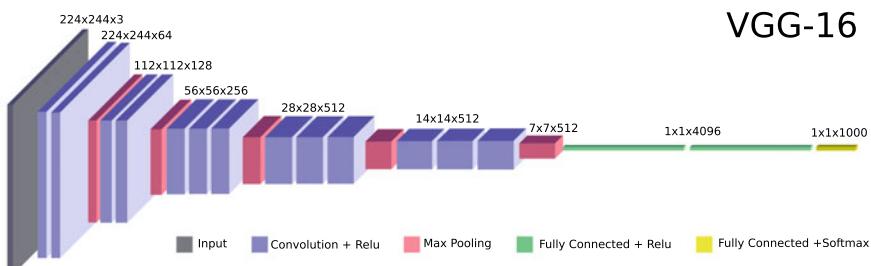
Visualization on actual images can provide many clues, especially when specific activations are propagated backward onto the input image [2], in an attempt to identify the features that may be related to various regions of the network. In this paper, we follow this idea and report some experiments on the backward propagation of the activations at specific neurons in the last layers, using various strategies for the backward flow. Nonlinearities and pooling sections are formalized with data-dependent binary masks that show how the information flows through the “holes” determined by connectivity and activations. Even if we focus on a specific pre-trained network, the VGG-16 [10], the results are to be considered typical as they would apply similarly to other convolutional architectures. We modify the so-called *Deconvnet* process [3], using also random parametrizations. We demonstrate that the information stored in the network is mostly related to the activation configurations, rather than to specific parameters. By looking at the field of view (FOV), we show that the local nonlinearities involved into specific activations, extract gradient-like statistics on image regions, rather than representing the storage of sub-images, as it would be if we had

designed a bank of matched filters. This seems to confirm why in computer vision SIFT-like algorithms [11], that use histogram of gradients (HOGs), were the state-of-the-art before being overcome by adaptive deep networks. The deep networks seem to learn through massive training similar kinds of image statistics.

In Sect. 19.2 we review the multi-layer architecture, and in Sect. 19.3 we describe the experiments with the backward flow. In Sects. 19.4 and 19.5, we compute best input selections and compute some activation statistics. Section 19.6 reports conclusions and points to future research directions.

## 19.2 Multi-Layer Convolutional Networks

A DCN for supervised learning on images is composed by a first part of convolutional layers, intermingled with pooling sections, and by a final stage of a few (at least two) fully connected layers for final classification. Figure 19.1 shows the popular network VGG-16 [10]. The first part, which is sometimes trained using unsupervised algorithms, and subsequently refined with supervised information backpropagated from the last layers, provides the mapping of the image to a feature space that is more prone to be used by the supervised part. It is well known that a classifier directly connected to the image would perform poorly, and that parameter extraction is the crucial component of any pattern recognition system. The understanding of this transformation is crucial for identifying the real peculiarity of the DCN, also in comparison with the many other parameter extraction criteria presented in decades of literature in machine vision. In a DCN at each layer  $i$ , we have a 3D array of size  $N^i \times M^i \times D^i$ ,  $i = 0, \dots, L$ . To fix ideas and numbering, we focus on the VGG-16 in Fig. 19.1 where we have layers 0, 1, 2, (P3), 4, (P5), 6, 7, (P8), 9, 10, (P11), 12, 13, (P), 14, 15, (16S), where (P) indicates pooling, (Pi) pooling + linear+ RELU and (iS) linear+ softmax. In the first 13 layers, we have progressively smaller images and larger depths ( $D^0, \dots, D^{13}$ ) = (3, 64, 64, 128, 128, 256, 256, 256, 512, ..., 512). Max pooling is applied after groups of layers on  $2 \times 2$  non-overlapping masks, causing each time



**Fig. 19.1** Typical VGG-16 multi-layer DCN structure

a reduction of the image size to a half (the max is taken on each feature slice separately). At the end, there are three fully connected layers that loose the spatial dimension  $(N^{14}, N^{15}, N^{16}) = (M^{14}, M^{15}, M^{16}) = (1, 1, 1)$  and have depths  $(D^{14}, D^{15}, D^{16}) = (4096, 4096, 1000)$ . All the linear layers in the first part are convolutional and use patches of size  $3 \times 3$  and stride (overlap) one. Convolutions are applied in such a way that the image size is preserved by padding the outside with zeros at the boundaries. After each convolution, a rectifier (RELU) is applied.

To write compact equations for the network, we reduce the 3D arrays to vectors  $x_i$  of sizes  $n_i = N^i M^i D^i$ ,  $i = 0, \dots, L$  (here  $L = 16$ ). Therefore, a layer without pooling is described by the equation

$$x_i = r_i \odot (W_i x_{i-1} + b_i), \quad (19.1)$$

where  $\odot$  is the Hadamard product (element-by-element),  $W_i$  is an  $n_i \times n_{i-1}$  real matrix, and  $b_i$  is an  $n_i$ -dimensional bias vectors. RELU rectification, i.e.,  $R(x) = \max(0, x)$ , is described by an  $n_i$ -dimensional binary vector  $r_i$  with zeros and ones. Clearly  $r_i$  is data-dependent and can be seen as an activation mask for every instance. A layer with pooling is described instead by the equation

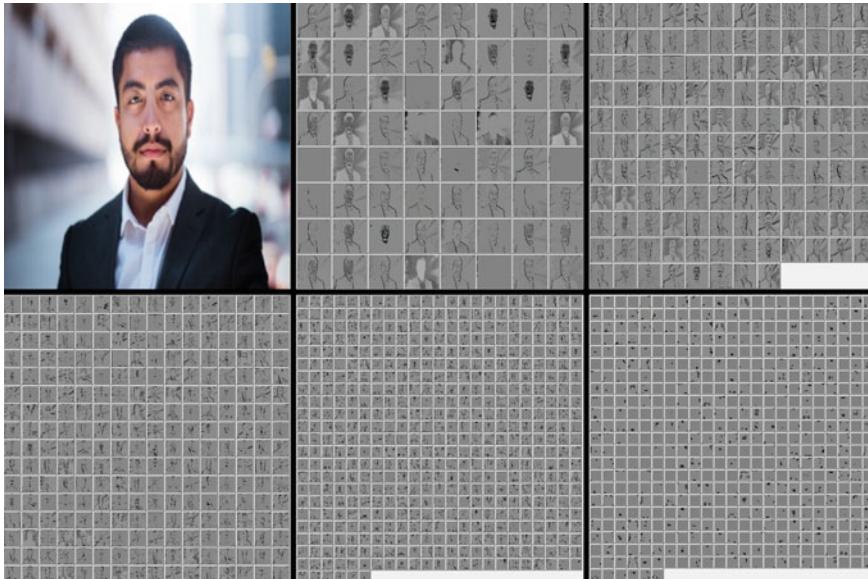
$$x_i = r_i \odot (W_i P_i x_{i-1} + b_i), \quad (19.2)$$

where  $W_i$  is an  $n_i \times n_i$  real matrix, and the pooling operation is represented by a  $n_i \times n_{i-1}$  sub-permutation matrix  $P_i$  (a sub-permutation matrix is such that each entry is either 1 or 0, each row contains only one 1 and each column contains at most one 1). Matrix  $P_i$  is clearly data-dependent and contains information on the position of all the local maxima. The last layer is described by the equation

$$x_L = S(W_L x_{L-1} + b_L), \quad (19.3)$$

where  $S(y)$  is the softmax operation applied to the vector  $y$ .

When an input is presented to the network (layer zero), the information is propagated forward, and only the linear combinations that produce a positive output for the RELUs influence the following stages. Max pooling produces a reduction in size letting through only one out of four activations. The code that emerges at the various layers is very sparse as shown in Fig. 19.2 where after the color input image, we show slices at various depths for layers 1, 4, 7, 9, and 13. The first layer seems to detect edges and elementary local configurations, while in the following stages progressive mapping of the input space at larger scales is obtained. The question is: What is the image feature detected at a specific location in layer 13? To provide a partial answer to this question, we observe that when a pixel in the last layer (13) is active, there is a complex tree of activations that causes that pixel to become alive. An activation tree is shown in Fig. 19.3 on a simplified network where the connections that potentially can contribute to a specific output in the last layer are colored. They identify the field of view (FOV) of that unit. However, for each input, a pattern of activations emerges and only a specific subset of these connections plays a role (circled in Fig. 19.3).

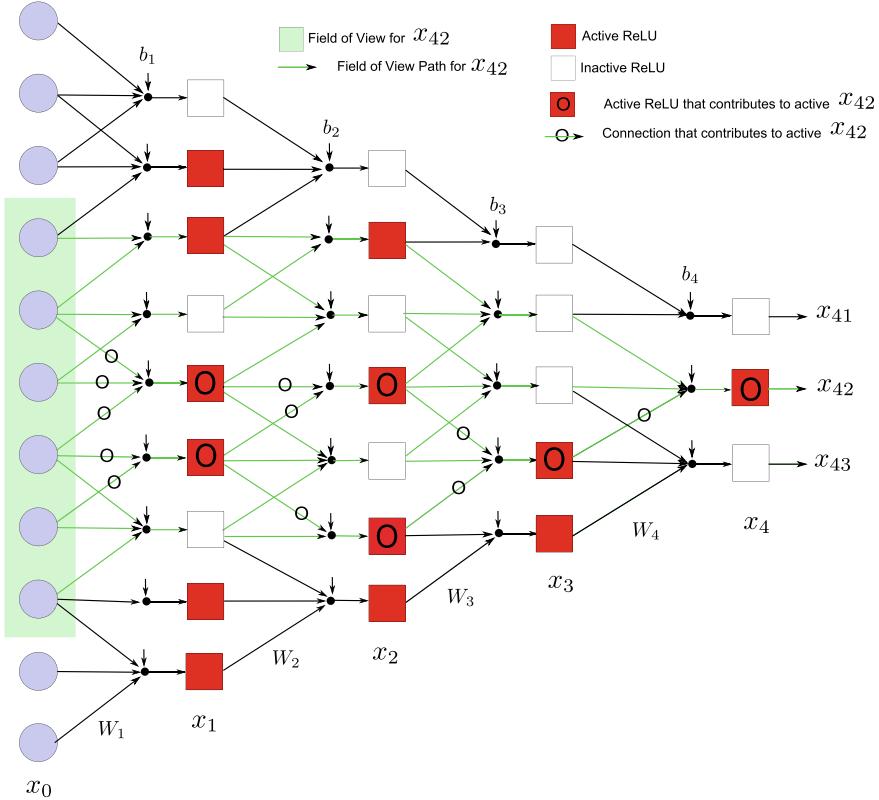


**Fig. 19.2** Activations in the VGG-16 for an image presented at the input. The images represent activations at layers 1, 4 , 7, 9, and 13, respectively

They identify a locally linear (affine) transformation that represents the specific filter for the pattern present in that FOV of the image. Observe that there may be many different configurations within that FOV tree that can cause the final unit to become active. Therefore, the unit at the end is a sort of logic OR of many intermediate configurations. The network performs a hierarchical coding in activating different sets of units in the intermediate layers. The understanding of the structure of the activations is crucial to see the nature of the transformation implemented by the network, which is peculiar to the multi-layer architecture in contrast to a shallow one-layer bank of filters. To shed light on some of these issues, we have performed a number of experiments in backpropagating the information from layer 13 backward to the image plane in the VGG-16.

### 19.3 Backward Reconstructions

**Deconvnet.** In a first set of experiments, after having presented an image to the network and produced the activations in all layers, we have used *Deconvnet*, an algorithm proposed in [4], to obtain the image reconstructed from active outputs. We do not propagate back all the activations, but we choose a specific location in layer 13 and a specific feature. From that element, we propagate  $x_{13}$  backward after setting to zero all the others elements of the vector. Such a vector is denoted with  $x_{13}^b$



**Fig. 19.3** Schematic network where for  $x_{42}$  is shown in the FOV with its connection tree (green). On a specific instance of  $x_0$  in which  $x_{42}$  becomes active, the pictures distinguish among the active ReLUs and the ones that contribute to activate  $x_{42}$ . Also the contributing weights within the FOV are marked with a circle

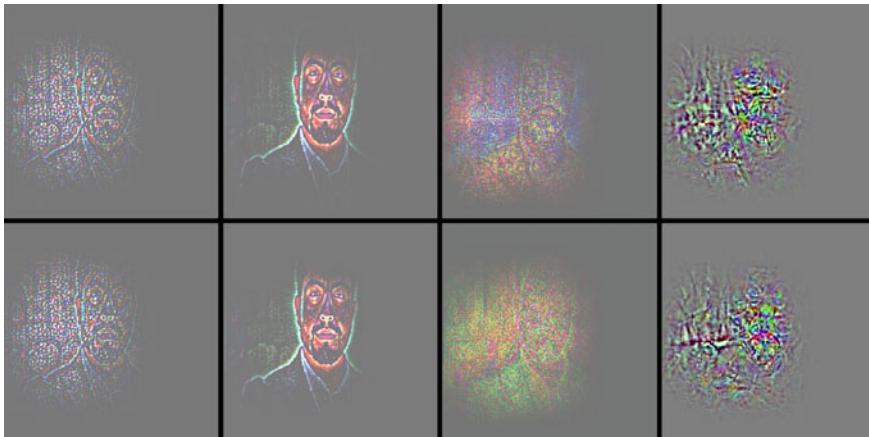
and similarly  $x_i^b$  denote all the intermediate activations in the backward flow. The process at the layer where no pooling is present can be compactly described by the equation

$$x_{i-1}^b = R(W_i^T x_i^b), \quad (19.4)$$

where  $R$  represents the RELU function applied during the backward propagation process, and in the layers with pooling by the equation

$$x_{i-1}^b = P_i^T R(W_i^T x_i^b). \quad (19.5)$$

All the activations are confined to the FOV of the chosen location. Note that the backward operation through the *Unpooling* ( $P_i^T$ ) needs uses local information from the forward flow. Furthermore, the RELU operations applied to the backward flow by no means guarantee that the same units are active in the backward operation



**Fig. 19.4** Results on backward reconstruction from a specific neuron in layer 13 and two different features (upper and lower row). The image is obtained using (col. 1) Deconvnet; (col. 2) Deconvnet with binary masks; (col. 3) Deconvolution with random weights and masks. Column 4 shows the exact filters activated in the FOV

in comparison with the forward flow; this is because the matrices are not orthonormal (the transpose is not the inverse). However, a faded and stylized reconstruction appears anyhow in the image plane within the FOV as in Fig. 19.4 (col. 1), for one location and two different features.

**Deconvnet With Masks.** In an attempt to better separate the roles of weights and activations, we have run the Deconvnet algorithm using also the activations stored from the forward flow, i.e., not relaying only on the RELUs applied to the backward flow. More specifically, if we conserve the binary masks  $r_i$  and the max pooling sub-permutations  $P_i$  for every layer, the backward flow at each layer with no pooling can be described by the equation

$$x_{i-1}^b = R(W_i^T(r_i \odot x_i^b)), \quad (19.6)$$

where  $\odot$  denote the Hadamard product (element-by-element), and in the layers with pooling by the equation

$$x_{i-1}^b = P_i^T R(W_i^T(r_i \odot x_i^b)). \quad (19.7)$$

Figure 19.4 (col. 2) shows the reconstructions revealing that the activation's masks influence greatly this attempted inversion. Vestiges of the image appear more clearly as information makes it better through the “holes” carved by the forward activations. This shows that the essence of the backward reconstruction is heavily based on the activation patterns, rather than on the specifics of the matrix parameters.

**Deconvnet with masks and random parameters.** To investigate further the issue, we have randomized the weights in the backward flow and used only the activation masks. The equations for the reconstructions are again (19.6) and (19.7). To keep values in a reasonable range, weights and biases have been randomly chosen according to a gaussian distribution with means and variances that match the learned network weights. Typical results are shown in Fig. 19.4 (col. 3). It is quite striking to see that an image in the FOV still appears confirming that the patterns of activations may be the most important means by which the image is coded in the network structure! Note that in this case no values are used, but only random numbers going through the masks.

### 19.3.1 Filter Reconstruction

A network made of linear combiners and RELUs implements functions that are piecewise linear (actually affine because of the biases). Every time an input is presented to the network, a region of linearity is involved in the very high-dimensional input space. We know that the number of regions that characterize a multi-layer network grows exponentially with the number of layers [7]. If we focus on one activation in the last layer, it must be then the composition of many linear regions. Now if linear region represents a feature revealed at the end of the chain on that neuron the natural question is: how do these linear regions look like in the input space of the FOV of that neuron? Do they resemble image patterns, or they provide a statistical account of that region?

In our attempt to shed some light on this issue, we have performed a number of reconstruction experiments in a way that are different from the above Deconvnet algorithm and its variations. More specifically, we have taken a delta activation in the last layer (1 for one location and one feature only, and zero else) and, using the masks memorized in the forward flow for a specific image, we have used the following equations in the backward process. At each layer with no pooling, we use the equation

$$x_{i-1}^b = (W_i^T(r_i \odot x_i^b)) \quad (19.8)$$

and in the layers with pooling the equation

$$x_{i-1}^b = P_i^T(W_i^T(r_i \odot x_i^b)). \quad (19.9)$$

The information goes backward through the activations and computes *exactly* the local linear combination activated on that specific instance. The results for two different features from the 13th layer are shown in Fig. 19.4 (col. 4). It is very interesting to look at the results because the activated filter is nothing like an image patch, rather

a very alternating pattern that seems to compute local statistics. This is particularly revealing because, just as in some the feature extractors used in the computer vision such as the SIFT algorithm [11], the filter seems to be computing local gradients. Extensive training of the VGG-16 network has automatically discovered this feature to be relevant for classification.

## 19.4 Best Input Selection

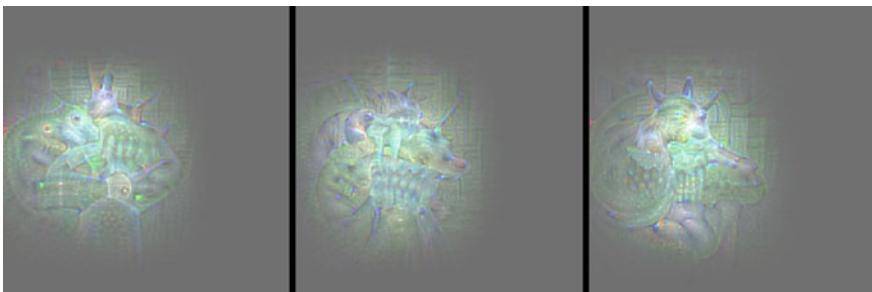
We have already pointed out that the activation of a specific neuron can be seen as the logic OR of possibly very many different image configurations. In order to reveal the kind of patterns that can activate a specific neuron, we have used an optimization algorithm to get a partial answer. More specifically, as suggested in [2], for the pre-trained VGG-16, we search for the image that maximally excites a specific neuron  $x_{13}^i$  in the last layer that is

$$x_0^* = \operatorname{argmax}_{x_0} x_{13}^i. \quad (19.10)$$

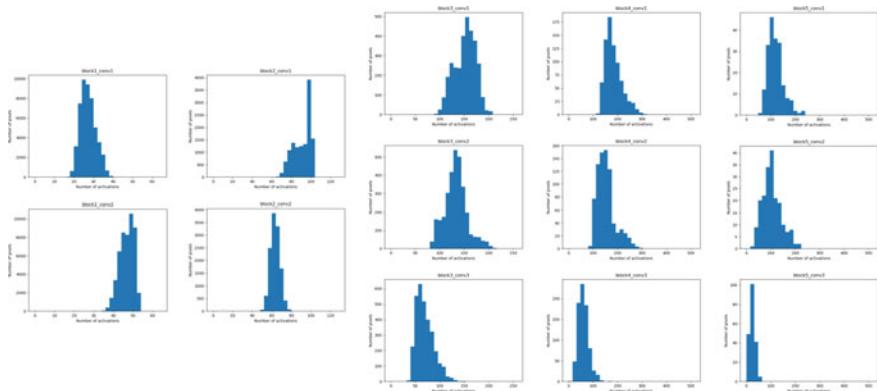
The algorithm starts from a random image, and it is updated according to a simple gradient ascent criterion. The value of  $x_{13}^i$  is backpropagated, just as in standard backpropagation, through the network from the 13th layer to the input. Every time the algorithm is started from a different random image, it converges to a different configuration. Figure 19.5 shows the results of three runs for the same neuron. The pictures show clearly within the FOV of that neuron a filter that does not resemble the image at all, but a rather complex linear transformation. This should be seen as a different way of obtaining a filter as in subsection 19.3.1 without reference to a specific image.

## 19.5 Activation Statistics

In this last section, we have investigated how the various features are activated at various layers in the VGG-16 for a typical image. This analysis has the objective of understanding the approximate cardinality of the coding that emerges in a multi-layer convolutional network. In the first convolutional layer, we see that about a half of the 64 features are activated. The number grows in the second convolutional layer and remains approximately the same in percentage after pooling, to return to about a half in the subsequent layer. After the second pooling, we see that the fraction remains about the same, tending to a more and more sparse code toward the end (Fig. 19.6).



**Fig. 19.5** Images resulting from the maximization of a specific neuron in the 13th layer



**Fig. 19.6** Histograms on the number of features activated at the various layer in the VGG-16 for a typical image

## 19.6 Conclusions and Future Directions

In this work, we have conducted a number of experiments on a deep convolutional neural network with the objective of gaining a better understanding of the inner transformations that are learned through extensive training. The motivation is in the striking performances that these systems provide in image recognition. We have focused on a specific pre-trained network, the VGG-16, and focused on the activations of specific neurons in layer 13 that are located at the end of the feature extraction part.

The experiments carried out in propagating backward the information through the network in various ways, using data-dependent binary activation masks, have revealed a number of interesting characteristics that should be common to most deep convolutional networks. The crucial role of the activations, and no so much their intensities, has been evidenced in our experiments also with randomly chosen parameters. Also the filters that become activated seem to compute statistics on image patches rather memorize specific patterns. The idea that the network at the various

layers learns specific image patterns seems to be confuted, because the activations are more gradient-like detectors. Much still needs to learned about the working of these complex architectures. Our current effort is devoted to gain a more comprehensive quantitative account for analysis and design.

## References

1. Poggio, T., Mhaskar, H., Rosasco, L., Miranda, B., Liao, Q.: Why and when can deep-but not shallow-networks avoid the curse of dimensionality: a review. *Int. J. Autom. Comput.* **14**, 503–519 (2017). Oct
2. Yosinski, J., Clune, J., Nguyen, A., Fuchs, T., Lipson, H.: Understanding neural networks through deep visualization. ArXiv e-prints (June 2015)
3. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *Computer Vision - ECCV 2014*, pp. 818–833. Springer International Publishing, Cham (2014)
4. Zeiler, M.D., Taylor, G.W., Fergus, R.: Adaptive deconvolutional networks for mid and high level feature learning. In: 2011 International Conference on Computer Vision, pp. 2018–2025 (Nov 2011)
5. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
6. van den Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., Kavukcuoglu, K.: WaveNet: A Generative Model for Raw Audio. ArXiv e-prints (Sept 2016)
7. Montúfar, G., Pascanu, R., Cho, K., Bengio, Y.: On the number of linear regions of deep neural networks. In: Proceedings of the 27th International Conference on Neural Information Processing Systems, NIPS’14, vol. 2, pp. 2924–2932. MIT Press, Cambridge, MA, USA (2014)
8. Montufar, G.F., Pascanu, R., Cho, K., Bengio, Y.: On the number of linear regions of deep neural networks. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (eds.) *Advances in Neural Information Processing Systems*, vol. 27, pp. 2924–2932, Curran Associates, Inc. (2014)
9. Cybenko, G.: Approximation by superpositions of a sigmoidal function. *Math. Control Signals Syst.* **2**, 303–314 (1989)
10. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. ArXiv e-prints (Sept 2014)
11. Lowe, D.G.: Object recognition from local scale-invariant features. *Proceedings of the Seventh IEEE International Conference on Computer Vision* **2**, 1150–1157 (1999)

# Chapter 20

## Neural Epistemology in Dynamical System Learning



Pietro Barbiero, Giansalvo Cirrincione, Maurizio Cirrincione,  
Elio Piccolo and Francesco Vaccarino

**Abstract** In the last few years, neural networks are effectively applied in different fields. However, the application of empirical-like algorithms as feed-forward neural networks is not always justified from an epistemological point of view [1]. In this work, the assumptions for the appropriate application of machine learning empirical-like algorithms to dynamical system learning are investigated from a theoretical perspective. A very simple example shows how the suggested analyses are crucial in corroborating or discrediting machine learning outcomes.

**Keywords** Epistemology · Time series learning · Feed-forward neural networks · Machine learning · Forecasting · Poincaré recurrence theorem · Bifurcation theory

### 20.1 The Need for an Epistemology

In the last few years, machine learning and neural networks are effectively applied in different fields. Experiments often show great results both on structured and unstructured data. However, when they obtain poor results, it is not straightforward from a human being point of view to understand why. In some respects, this need can seem absurd. We typically employ machine learning to solve tasks which are not solvable by human beings. Nevertheless, we require to understand how they make decisions

---

P. Barbiero (✉) · F. Vaccarino

Department of Mathematical Sciences, Politecnico di Torino, 10126 Torino, Italy  
e-mail: [pietro.barbiero@studenti.polito.it](mailto:pietro.barbiero@studenti.polito.it)

G. Cirrincione

University of Picardie Jules Verne Lab. LTI, Amiens, France  
e-mail: [exin@u-picardie.fr](mailto:exin@u-picardie.fr)

M. Cirrincione

University of South Pacific, Suva, Fiji

E. Piccolo

Department of Control and Computer Engineering,  
Politecnico di Torino, 10126 Torino, Italy  
e-mail: [elio.piccolo@polito.it](mailto:elio.piccolo@polito.it)

and, above all, we want to understand the circumstances in which they struggle to get good results. Generally, when we observe such situations, the spontaneous reaction is to charge the architecture, the hyperparameter tuning, or the optimization algorithm. This is the typical “scapegoat” answer where the impeachment is charged to the algorithm. *However it does not question the use of the algorithm itself.* Can we really be so sure? In this work, we try to collect some justifications that can support the use of machine learning algorithms to dynamical system learning. The kind of problem we are stating is epistemological. Epistemology is a branch of philosophy which investigates the origin, nature, methods, and limits of knowledge. Given the strict correlation between science and knowledge, epistemology is probably the closest branch of philosophy to science. Apart from the diatribes between justificationism and critical rationalism, machine learning needs epistemology for two main reasons. On the one hand, researchers will be aware in advance about the limits of machine learning algorithms before they dive into a new problem. On the other hand, epistemology can have a social impact in increasing the confidence of external users to machine learning.

## 20.2 Mathematical Models of Dynamical Systems

From Newton and Leibniz on, the standard way of mathematicians to represent real-world phenomena is based on calculus [2]. More specifically, time-variant phenomena are usually described either by differential equations or by recurrence relations. Both approaches represent physical quantities  $q$  and their rate of change  $q_r$  by means of equations:

$$q = f(q_r) \quad (20.1)$$

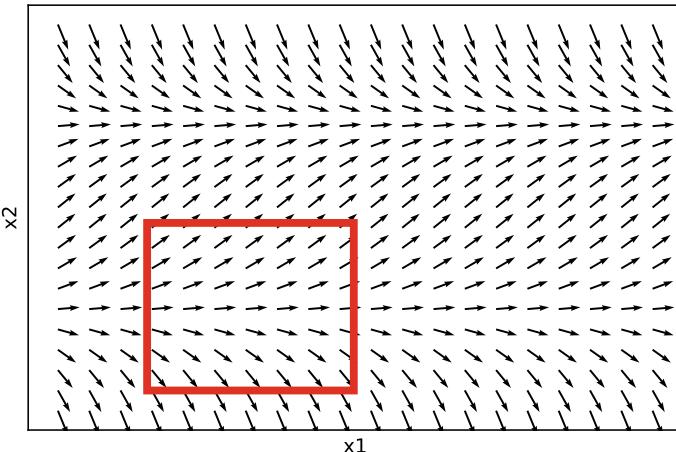
This kind of representations are very straightforward to interpret by humans. Thus, they have emerged as references both in purely theoretical subjects and in making hypotheses for new experiments. However, their compact representation hides the global understanding of the phenomenon. One of the most used approaches consists in solving the system of equations and drawing the corresponding phase space [3]. The phase space is a space in which all possible states of a system are represented, with each possible state corresponding to one unique point in the phase space. The solution of the dynamical equations can be represented as a trajectory in the phase space. The resulting plot gives qualitative but global information about the system behavior. In the following, we will focus on two trajectory characteristics: boundedness and periodicity. Periodic trajectories are curves that repeat their values in regular intervals. A bounded trajectory, instead, is a curve which can be represented in a finite volume. Dynamical systems having bounded or periodic solutions are fully observable (at least theoretically); i.e., we can plan an experiment in order to observe each possible state of the system or, for periodic ones, we can collect enough data to infer the system behavior outside the experiment window.

### 20.3 The Poincaré Recurrence Theorem

The Poincaré recurrence theorem points out an interesting property of dynamical systems. Indeed, it gives the assumptions under which a dynamical system returns close to previous states in a finite time interval.

**Theorem 1** *If a dynamical system has a bounded domain in the phase space, then for each open neighborhood  $I_P$  of a point  $P$  in the phase space there exists a point  $P' \in I_P$  that the system will encounter in a finite time [2, 4] (Fig. 20.1).*

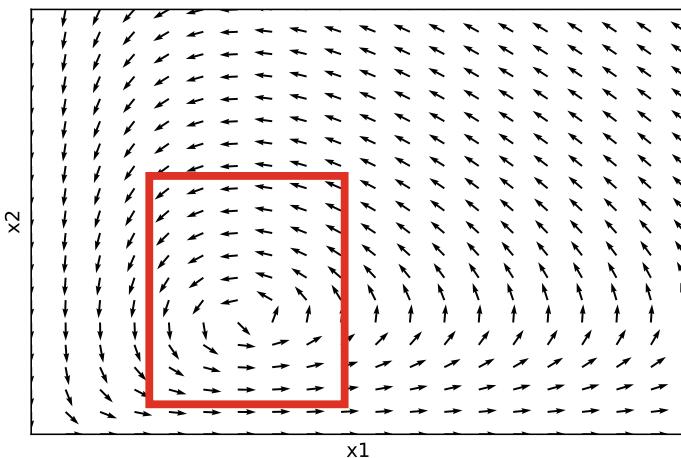
The impact of this theorem on machine learning epistemology is enormous. Ensuring that *in a finite time* the system will return near a previous state, the theorem guarantees the existence of recurrence which is the building block of statistics and machine learning which literally learn associations (see, e.g., the i.i.d. assumptions in [5]). This recurrence property of bounded phase spaces can be extended to periodic ones (even if unbounded) in the sense that recurrence is implicit in repetition. By analyzing the solutions of the dynamical equations, we can have an insight about what could be the limits of machine learning algorithms when applied to the problem.



**Fig. 20.1** Phase space of the system of equations:  $x'_1 = 1$  and  $x'_2 = x_2(1 - x_2)$ . The arrows represent possible trajectories of the dynamical system. The red box represents a possible bound to trajectories. Observe that it does not exist a box which can bound these trajectories. The system behavior cannot be inferred by analyzing the states inside the experiment window

## 20.4 What Can Go Wrong

Despite the fact that most of real-world phenomena have a bounded phase space, there are some aspects that can get machine learning into trouble. First of all, experiments may (and often) have practical limits in observing phenomena. Measuring instruments have limited ranges, and/or recurrences may occur in undetectable time intervals. Thus, the retrieved data may lack in fundamental associations needed by algorithms. Another kind of problem may arise in real-time applications when the algorithms process online data. Suppose you have modeled a phenomenon, you have carried out an experiment and used the output data to train an algorithm. Then, you want to apply the trained algorithm to process a stream of new data. In this case, if the original model is not perfect or does not take into account some important variables, then the data used to train the algorithm may have a different distribution from real-time ones (violating the i.i.d. assumptions [5]). In such cases, machine learning algorithms try to extrapolate inferring outside the training domain. Unfortunately, the results are usually decidedly worse. Finally, dynamical systems may be affected by bifurcations [1, 6]. Bifurcations occur when a small smooth change made to the state of the system causes a sudden change in its behavior. In such cases, most machine learning algorithms struggle to get meaningful results. Thus, researchers should analyze both the model and the experiment carefully before applying machine learning in such contexts (Fig. 20.2).



**Fig. 20.2** Phase space of the system of equations:  $x'_1 = ax_1 - bx_1x_2$  and  $x'_2 = dx_1x_2 - cx_2$ . This system is the Lotka–Volterra model also known as the predator–prey model. The arrows represent possible trajectories of the dynamical system. Observe that the red square is a bound for some trajectories

## 20.5 How Machine Learning Sees Dynamical Systems

Having modeled a dynamical phenomenon and designed the corresponding experiment, the outcome is a sequence of data usually organized as a time series [7]. A time series is an ordered sequence of data points, where the order refers to time. In practice, a certain quantity (e.g., the temperature) is sampled at successive (often) equally spaced time instants. More formally, a unidimensional time series can be described as:

$$a_n = (x_0, x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_T) \quad x_i \in \mathbb{R} \quad \forall i \in [0, T] \quad (20.2)$$

where the data point  $x_i$  is a real number and it refers to the  $i$ th time instant. The most common use of such data is forecasting analysis. Forecasting refers to the use of a mathematical model that tries to predict future values based on observed ones [8]. More formally, having defined a time window  $w \in \mathbb{N}$ , a time series forecasting model is a function:

$$f : \mathbb{R}^w \rightarrow \mathbb{R} \quad (20.3)$$

that uses the observed points  $(x_i, \dots, x_{i+w-1})$  in order to predict the value of  $x_{i+w}$ . From a machine learning point of view, each observed point corresponds to a feature and the future value is the corresponding target. Following this perspective, it is possible to reshape the time series in order to build a more suitable database for machine learning algorithms:

$$(x_0, \dots, x_{w-1}) \rightarrow y_0 = x_w \quad (20.4)$$

$$(x_1, \dots, x_w) \rightarrow y_1 = x_{w+1} \quad (20.5)$$

$$(x_2, \dots, x_{w+1}) \rightarrow y_2 = x_{w+2} \quad (20.6)$$

$$\vdots \quad (20.7)$$

In practice, a machine learning algorithm should learn the associations from a training set of the above database.

## 20.6 A Song Experiment: The Frère Jacques Song

“Frère Jacques” is a famous French child song [9]. Figure 20.3 shows the music sheet of the song.

For the purpose of the experiments, assume that the music sheet is repeated from the beginning several times, thus simulating a bounded and periodic dynamical system. For the sake of visualization, initially consider a time window  $w = 2$ . Indeed, with such a choice it is possible to visualize predictions in a 2-dimensional plane. In order to set up the database, the following one-hot encoding is chosen:



**Fig. 20.3** Frère Jacques music sheet

$$C = 1000000$$

$$D = 0100000$$

$$E = 0010000$$

$$F = 0001000$$

$$G = 0000100$$

$$A = 0000010$$

$$B = 0000001$$

Given the encoding choice, it is possible to build a database iteratively in the following way [7, 8]:

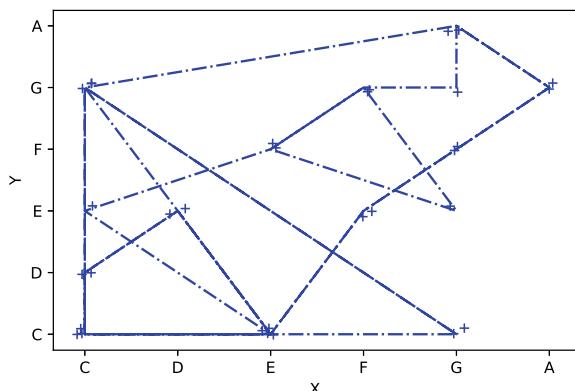
$$x_1 = (C, D) \rightarrow y_1 = (E) \quad (20.8)$$

$$x_2 = (D, E) \rightarrow y_2 = (C) \quad (20.9)$$

$$\vdots \quad (20.10)$$

where the note symbols are replaced with their corresponding binary codes. The phase state of the described problem is limited, and its trajectories are bounded in a finite volume as shown in Fig. 20.4.

**Fig. 20.4** Phase space of the Frère Jacques song (window size  $w = 1$ ). Observe that the trajectories along which the dynamical system evolves are bounded in a finite volume

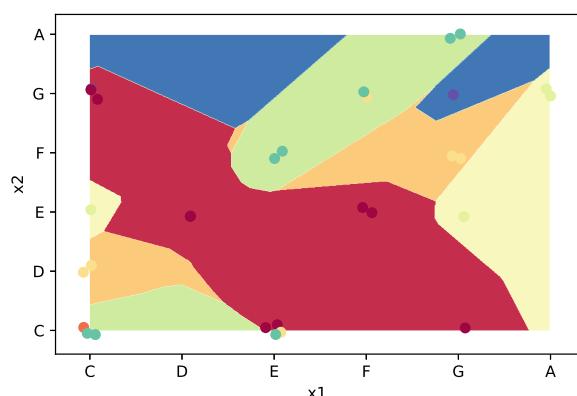


Indeed, the system is periodic and exactly returns in a previous state after a finite (and short) time interval. Since the learning assumptions are fulfilled, any machine learning algorithm can be correctly applied to the problem. In particular, we are interested in how multilayer perceptron (MLP) tackles time series forecasting [10–12]. For the purpose of this experiment, the MLP is equipped with the cross-entropy error function and a SoftMax output layer in order to perform multiclass recognition. The optimization algorithm used is the adaptive momentum estimation optimizer (Adam). It is an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments [13]. The objective is forecasting the next note given the previous ones. So, each unit in the output layer corresponds to a note and the network prediction corresponds to the output unit with the highest activation. The MLP architecture is composed of the input layer with 2 units (equal to the time window size), a hidden layer, and an output layer with 7 units (one for each note). The experiments are carried out with 30 units in the hidden layer.

Figure 20.5 shows the input space  $x_1-x_2$  having chosen a window of size  $w = 2$ . In the figure, there are 7 regions with different colors. They are delimited by the estimated decision boundaries learned by the network. The MLP tries to delimit the input space with straight lines (or hyperplanes) in order to label each region of the space. Each point of a region is identified by the label of the region it belongs. This label is compared with the true target in order to evaluate the mismatch. If the predicted labels correspond to targets, then the cost is low; otherwise, it increases for each mislabeled sample. The experiment is repeated 100 times for different window sizes. The training accuracy increases according to the window size as shown in table 20.1.

These results can be explained by analyzing the phase space of the song. For example, when  $w = 6$ , there exists 2 sequences  $X = [G, A, G, F, E, C]$  (“Sonnez les matines!”) in the training set. However, the first one has  $Y = G$  as target (the next note after the sequence), while the second one has  $Y = C$ . This is the only sequence in which the MLP fails with this window size. Indeed, the two sequences are the

**Fig. 20.5** Decision boundaries learned by a MLP for the Frère Jacques song with a window size  $w = 2$



**Table 20.1** Training accuracy of the Frère Jacques song experiment for different window sizes

Window size	Training accuracy
$w = 1$	$39.4 \pm 1.9\%$
$w = 2$	$82.9 \pm 1.1\%$
$w = 3$	$86.2 \pm 0.0\%$
$w = 4$	$92.9 \pm 0.0\%$
$w = 5$	$96.3 \pm 0.0\%$
$w = 6$	$96.2 \pm 0.0\%$
$w = 7$	$100.0 \pm 0.0\%$

same, but with different targets. Hence, MLP has no information to disambiguate the forecast. This can be regarded as a discrete bifurcation. The experiment window is not enough to allow the disambiguation and the MLP fails the prediction even if the phase space is bounded. Carrying out the same analysis for all the window sizes, it turns out that the training performances listed above correspond to the performances of a Bayes classifier; i.e., they are the optimal ones.

## 20.7 Conclusion

In this work, we pointed out the need for an epistemology for machine learning. To begin with, we focused on dynamical systems. We described the typical scientific process used to approach the problem from the design of mathematical models to the application of machine learning on the experiment outcomes. We observed how the Poincaré recurrence theorem could be a reliable building block for the foundation of such epistemology. Finally, we underlined how further analyses should go with the application of machine learning to dynamical systems. Specifically, they are related to: the boundedness of the trajectories with regard to the experiment window, the differences between the expected data distribution and the real one, and the presence of bifurcations. We illustrate with a simple experiment how the underlined problems may emerge. The example shows how the suggested analyses are crucial in corroborating or discrediting machine learning outcomes.

## Appendix: Source Code

The source code of the Frère Jacques song experiment is downloadable at <http://www.pietrobarbiero.eu/>.

## References

1. Ellingsen, B.K., Grimen, H., et al.: The Epistemology of Learning in Artificial Neural Networks. Citeseer (1994)
2. Barreira, L.: Poincaré recurrence: old and new. In: XIVth International Congress on Mathematical Physics (2006). <https://doi.org/10.1142/97898127040160039>
3. Hirsch, M.W., Smale, S., Devaney, R.L.: Differential Equations, Dynamical Systems, and an Introduction to Chaos. Academic press (2012)
4. Poincaré, H.: Sur le problème des trois corps et leséquations de la dynamique. In: Acta Mathematica (1890). <https://projecteuclid.org/euclid.acta/1485881725>
5. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016). <http://www.deeplearningbook.org>
6. Poincaré, H.: L'quilibre d'une masse uide animée d'un mouvement de rotation. In: Acta Mathematica (1885). [http://smf4.emath.fr/Publications/Gazette/2005/104/smf\\_gazette\\_104\\_71-76.pdf](http://smf4.emath.fr/Publications/Gazette/2005/104/smf_gazette_104_71-76.pdf)
7. Dietterich, T.G.: Machine learning for sequential data: a review. In: Structural, Syntactic, and Statistical Pattern Recognition, pp. 15–30. Springer, Heidelberg (2002). ISBN: 978-3-540-70659-5. [https://doi.org/10.1007/3-540-70659-3\\_2](https://doi.org/10.1007/3-540-70659-3_2)
8. Bontempi, G., Taieb, S.B., Le Borgne, Y.-A.: Machine learning strategies for time series forecasting. In: Business Intelligence: Second European Summer School, eBISS 2012, Brussels, Belgium, July 15–21, 2012, Tutorial Lectures, pp. 62–77. Springer, Heidelberg (2013). [https://doi.org/10.1007/978-3-642-36318-4\\_3](https://doi.org/10.1007/978-3-642-36318-4_3).
9. Frère Jacques Song. <https://en.wikipedia.org/wiki/Fr>
10. Bishop, C.: Neural Networks for Pattern Recognition. Clarendon Press (1995). ISBN 978-0198538646
11. Tensorow. <https://www.tensorflow.org/>
12. Funahashi, K., Nakamura, Y.: Neural networks, approximation theory, and dynamical systems. In: Structure and Bifurcations of Dynamical Systems-Proceedings Of The Rims Conference (1992)
13. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: International Conference for Learning Representations (2014)

# Chapter 21

## Assessing Discriminating Capability of Geometrical Descriptors for 3D Face Recognition by Using the GH-EXIN Neural Network



**Gabriele Ciravegna, Giansalvo Cirrincione, Federica Marcolin, Pietro Barbiero, Nicole Dagnes and Elio Piccolo**

**Abstract** In pattern recognition, neural networks can be used not only for the classification task, but also for feature selection and other intermediate steps. This paper addresses the 3D face recognition problem in order to select the most meaningful geometric descriptors. At this aim, the classification results are directly integrated in a biclustering process in order to select the best leaves of a neural hierarchical tree. This tree is created by a novel neural network GH-EXIN. This approach results in a new criterion for the feature selection. This technique is applied to a database of face expressions where both traditional and novel geometric descriptors are used. The results state the importance of the curvedness novel descriptors and only of a few Euclidean distances.

### 21.1 Introduction

3D face recognition has been deeply investigated in the last decades due to the large number of applications in both security and safety domains, even in real-time scenarios. The third dimension improves accuracy and avoids problems like lighting

---

G. Ciravegna (✉)  
DINFO, Università degli studi di Firenze, Firenze, Italy  
e-mail: [gabriele.ciravegna@unifi.it](mailto:gabriele.ciravegna@unifi.it)

G. Cirrincione  
Université de Picardie Jules Verne, Lab. LTI, Amiens, France

G. Cirrincione  
University of South Pacific, Suva, Fiji

F. Marcolin · N. Dagnes  
DIGEP, Politecnico di Torino, Torino, Italy

P. Barbiero  
DISMA, Politecnico di Torino, Torino, Italy

E. Piccolo  
DAUIN, Politecnico di Torino, Torino, Italy

and make-up variations. In addition, it allows the adoption of geometrical features to study and describe the facial surface.

This research makes use of seven novel geometrical descriptors which rely on shape and curvedness index and the coefficients of the first fundamental forms. These descriptors have been presented in [1]. The face model is a “mean face” evaluated with a 100 neutral training faces from the Bosphorus database [2]. Formulas and facial mappings of the novel descriptors are shown in Table 21.1.

In addition to these novel features, other descriptors, presented in [3], have been used including Euclidean and geodesic distances between landmarks, the nose volume, and the shape index [4]. Overall, a set of 11 feature types was generated. All geometrical descriptors, i.e., those reported in Table 21.1 and the shape index, are adopted in this work in the form of histograms.

## 21.2 Database Creation and Goals

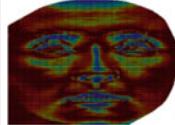
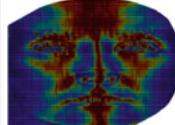
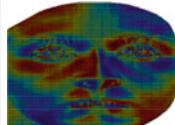
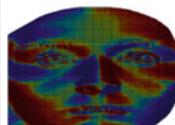
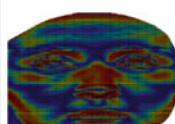
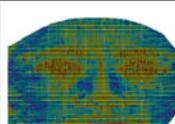
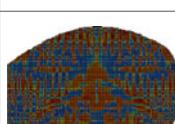
The global set of features for each face is given by 171 features: 12-bins histograms of six novel geometrical descriptor (72 features); 7-bins histograms of the shape index and the personal shape index (14); Euclidean distances between landmarks (62); geodesic distances (22); and nose volume (1). The faces in this dataset belong to 62 subjects of the Bosphorus database [2], chosen in such a way to create a different dataset, composed of seven facial expressions each (Ekman’s basic emotions [5, 6]), meaning 434 faces overall.

The methodology proposed in this work investigates the capabilities of features to discriminate between different individuals, so that the inter-person variability (between subjects) is maximized and intra-person variability (between different expressions of the same subjects) is minimized. Selecting discriminating features for subject identification and recognition is desirable and has been recently addressed in the 3D context [7]. Specific methodologies have been developed to improve stability [8], interpretability, and predictability [9], but the advances in the 3D domain are still underway.

## 21.3 Data Analysis

This study proposes a novel feature selection technique based on an original self-organizing neural network, the Growing Hierarchical EXIN algorithm (GH-EXIN, [10]), which builds a hierarchical tree in a divisive way, by adapting its architecture to data. This clustering technique is integrated in a biclustering framework in which face and descriptor (feature) clusterings are alternated and controlled by the quality of the bicluster. In our case, biclustering allows to get meaningful insights into what are the

**Table 21.1** Formulas of new descriptors and respective point-by-point mappings on a Bosphorus facial depth map.  $Z_x$  and  $Z_y$  are derivatives with respect of  $x$  and  $y$ , respectively, of the facial surface  $Z$ . Complete formulas of  $E$ ,  $F$ ,  $G$  are given in [1];  $C$  is the curvedness index theorized by Koenderink and vanDoorn [4]

Descriptor formula	Facial map
$E_{\text{den}2} = \frac{E}{1+Z_x^2+Z_y^2}$	
$G_{\text{den}2} = \frac{G}{1+Z_x^2+Z_y^2}$	
$S_{\text{fond}1} = -\frac{2}{\pi} \arctan \frac{E+F+G}{E+G-F}$	
$\arctan F$	
$\arctan G$	
$\log C$	
$S_{\text{pers}} = S[Z + (Z - Z_F M)]$	

most interesting features by analyzing the subspaces composed of the clustered faces. Clustering techniques have been already applied with success in the field of 3D face recognition [11] with a different aim: facial feature identification and localization.

### 21.3.1 Biclustering

Biclustering, also known as two-way clustering or manifold (subspace) clustering, is the key of this work. This technique was introduced in the 1960s, but has been properly defined only by Cheng and Church in 2000 [12]. Basically, clustering can be applied to either the rows or the columns of the data matrix, separately. Biclustering performs clustering in both dimensions, at the same time, as shown in Fig. 21.1.

It has several advantages over clustering since it groups items based on a subset of the features so that it does not only perform grouping but also discovers the context (subspace) in which the groups are found. Furthermore, the projection of the biclusters into the features or the samples space allows to analyze the results as grouping of samples or features, respectively. Biclustering has been deeply used as a technique to study the coregulation of genes in DNA microarray analysis [10, 12]. In these works, biclusters were projected into the sample space in order to discover in which conditions—i.e., for which individuals—different genes coregulate. In our case, instead, biclustering allows to understand what are the descriptors that better group faces of the same person. This is achieved by projecting the biclusters into the feature (descriptor) space. Further details will be given in Sect. 21.4.

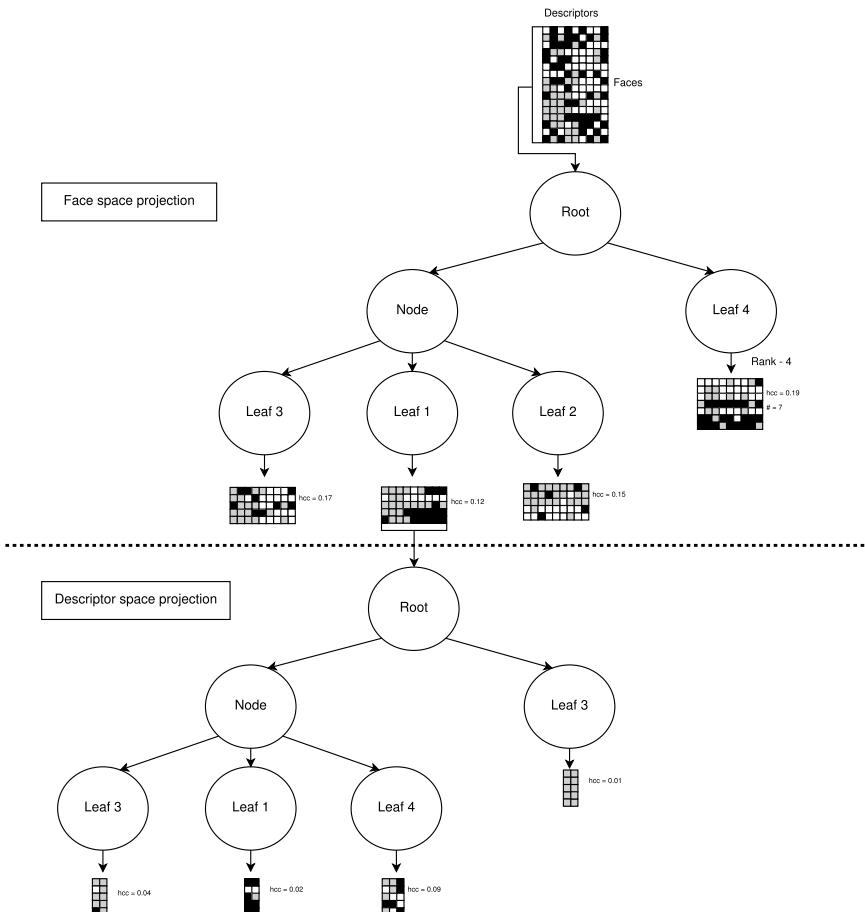
Biclustering generally searches for biclusters with constant values, with constant values on rows or columns and with coherent values, respectively. For the database at hand, our approach searches for constant column biclusters, that is submatrices in which the elements along the same column are similar, but may differ with regard to the other columns for a constant. The  $H_{cc}$  index, introduced by Cheng and Church [12], is used to control the quality of the bicluster. It takes into account the noise in data and is expressed as:

$$H_{cc} = \frac{\sum_i^{N_r} \sum_j^{N_c} r_{ij}^2}{N_r N_c} \quad (21.1)$$

where  $N_c$  represents the total number of columns of the matrix,  $N_r$  represents the total number of rows and  $r_{i,j}$  is the residue, which is calculated as:

$$r_{ij} = a_{ij} - \frac{\sum_k^C a_{ik}}{C} - \frac{\sum_h^R a_{hj}}{R} + \frac{\sum_i^R \frac{\sum_j^C a_{ik}}{C}}{R} \quad (21.2)$$

The terms  $a_{i,j}$  are the elements of the matrix (rows and columns represent faces and descriptors).  $C$  and  $R$  are the number of columns and of rows of the bicluster at hand, respectively. The second term is the average value of the  $i$ th row, the third term is the average value of the  $j$ th column, while the last one is the average value

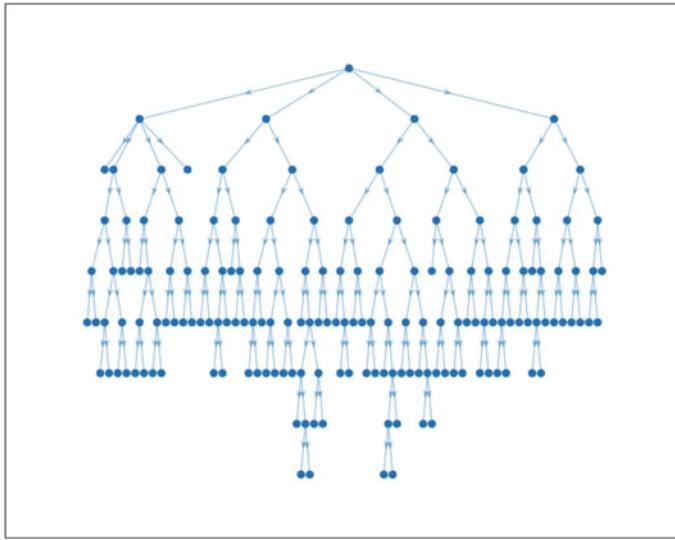


**Fig. 21.1** Neural bicluster

of the whole bicluster. This index decreases as the values in the bicluster tend to be constant, differing for a constant on the rows or a constant on the columns. It goes to zero for the trivial  $1 \times 1$  bicluster. This drawback implies additional controls on the minimum bicluster cardinality.

### 21.3.2 The GH-EXIN Neural Network

Biclustering is here based on the GH-EXIN neural network. It is a hierarchical variant of the quantization layer of the GCCA neural network [13]. This structure is useful as it easily allows to select the desired cluster resolution level, by simply stopping the



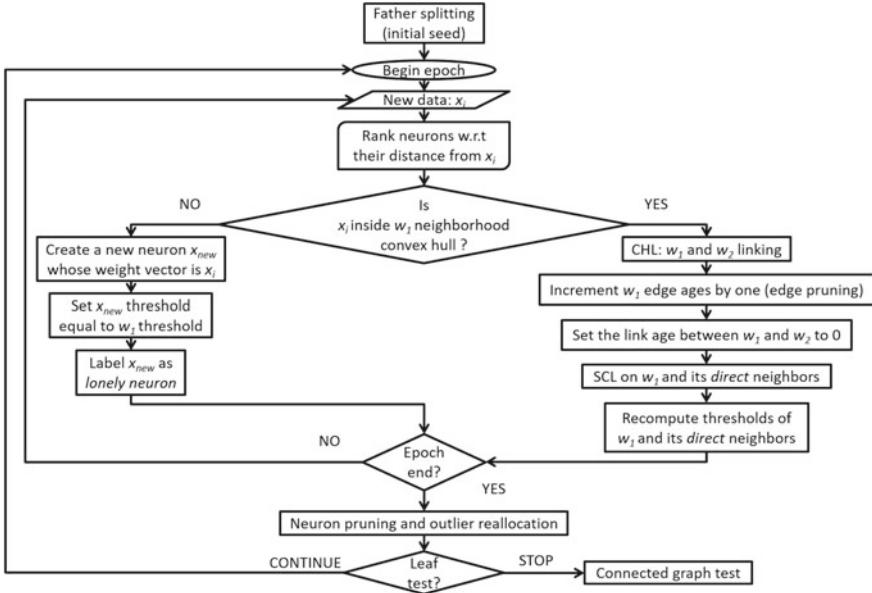
**Fig. 21.2** GH-EXIN tree for the first step of the biclustering (face clustering)

network when a certain height in the hierarchy has been reached. This is fundamental in biclustering because it allows to alternate the clustering in the two spaces several times until the best biclusters are found.

GH-EXIN builds a hierarchical (divisive) tree whose vertices correspond to its neurons as it is shown in Fig. 21.2. Each neuron is equipped with a weight vector whose dimensionality is the same of the input space. For each father neuron, a neural network is trained on its corresponding Voronoi set (set of data represented by the father neuron). The sons are the neurons of the associated neural network and determine a subdivision of the father Voronoi set. For each leaf, the procedure is repeated. The initial structure of the neural network is a seed, i.e., a pair of neurons, which are linked by an edge, whose age is set to zero.

The GH-EXIN architecture is data-driven, in the sense that the number of neurons is determined by the training set by means of node creation or pruning. For each epoch (presentation in a random way of the whole training set to the network), the basic iteration starts at the presentation of a new data, say  $x_i$ . All neurons are ranked according to the Euclidean distances between  $x_i$  and their weights. The neuron with the shortest distance is the winner  $w_1$ . If its distance is higher than the scalar threshold of the neuron (novelty test), a new neuron is created with weight vector given by  $x_i$  (left branch of Fig. 21.3 diagram). The initial weight vectors and neuron thresholds are given by heuristics.

Otherwise, there is a weight adaptation and the creation of an edge (right branch of Fig. 21.3 diagram). The weight computation (training) is based on the Soft Competitive Learning (SCL) [14] paradigm, which requires a winner-take-most strategy: At each iteration, both the winner and its neighbors change their weights but in different



**Fig. 21.3** GH-EXIN flowchart

ways:  $w_1$  and its direct topological neighbors are moved toward  $x_i$  by fractions  $\alpha_1$  and  $\alpha_n$  (learning rates), respectively, of the vector connecting the weight vectors to the datum.

This law requires the determination of a topology (neighbors) which is achieved by the Competitive Hebbian Learning (CHL) rule [14], which is used for creating the neuron connections: Each time a neuron wins, an edge is created, linking it to the second nearest neuron, if it does not exist yet. If there was an edge, its age is set to zero and the same age procedure as in [13] is used as follows. The age of all other links emanating from the winner is incremented by one; if a linkage is greater than the *agemax* scalar parameter, it is eliminated (pruning).

The thresholds of the winner and second winner are recomputed as the distance to their farthest neighbor.

At the end of each epoch, if a neuron remains unconnected (no neighbors), it is pruned, but the associated data are analyzed by a new ranking of all the neurons of the network (i.e., also the neurons of the neural networks of the other leaves of the hierarchical tree). If it is outside the threshold of the new winner, it is labeled as outlier and pruned. If, instead, it is inside, it is assigned to the winner Voronoi set.

Each leaf neural network is controlled by  $H_{cc}$  because GH-EXIN is searching for biclusters (it is estimated by using the data of each Voronoi set). In particular, the training epochs are stopped when the estimated value of this parameter falls below a percentage of the value for the father leaf.

This technique builds a vertical growth of the tree. The horizontal growth is generated by the neurons of each network. However, a simultaneous vertical and horizontal growth is possible. At the end of a training, the graphs created by the neuron edges are checked. If connected subgraphs are detected, each subgraph is considered as a father, by estimating the centroid of the cluster (vertical growth) and the associated neurons as the corresponding sons (horizontal growth).

The whole neural approach requires two groups of user dependent parameters:

1. The GH-EXIN parameters, i.e., the two learning rates and the scalar  $agemax$  for edge pruning; the two rates are constant values for SCL. However, they can be made decreasing in time and automatically scaled by using the Voronoi cardinality (conscience). The last one has to be lowered if more edges (and neurons) have to be pruned. In a sense, it controls, in an indirect way, the leaf cardinality.
2. The biclustering quality indices, i.e., the percentage of  $H_{cc}$ , its maximum value and the minimum cardinality of leaves. They control the search and require a deep analysis (out of the scope of the paper).

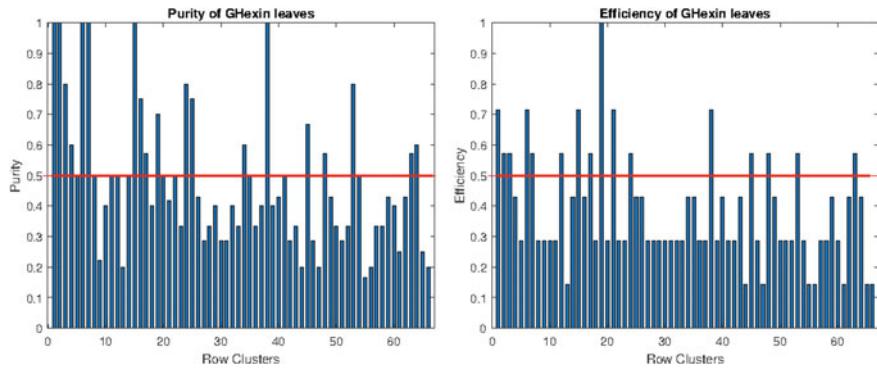
## 21.4 Analysis of the Database

In this work, biclustering is used for estimating the descriptor subspaces in which clusters of faces (mostly same person) are detected. These subspaces are meaningful for 3D detection. It can be argued that their intersection is the core for a correct classification. Here, this intersection is related to the number of times (frequency) a feature is found in each bicluster. At this aim, a two-step approach is proposed.

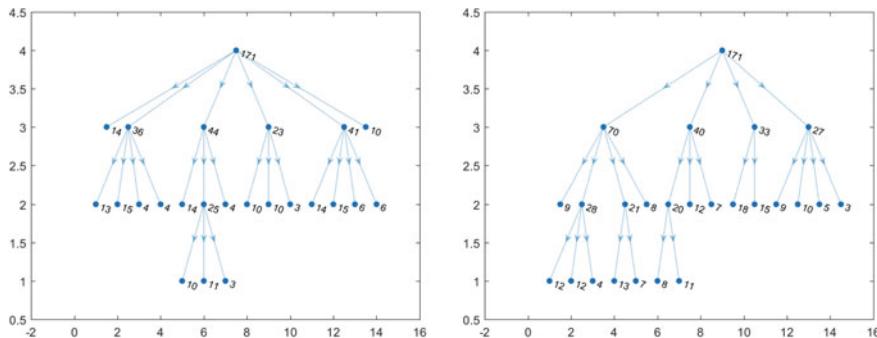
Firstly, a hierarchical clustering of the samples (434 faces) is performed by using GH-EXIN (Fig. 21.2). Then, some leaves are selected according to the criteria of efficiency and purity. They are useful indexes in clustering analysis and give an estimate about the classification accuracy of the algorithm. They both compute the number of elements in a cluster belonging to the same class. While purity compares it with the cardinality of the cluster, efficiency compares it with the cardinality of the class in the whole dataset. A common issue using purity and efficiency separately is that they may select clusters composed of too few or too many elements, respectively. The use of both indexes at the same time avoids a further check on the cardinality of the clusters. This approach allows the selection of those clusters whose Voronoi set contains faces belonging mostly to the same person. As shown in Fig. 21.4, 12 leaves are retained: Both efficiency and purity indexes are above 0.5 (red line).

The Voronoi sets of the neurons of the best leaves have been found by using all the descriptors of the faces. In order to find the most meaningful features, the role of samples and descriptors is reversed and, for each leaf, a hierarchical tree is created by GH-EXIN. Figure 21.5 shows two of these neural structures.

The goal of the second step is the clustering of descriptors for each selected leaf (mostly one person, proportional to its purity). The resulting biclusters identify the



**Fig. 21.4** Purity and efficiency after the first GH-EXIN



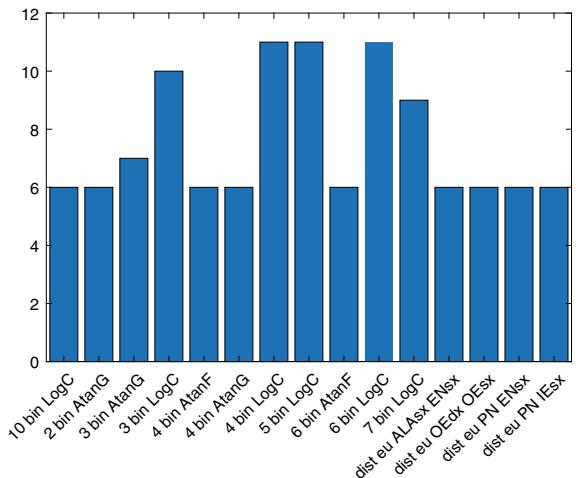
**Fig. 21.5** GH-EXIN tree of the second step of the biclustering (descriptor clustering)

best subspace for the faces of the leaf. Indeed, they are identified by lower values of  $H_{cc}$  index (an upper threshold of 0.1 is adopted).

The descriptors shown in Fig. 21.6 are those that are found at least six times in the 12 selected leaves (the best ones of the first clustering). These descriptors have, therefore, the highest discriminative power for their capability in assigning similar values to faces of the same person, which implies to distinguish persons.

Among all descriptors, this study revealed the discriminative capability of LogC. Its bins have been selected until 11 times over 12. As shown in Table 21.1, the corresponding figure clearly reveals the intrinsic capability of the descriptor to display the most important trait of a person. As previously said in the introduction, this descriptor has been recently introduced in [1] and further research in the future will be done in order to analyze its importance. The bins of AtanF and AtanG are also selected many times. AtanF is a descriptor capable to highlight all the critical points of the face, like the nasion, and to take into account face asymmetries. AtanG, instead, is used to show the curvedness of the face. In Table 21.1, the blue, yellow, and red colors represent the negative curvedness, the flat surface and the positive curvedness, respectively. The remaining selected geometrical descriptors are only a few Euclidean

**Fig. 21.6** Histograms of most discriminative descriptors



distances. Nevertheless, they are among the most important in literature, like the distance between the pronasal landmark and the inner eyebrow landmark.

## 21.5 Conclusion

This work faces the problem of the discriminative power of the 3D face descriptors in an original way. Indeed, instead of taking into account the classification results, a neural network is created for solving a biclustering problem, by exploiting the purity and efficiency results for the choice of the meaningful leaves. In this sense, it can be stated that the power of the descriptors is integrated in the neural architecture. The consequent approach of selecting the most frequent features in the biclusters results in the assessment of the importance of the novel curvedness descriptors and only of a very few Euclidean distances, in accordance with the analysis in [3], where the low intrinsic dimensionality of the corresponding manifold is determined.

Feature work will deal with the analysis of the manifold of the novel descriptor features and their impact in a new 3D face neural classifier.

## References

1. Marcolin, F., Vezzetti, E.: Novel descriptors for geometrical 3D face analysis. *Multimedia Tools Appl.* **76**(12), 13805–13834 (2017)
2. Savran, A., Alyüz, N., Dibeklioğlu, H., Çeliktutan, O., Gökberk, B., Sankur, B., Akarun, L.: Bosphorus database for 3D face analysis. In: European Workshop on Biometrics and Identity Management, pp. 47–56 (May 2008)
3. Cirrincione, G., Marcolin, F., Spada, S., Vezzetti, E.: Intelligent quality assessment of geometrical features for 3D face recognition. In : 27th Italian Workshop on Neural Networks (WIRN). Vietri sul mare, Salerno (2017)

4. Koenderink, J., van Doorn, A.: Surface shape and curvature scales. *Image Vision Comput.* **10**(8), 557–564 (1992)
5. Ekman, P.: Universal facial expressions of emotions. *Calif. Mental Health Res. Dig.* **8**(4), 151–158 (1970)
6. Ekman, P., Keltner, D.: *Facial Expressions of Emotions*. Lawrence Erlbaum Associates Publisher, Mahwah, New Jersey (1997)
7. Gunlu, G., Bilge, H.: Feature extraction and discriminating feature selection for 3D face recognition. In: 24th International Symposium on Computer and Information Sciences, pp. 44–49 (September 2009)
8. Bevilacqua, V., et al.: 3D head pose normalization with face geometry analysis, genetic algorithms and PCA. *J. Circuits Syst. Comput.* **18**, 1425–1439 (2009). <https://doi.org/10.1142/S0218126609005769>
9. Guo, Y.: SIP-FS: a novel feature selection for data representation. *EURASIP J. Image Video Process.* **1**, 14 (2018)
10. Barbiero, P., Bertotti, A., Ciravegna, G., Cirrincione, G., Cirrincione, M., Piccolo, E.: Neural biclustering in gene expression analysis. In: International Conference on Computational Science and Computational Intelligence 2017, Las Vegas, USA
11. Bevilacqua, Vitoantonio, Mastronardi, Giuseppe, Santarcangelo, Vito, Scaramuzzi, Rocco: 3D nose feature identification and localization through self-organizing map and graph matching. *J. Circuits Syst. Comput.* **19**(1), 191–202 (2010). <https://doi.org/10.1142/S0218126610006062>
12. Cheng, Y., Church G.: Biclustering of expression data. In: Proceedings of the Eight International Conference Intelligent Systems for Molecular Biology (ISMB), pp. 93–103 (2000)
13. Cirrincione, G., Randazzo, V., Pasero, E: The Growing Curvilinear Component Analysis (GCCA) neural network. In: Neural Network, vol. 103, pp. 108–117 (2018). ISSN 0893-6080
14. Bevilacqua, V., et al.: 3D head pose normalization with face geometry analysis, genetic algorithms and PCA. *J. Circuits Syst. Comput.* **18**, 1425–1439 (2009). <https://doi.org/10.1142/S0218126609005769>

# Chapter 22

## Growing Curvilinear Component Analysis (GCCA) for Stator Fault Detection in Induction Machines



**Giansalvo Cirrincione, Vincenzo Randazzo, Rahul R. Kumar,  
Maurizio Cirrincione and Eros Pasero**

**Abstract** Fault diagnostics for electrical machines is a very difficult task because of the non-stationarity of the input information. Also, it is mandatory to recognize the pre-fault condition in order not to damage the machine. Only techniques like the principal component analysis (PCA) and its neural variants are used at this purpose, because of their simplicity and speed. However, they are limited by the fact they are linear. The GCCA neural network addresses this problem; it is nonlinear, incremental, and performs simultaneously the data quantization and projection by using the curvilinear component analysis (CCA), a distance-preserving reduction technique. Using bridges and seeds, it is able to fast adapt and track changes in the data distribution. Analyzing bridge length and density, it is able to detect a pre-fault condition. This paper presents an application of GCCA to a real induction machine on which a time-evolving stator fault in one phase is simulated.

### 22.1 Introduction

Data mining is more and more facing the extraction of meaningful information from big data (e.g., from Internet), which is often very high dimensional. For both visual-

---

G. Cirrincione  
University of Picardie Jules Verne, Lab. LTI, Amiens, France  
e-mail: [exin@u-picardie.fr](mailto:exin@u-picardie.fr)

G. Cirrincione · R. R. Kumar · M. Cirrincione  
University of South Pacific, SEP, Suva, Fiji Islands  
e-mail: [rahul.kumar@usp.ac.fj](mailto:rahul.kumar@usp.ac.fj)

M. Cirrincione  
e-mail: [maurizio.cirrincione@usp.ac.fj](mailto:maurizio.cirrincione@usp.ac.fj)

V. Randazzo (✉) · E. Pasero  
Politecnico di Torino, DET, Turin, Italy  
e-mail: [vincenzo.randazzo@polito.it](mailto:vincenzo.randazzo@polito.it)

E. Pasero  
e-mail: [eros.pasero@polito.it](mailto:eros.pasero@polito.it)

ization and automatic purposes, their dimensionality has to be reduced. This is also important in order to learn the data manifold, which, in general, is lower dimensional than the original data. Dimensionality reduction (DR) also mitigates the curse of dimensionality: e.g., it eases classification, analysis, and compression of high-dimensional data.

Most DR techniques work offline; i.e., they require a static database (batch) of data, whose dimensionality is reduced. They can be divided into linear and nonlinear techniques, the latter being in general slower, but more accurate in real-world scenarios. See [1] for an overview.

However, the possibility of using a DR technique working in real time is very important, because it allows not only having a projection after only the presentation of few data (i.e., a very fast projection response), but also tracking non-stationary data distributions (e.g., time-varying data manifolds). This can be applied, for example, to all applications of real-time pattern recognition, where the data reduction step plays a very important role: fault diagnosis, novelty detection, intrusion detection for alarm systems, speech, face and text recognition, computer vision and scene analysis, and so on.

In recent years, research in the field of fault diagnosis (FD) and condition monitoring (CM) of electrical machines has attracted researchers all over the world. This is because of its involvement in an endless number of industrial applications. The concept of FD and CM has always been a key issue for industries when it comes to maintaining the assets, especially large motors or generators, whose possible failures may pose serious repercussions in both monetary terms and non-monetary terms.

Early identification of incipient faults results in a quick maintenance and short downtime for processes under consideration. An ideal FD and CM system must be able to extract the required data and correctly detect and classify the fault incurred in the motor. In the most recent years, there has been a lot of research in the development of new CM schemes for electrical machines and drives, overseeing the downsides of the conventional techniques.

According to the authors of [2–4], the quantity of working machines in the world was expected to be around 16.1 billion in 2011, with a rapid development of 50% w.r.t. the preceding five years. Among these machines, induction machines (IMs) are the most common ones and are widely used in the industry. This derives from the fact that IMs are rugged, cheap, reasonably portable, sensibly high effective, and conform to the available power supplies. They are reliable in operations, yet are liable to various sorts of undesirable faults, which can be categorized as follows: mechanical faults, electrical faults, and outer motor drive faults. In view of rotating magnetic field, the IMs are incredibly symmetrical electrical systems, so any fault occurrence changes its symmetrical properties.

As per the statistics available from IEEE and EPRI for motor faults [5–7], stator-winding faults contribute to as much as 26% of the total number of failures in IMs. The stator-winding faults begin as an inter-turn short circuit, which evolves over time into a short circuit between coils and phase windings. Thus, it is fundamental that a diagnosis be made able to track them in real time [8, 9].

Working in real time requires a data stream, a continuous input for the DR algorithms, which are defined as online or, sometimes, incremental (synonym for non-batch). They require, in general, data drawn from a stationary distribution. The fastest algorithms are linear and use the principal component analysis (PCA [10]) by means of linear neural networks, like the generalized Hebbian algorithm (GHA [11]) and the incremental PCA (candid covariance-free CCIPCA [12]). Nonlinear DR techniques are not suitable for online applications. Many efforts have been tried in order to speed up these algorithms: updating the structure information (graph), new data prediction, embedding updating. However, these incremental versions (e.g., iterative LLE [13]) require too a cumbersome computational burden and are useless in real-time applications. Neural networks can also be used for data projection. In general, they are trained offline and used in real time (recall phase). In this case, they work only for stationary data and can be better considered as implicit models of the embedding. Examples are the self-organizing maps (SOMs) [14] and their variants [15–18].

For data drawn from a non-stationary distribution, as it is the case for fault and pre-fault diagnosis and system modeling, the online curvilinear component analysis (onCCA [19]) and the growing curvilinear component analysis (GCCA) have been proposed in [20, 21]. They both track non-stationarity by using an incremental quantization synchronously with a fast projection based on the curvilinear component analysis (CCA [22, 23]).

The purpose of this paper is the presentation of an application of GCCA to the stator-winding fault problem previously described with the purpose of detecting and following in real time the evolution of a fault in a phase of IM.

After the presentation of GCCA in Sect. 22.2, Section 22.3 shows the results of a fault simulation on a stator winding on a real IM. Finally, Section 22.4 presents the conclusions.

## 22.2 The Growing CCA (GCCA)

The growing CCA is an incremental supervised neural network whose number of neurons is determined by the quantization of the input space. Each neuron has associated two weight vectors: one in the input space (X-weight) and the other one in the latent space (Y-weight) which yields the data projection. Each neuron is equipped with a threshold which represents its Voronoi region in the data space. It is computed as the distance in the X-space between the neuron and its farthest neighbor (neighbors are defined by the edge graph) and is used for determining the novelty of the input data. If the input data passes the novelty test, a new neuron is created; otherwise, the closest neuron (the first winner) in the X-space and its neighbors adjust their weight vectors according to the soft competitive learning (SCL [19, 20]).

Neurons can be connected in two ways: through edges, which define the manifold topology according to the competitive Hebbian learning (CHL [24]), or through bridges, which track a change in the input distribution (e.g., a jump). GCCA uses bridges and seeds to understand how the input evolves over time. A bridge is a

particular kind of neurons link created to connect a new neuron to the already existing network. It is a directional link toward the new neuron. In this sense, it points toward the change in the input data. A seed is a pair of neurons made of a neuron and its doubled (whose weight is computed using the hard competitive learning, HCL [19, 20]). Neuron doubling is performed each time the first winner is the top of a bridge with the second-close neuron (the second winner). On the contrary, if the first winner is the tail of the bridge, that connection becomes an edge.

GCCA is incremental, it can increase or decrease (pruning by age) the number of neurons.

The projection algorithm is based on CCA. It uses a distance-preserving function which aims to preserve in the Y-space distances whose length is less than  $\lambda$ .

## 22.3 Stator-Winding Fault Experiment

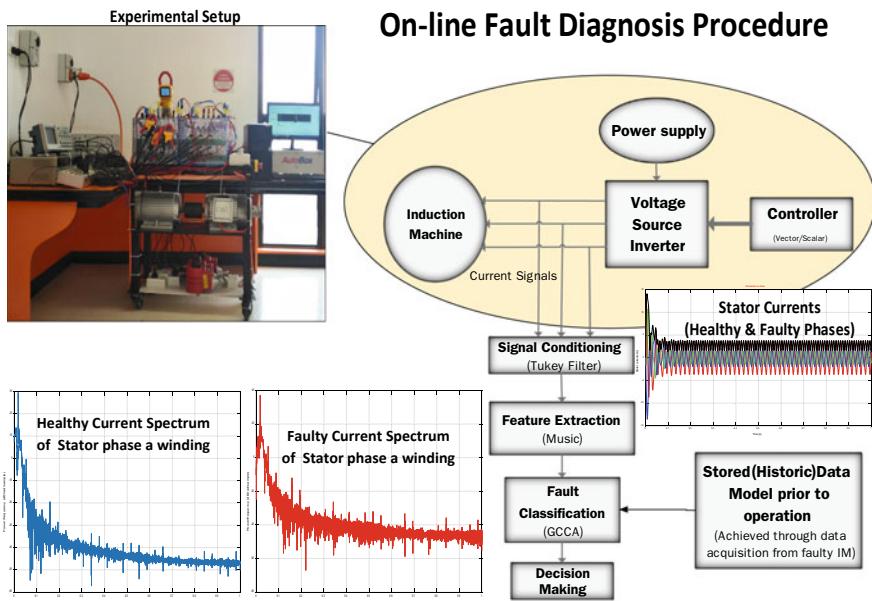
Using model-based techniques, a stator fault has been modeled and the temporal evolution of its current has been compared with the healthy case.

The dataset is generated for both the cases: healthy and faulty conditions of a three-phase squirrel cage IM which is of 1.1 kW rating and connected to a 60 Hz voltage supply. By using a preprocessing based on the Tukey filter, the signal-to-noise ratio (SNR) is increased from the acquired current signal. Thereafter, by using statistical signal processing, the frequencies of interest are extracted (see Fig. 22.1). Both the healthy and the faulty IMs are dynamically modeled in MATLAB®, and the current signature is acquired. The dataset consists of 35,685 samples taken in a span of seven seconds.

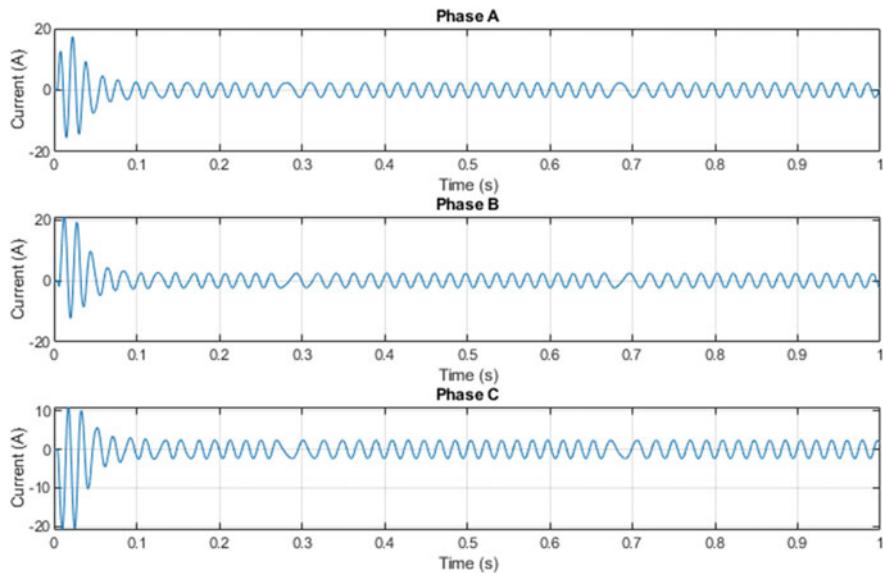
Figures 22.2 and 22.3 show, respectively, the three-phase current of the IM and the space vector representation (i.e., a two-phase current transformation by means of direct and quadrature currents) in the healthy case. Both figures are characterized by an initial transient (large oscillations in the current signature and corresponding decreasing spirals in the space vector representation), followed by a steady state (regular oscillation in Fig. 22.2 and circles in Fig. 22.3).

The inter-turn short circuit fault is induced in the IM by introducing a variable resistor in parallel with the phase A of the IM. The resistance was varied to correspond to a percentage of fault in the stator. From the starting ( $t = 0$ ), the IM is in healthy condition for one second, and after every second, the percentage of stator inter-turn fault rises by 5%. In particular, the current signature (see Fig. 22.4) in phase A rises every second, first portraying a transient stage (a spike in current signature each second) and then moving to a steady state. The other phases are also affected by a transient stage as the fault severity changes as shown in Fig. 22.4.

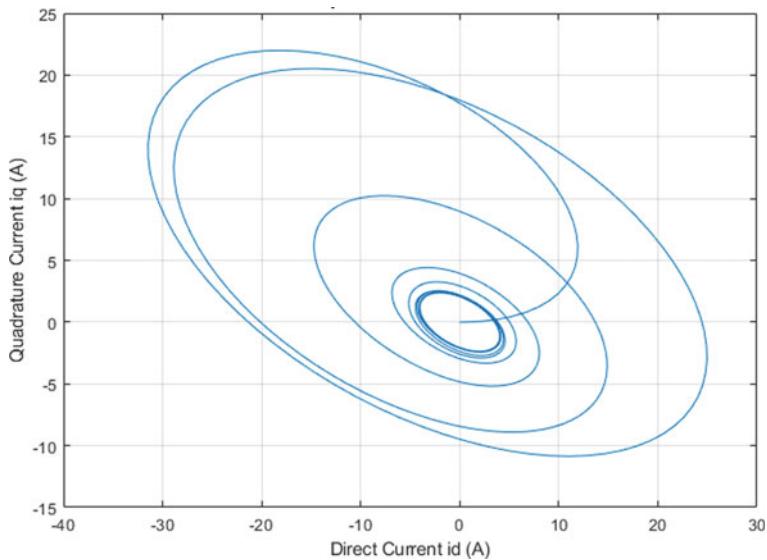
The interchange between transient and steady phases, i.e., the fault evolution over time, is also observed in the space vector representation (see Fig. 22.5 and its zoom in Fig. 22.6). As in the previous case, the space vector trajectories follow the same loci as before but with larger radii (they are larger and larger as the fault evolves).



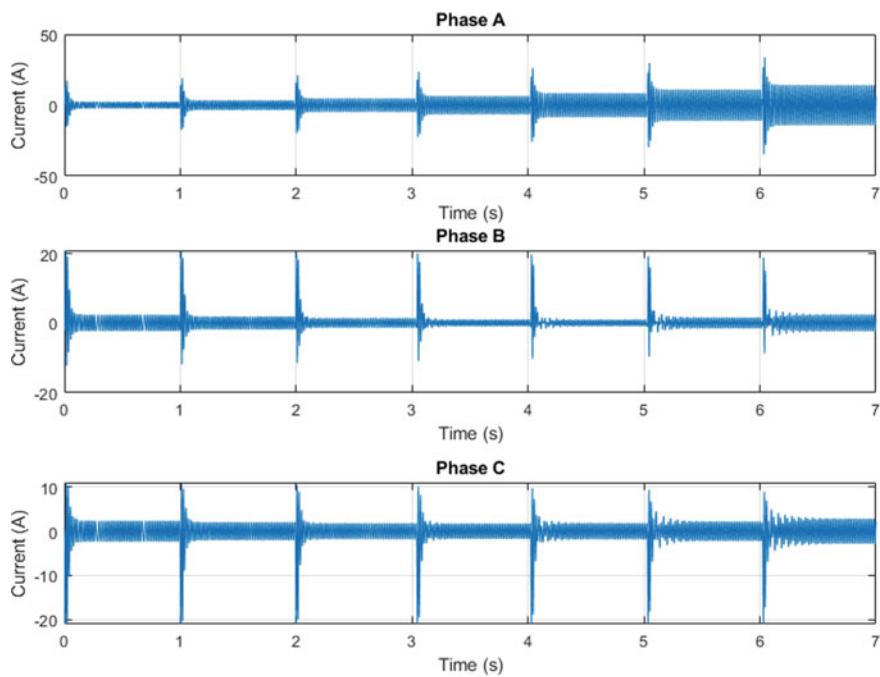
**Fig. 22.1** Proposed methodology



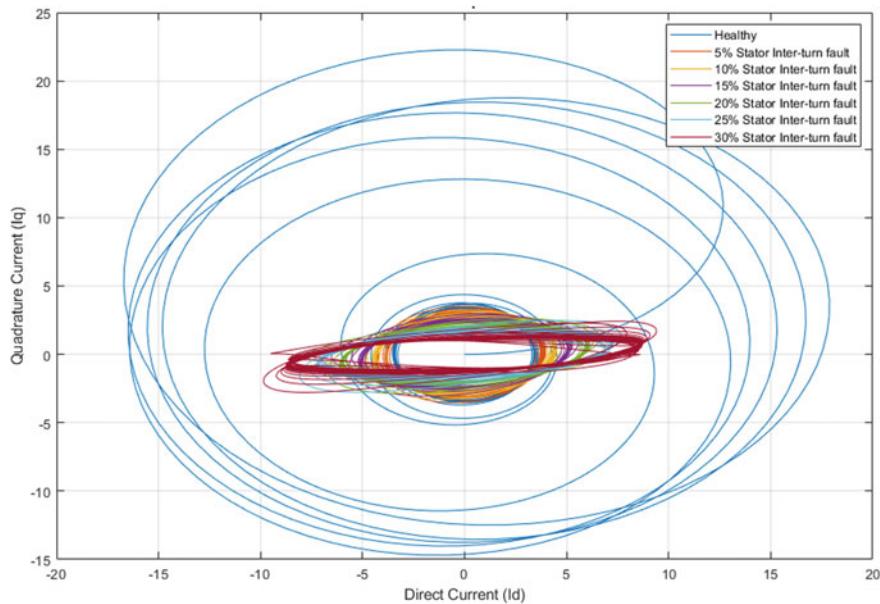
**Fig. 22.2** Three-phase current signature of a healthy IM



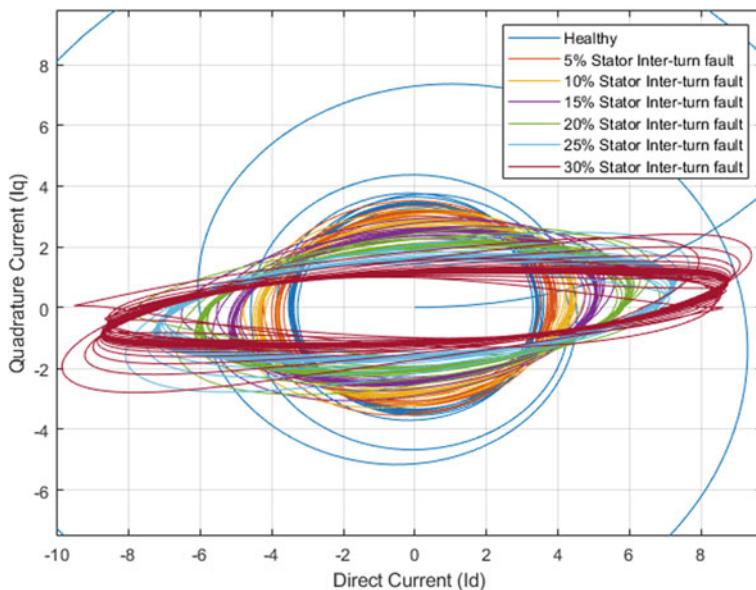
**Fig. 22.3** Space vector loci of stator current for a healthy IM



**Fig. 22.4** Fault evolution in IM from healthy to 30% stator inter-turn fault



**Fig. 22.5** Space vector loci—fault evolution from 0 to 30% stator inter-turn fault

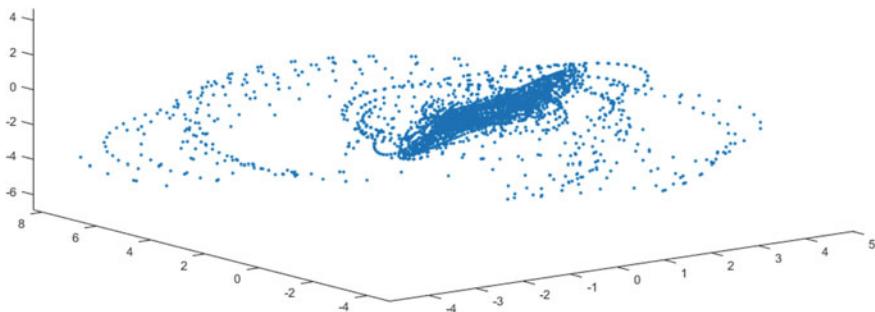


**Fig. 22.6** Fault evolution from 0 to 30% stator inter-turn fault: zoom around the origin

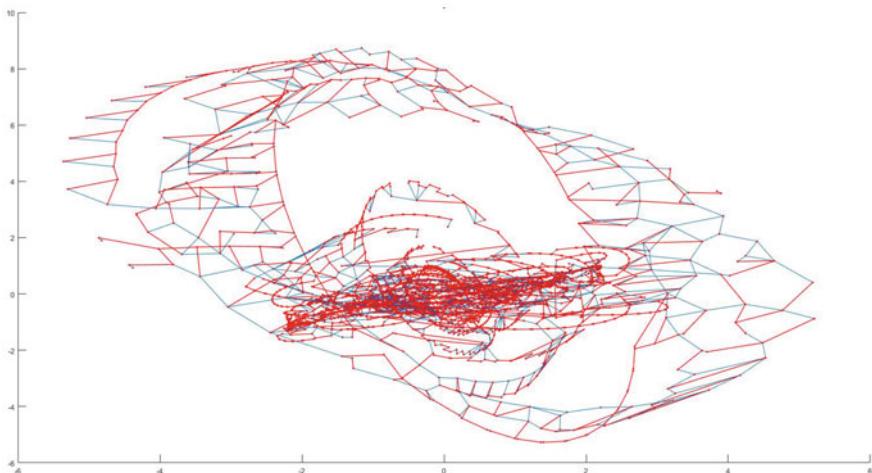
GCCA has been applied to this problem. The parameters of GCCA are the following:  $\alpha = 0.01$ ,  $\lambda = 0.5$ ,  $\alpha_1 = 0.2$ ,  $\alpha_n = 0.04$ ,  $\text{age}_{\max} = 4$ , epochs = 5. GCCA is trained with the phase current information and evolves with it. It also projects in the latent current space in real time.

Figure 22.7 shows the quantization made by the first layer of weights of GCCA (connections are not shown for clarity).

The trajectories have been modeled (tracked) accurately, and spirals and circles are visible. Figure 22.8, instead, illustrates the linking phase of the neural network. The first transient is represented by small edges and bridges, which are also orthogonal to the true current projection. This is due to the rapidity of the transient which does not allow their pruning. However, they build a fine small size network, which is typical of self-organization. As the transient evolves, more and more bridges track the changes



**Fig. 22.7** GCCA—fault evolution from 0 to 30% stator inter-turn fault—X-space quantization



**Fig. 22.8** GCCA—fault evolution from 0 to 30% stator inter-turn fault (edges in blue, bridges in red)

in current, which are represented by their density. Indeed, the appearance of bridges detects the onset of a non-stationarity. If the time change is abrupt, more and more bridges are created in a correlated way. This is the reason that the inner part of the plot is denser and denser.

Unlike other neural networks which need constant parameters in order to track non-stationarity, GCCA does not only recognize the pre-fault situation, but also records the whole story of the machine.

## 22.4 Conclusions

Time signals extracted from a non-stationary distribution, as in the case of the stator phase current, which evolves in the same way as the machine (faults and deterioration), are not easy to handle, above all in real time. This could have important applications, as, for instance, the possibility to stop the motor well before the fault or for maintenance. In the literature, only linear techniques are used, because of their speed and simplicity. Nonlinear techniques and neural networks are too cumbersome and time-consuming. The GCCA neural network is the only neural method able to track a non-stationary input distribution and to project it in a lower-dimensional space. In a sense, GCCA learns a time-varying manifold. It has been applied in a difficult test, like the tracking of evolutive faults on an electrical machine. It has been shown that it learns (and represents) the machine life, from the first transient to the last fault. However, this can be automatically exploited, by means of the bridge length and density estimation, in order to stop working for avoiding damages. Future work will deal with the exploitation of the stator current spectrum (by using algorithms like MUSIC) and observing the changes with respect to the stator current spectrum of a healthy IM. Because of this preprocessing step on the stator currents, GCCA should perform a faster and more reliable fault detection. It will also help in fault classification.

**Acknowledgements** This work has been partly supported by OPLON Italian MIUR project.

## References

1. Van der Maaten, L., Postma, E., Van der Herik, H.: Dimensionality reduction: a comparative review. TiCC TR 2009-005, Delft University of Technology (2009)
2. Henao, H., Capolino, G.A., Fernandez-Cabanas, M., Filippetti, F., Bruzzese, C., Strangas, E., et al.: Trends in fault diagnosis for electrical machines: a review of diagnostic techniques. IEEE Ind. Electron. Mag. **8**, 31–42 (2014)
3. Filippetti, F., Bellini, A., Capolino, G.A.: Condition monitoring and diagnosis of rotor faults in induction machines: state of art and future perspectives. In: 2013 IEEE Workshop on Electrical Machines Design, Control and Diagnosis (WEMDCD) (2013)

4. IEEE Recommended Practice for the Design of Reliable Industrial and Commercial Power Systems—Redline, IEEE Std 493-2007 (Revision of IEEE Std 493-1997)—Redline, pp. 1–426 (2007)
5. Karmakar, S., Chattopadhyay, S., Mitra, M., Sengupta, S.: Induction Motor Fault Diagnosis
6. Singh, G.: Induction machine drive condition monitoring and diagnostic research—a survey. *Electr. Power Syst. Res.* **64**, 145–158 (2003)
7. Toliyat, H.A., Nandi, S., Choi, S., Meshgin-Kelk, H.: Electric machines: modeling, condition monitoring, and fault diagnosis. CRC press (2012)
8. Stavrou, A., Sedding, H.G., Penman, J.: Current monitoring for detecting inter-turn short circuits in induction motors. *IEEE Trans. Energy Convers.* **16**, 32–37 (2001)
9. Pietrowski, W.: Detection of time-varying inter-turn short-circuit in a squirrel cage induction machine by means of generalized regression neural network. *COMPEL Int. J. Comput. Math. Electr. Electron. Eng.* **36** (2017)
10. Diamantaras, K.I., Kung, S.Y.: Principal Component Neural Networks: Theory and Applications. Wiley, Hoboken (1996)
11. Sanger, T.D.: Optimal unsupervised learning in a single-layer neural network. *Neural Netw.* **2**, 459–473 (1989)
12. Weng, J., Zhang, Y., Hwang, W.S.: Candid covariance-free incremental principal components analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(8), 1034–1040 (2003)
13. Kong, D., Ding, C.H.Q., Huang, H., Nie, F.: An iterative locally linear embedding algorithm. In: Proceedings of the 29th International Conference on Machine Learning (ICML) (2012)
14. Qiang, X., Cheng, G., Li, Z.: A survey of some classic self-organizing maps with incremental learning. In: 2nd International Conference on Signal Processing Systems (ICSPS), pp. 804–809 (2010)
15. Martinetz, T., Schulten, K.: A “Neural Gas” Network Learns Topologies, pp. 397–402. Artificial Neural Networks, Elsevier (1991)
16. Fritzke, B.: A growing neural gas network learns topologies. In: Advances in Neural Information Processing System, vol. 7, pp. 625–632. MIT Press (1995)
17. Martinetz, T., Schulten, K.: Topology representing networks. *Neural Netw.* **7**(3), 507–522 (1994)
18. Estevez, P., Figueira, C.: Online data visualization using the neural gas network. *Neural Netw.* **19**, 923–934 (2006)
19. Cirrincione, G., Héault, J., Randazzo, V.: The on-line curvilinear component analysis (onCCA) for real-time data reduction. In: International Joint Conference on Neural Networks (IJCNN), pp. 157–165 (2015)
20. Cirrincione, G., Randazzo, V., Pasero, E.: Growing Curvilinear Component Analysis (GCCA) for dimensionality reduction of nonstationary data. In: Esposito, A., Faudez-Zanuy, M., Morabito, F., Pasero, E. (eds.) Multidisciplinary Approaches to Neural Computing. Smart Innovation, Systems and Technologies, vol. 69. Springer, Cham (2018)
21. Kumar, R.R., Randazzo, V., Cirrincione, G., Cirrincione, M., Pasero, E.: Analysis of stator faults in induction machines using growing curvilinear component analysis. In: 2017 20th International Conference on Electrical Machines and Systems (ICEMS), pp. 1–6, Sydney, NSW (2017)
22. Demartines, P., Héault, J.: Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of data sets. *IEEE Trans. Neural Netw.* **8**(1), 148–154 (1997)
23. Sun, J., Crowe, M., Fyfe, C.: Curvilinear components analysis and Bregman divergences. In: Proceedings of the European Symposium on Artificial Neural Networks—Computational Intelligence and Machine Learning (ESANN), pp. 81–86, Bruges (Belgium) (2010)
24. White, R.: Competitive Hebbian learning: algorithm and demonstrations. *Neural Netw.* **5**(2), 261–275 (1992)

**Part III**

**Neural Networks and Pattern Recognition**

**in Medicine**

# Chapter 23

## A Neural Based Comparative Analysis for Feature Extraction from ECG Signals



Giansalvo Cirrincione, Vincenzo Randazzo and Eros Pasero

**Abstract** Automated ECG analysis and classification are nowadays a fundamental tool for monitoring patient heart activity properly. The most important features used in literature are the raw data of a time window, the temporal attributes and the frequency information from the eigenvector techniques. This paper compares these approaches from a topological point of view, by using linear and nonlinear projections and a neural network for assessing the corresponding classification quality. The nonlinearity of the feature data manifold carries most of the QRS-complex information. Indeed, it yields high rates of classification with the smallest number of features. This is most evident if temporal features are used: Nonlinear dimensionality reduction techniques allow a very large data compression at the expense of a slight loss of accuracy. It can be an advantage in applications where the computing time is a critical factor. If, instead, the classification is performed offline, the raw data technique is the best one.

### 23.1 Introduction

The standard procedure used by physicians to monitor heart is to measure and record its electrical activity through an electrocardiogram (ECG). A healthy ECG, shown in Fig. 23.1, presents six fiducial points (P, Q, R, S, T, U) which are correlated to the four principal stages of activity of a cardiac cycle: isovolumic relaxation, inflow, isovolumic contraction and ejection.

---

G. Cirrincione

University of Picardie Jules Verne, Lab. LTI, Amiens, France

e-mail: [exin@u-picardie.fr](mailto:exin@u-picardie.fr); [giansalvo.cirrincione@usp.ac.fj](mailto:giansalvo.cirrincione@usp.ac.fj)

University of South Pacific, SEP, Suva, Fiji Islands

V. Randazzo (✉) · E. Pasero

Politecnico di Torino, DET, Turin, Italy

e-mail: [vincenzo.randazzo@polito.it](mailto:vincenzo.randazzo@polito.it)

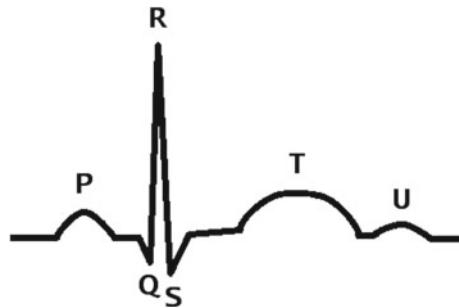
E. Pasero

e-mail: [eros.pasero@polito.it](mailto:eros.pasero@polito.it)

© Springer Nature Singapore Pte Ltd. 2020

A. Esposito et al. (eds.), *Neural Approaches to Dynamics of Signal Exchanges*,  
Smart Innovation, Systems and Technologies 151,

[https://doi.org/10.1007/978-981-13-8950-4\\_23](https://doi.org/10.1007/978-981-13-8950-4_23)

**Fig. 23.1** Healthy ECG

This path should repeat itself constantly over the time; otherwise, a person suffers from arrhythmias.

ECG recording is usually performed with the use of ten electrodes attached to a human body to analyze, at the same time, twelve leads, both peripherals (I, II, III, aVR, aVL, aVF) and precordials (V1, V2, V4, V5, V6). The recordings are, then, visually inspected by an expert, e.g., a cardiologist, looking for anomalies, i.e., diseases.

Several techniques for an automated ECG analysis have been proposed in literature. Adaptive filtering for noise canceling and arrhythmia detection is suggested in [1]. A fuzzy  $K$ -nearest neighbor classifier is used in [2]. Finally, ECG classification based on artificial neural networks has been adopted in [3–5]; an extensive review can be found in [6].

A fundamental phase prior to classification is the feature extraction. Indeed, high percentages of misclassifications are often due to an inappropriate feature selection [7–9]. Depending on the algorithm used for their extraction, features can be classified into two primary areas: temporal-based and eigenvectors-based. The former aims at exploiting the temporal evolution of the ECG signal (e.g., R-R variance); some examples can be found in [10, 11]. The latter, i.e., the Eigenvector method, is used for estimating frequencies of signals from noise-corrupted measurements; it is based on an eigen-decomposition of the correlation matrix of the signal. The two most used methods within this class are: Pisarenko [12] and MUSIC [13]. An application of these two to ECG classification can be found in [14–16].

This paper presents a comparative analysis about the classification performances of a multilayer perceptron (MLP) trained on six different datasets: ECG raw data, temporal features, eigenvector features and their projections using the curvilinear component analysis (CCA) [17]. First, Section 23.2 describes the proposed approach. Then, the results of the experiments are presented and discussed in Sect. 23.3.

## 23.2 The Proposed Approach

Unlike the traditional approach to ECG, which aims to improve the classification quality, and by considering that, in general, the results are very good, here the characteristics of the most important feature extraction techniques are analyzed in itself, for having a deeper insight in how they represent the QRS complex. At this aim, neural networks are used as tools for assessing the quality of the representation in order to evaluate the validity and properties of each technique. Here, the MLP network is used because it is well suited for pattern recognition [18]. At this purpose, it has a single hidden layer and five output units equipped with the soft-max activation function [18]. Because of the use of the cross-entropy error function [18], they yield the probability of membership for the following classes: normal beat, right bundle branch block beat, premature ventricular contraction, atrial premature contraction and other anomalies.

Each approach results in a different data manifold, which is here studied both by means of its intrinsic dimensionality and its level of nonlinearity by using CCA and the corresponding projected space visualization through the *dy-dx* diagram. CCA is a neural network which is able to project its input into a space of reduced dimensionality while preserving the manifold topology by means of local distance preservation. In this sense, it can be used to reduce the number of features without altering the original manifold. This is validated by the *dy-dx* plot, which is the plot of the distances of samples in the latent space (*dy*) versus the distances of corresponding samples in the data space (*dx*). In this scenario, it acts as a tool for the detection and analysis of nonlinearities. Generally, the more the deviation of data cloud with respect to the bisector, the more nonlinear the manifold is. Therefore, the input space can be reduced without losing information about the data.

The three main techniques, i.e., ECG raw data, temporal and eigenvector features, are then analyzed according to the number of features they require, the geometry of the representation (linear or nonlinear), the accuracy of the classification and the validity of their possible reductions (feature extraction).

## 23.3 Feature Analysis and Comparison

To test the proposed approach and its classification performance, several experiments have been conducted on the MIT-BIH Arrhythmia dataset [19–21]. First of all, it has been chosen because of its widespread use in research and the wide range of diseases covered. Moreover, each QRS complex within each record is labeled; hence, a supervised learning approach is quite straightforward. Also, the entire dataset is very well documented.

The chosen records are [22]: 106, 119, 200, 203, 207, 208, 209, 212, 231, 232, 233. For sake of simplicity, only the first 250.000 samples of the L2 lead of these records have been used for training and testing purposes. This should be not considered,

**Table 23.1** MLP classification results

	Original space (# Features)	Reduced space (# Features)
ECG raw data	99.1 (42)	89.4 (4)
<b>Temporal features</b>	96.0 (16)	<b>93.5 (6)</b>
<b>Eigenvector features</b>	90.3 (9)	88.4 (6)

at all, as a limitation of the proposed approach; indeed, L2 is, typically, the lead which carries most of information and is, in general, used as a reference for the interpretation of the others.

Six different datasets have been used to train and test the MLP: ECG raw data, temporal features, eigenvector features and their projections using CCA. The goal is the analysis of the dataset manifolds and the study of the most relevant subset of features for classification. Finally, in all the above cases, two-thirds of data have been used to train the network, while the remaining one-third has been used to test it.

### 23.3.1 ECG Raw Data

The first experiment deals with data extracted directly, i.e., without the feature extraction phase, from the MIT-BIH database. Each one of the above-cited records has been parsed in order to extract its QRS complexes. At this purpose, labels, which point to R-peaks time instants, have been used as the center of a 41-time instants window (twenty time instants before the one pointed by the label and twenty after it). In addition, the R-R time, i.e., the time between two consecutive R-peaks, has been added as last feature of this initial set. Consequently, the resulting training set is a matrix made of forty-two columns and as many rows as the number of QRS complexes. Then, two-thirds of this set, i.e., the training set, has been fed to the MLP with an input layer composed of 42 neurons and a hidden layer composed of 100 neurons. The confusion matrix resulting from the testing is shown in Fig. 23.2a. An overall accuracy of 99.1% is reached (see Table 23.1). This classification is very accurate. However, this method requires a lot of attributes, which are the raw sampled data of the temporal window. In this sense, there is no feature creation, which implies a very time-consuming algorithm.

The analysis of the data manifold by means of the principal component analysis (PCA) [18] suggests the intrinsic dimensionality is probably four (96.42% explained), as seen in Fig. 23.3. However, the nonlinearity of data has to be considered. At this aim, CCA is performed ( $\lambda = 70$ , epochs = 10), in order to project to a four-dimensional space. The corresponding  $dy-dx$  diagram, see Fig. 23.8 left, is concentrated around the bisector, which proves the manifold is nearly a hyperplane. If the projected data are fed to an MLP with one hidden layer of 20 neurons, the overall test performance is decreased to 89.4% (see Table 23.1 and Fig. 23.2b), that is, 9.78% loss of accuracy.

		Confusion Matrix						
		1	2	3	4	5		
Output Class	1	1937 20.1%	1 0.0%	9 0.1%	25 0.3%	8 0.1%	97.8% 2.2%	
	2	2 0.0%	1860 19.3%	0 0.0%	8 0.1%	1 0.0%	99.4% 0.6%	
	3	8 0.1%	0 0.0%	1863 19.3%	0 0.0%	7 0.1%	99.2% 0.8%	
	4	12 0.1%	0 0.0%	0 0.0%	1972 20.5%	0 0.0%	99.4% 0.6%	
	5	3 0.0%	0 0.0%	7 0.1%	0 0.0%	1907 19.8%	99.5% 0.5%	
		98.7% 1.3%	99.9% 0.1%	99.1% 0.9%	98.4% 1.6%	99.2% 0.8%	99.1% 0.9%	
		1	2	3	4	5		
		Target Class	Target Class	Target Class	Target Class	Target Class		

		Confusion Matrix						
		1	2	3	4	5		
Output Class	1	1576 16.4%	107 1.1%	44 0.5%	228 2.4%	25 0.3%	79.6% 20.4%	
	2	65 0.7%	1754 18.2%	5 0.1%	37 0.4%	10 0.1%	93.7% 6.3%	
	3	32 0.3%	12 0.1%	1643 17.1%	6 0.1%	185 1.9%	87.5% 12.5%	
	4	147 1.5%	2 0.0%	0 0.0%	1831 19.0%	4 0.0%	92.3% 7.7%	
	5	45 0.5%	0 0.0%	57 0.6%	14 0.1%	1801 18.7%	93.9% 6.1%	
		84.5% 15.5%	93.5% 6.5%	93.9% 6.1%	86.5% 13.5%	88.9% 11.1%	89.4% 10.6%	
		1	2	3	4	5		
		Target Class	Target Class	Target Class	Target Class	Target Class		

Fig. 23.2 ECG raw data confusion matrix: original (a) and reduced (b) space cases

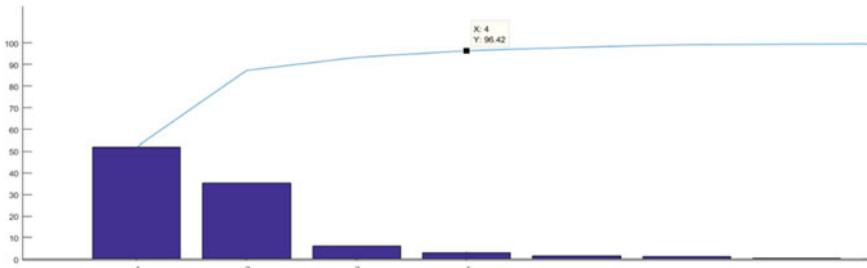


Fig. 23.3 ECG raw data PCA analysis

### 23.3.2 Temporal Features

The second dataset used to test the proposed approach is made of fifteen statistical features extracted from each record of the ECG raw data dataset. The selected features are the following: mean, max value, root mean square, square root mean, standard deviation, variance, shape factor (with RMS), shape factor (with SRM), crest factor, latitude factor, impulse factor, skewness, kurtosis, normalized 5th central moment and normalized 6th central moment. As before, the R-R time, i.e., the time between two consecutive R-peaks, has been added as last feature of this set. Data are statistically normalized (z-score). Two-thirds of this set, i.e., the training set, has been fed to the MLP. Here, the input layer is composed of 16 neurons and the hidden layer by 40.

The confusion matrix resulting from the testing is shown in Fig. 23.4a. An overall accuracy of 96.0% is reached (see Table 23.1). The classification is worsened with regard to previous method. However, this method requires fewer attributes: from 42 to 16, that is a nearly 61.9% reduction (Fig. 23.4).

		Confusion Matrix							
		1	2	3	4	5			
Output Class	1	1775 18.4%	29 0.3%	36 0.4%	46 0.5%	16 0.2%	93.3% 6.7%		
	2	13 0.1%	1907 19.8%	0 0.0%	18 0.2%	3 0.0%	98.2% 1.8%		
	3	40 0.4%	0 0.0%	1845 19.2%	5 0.1%	53 0.6%	95.0% 5.0%		
	4	46 0.5%	3 0.0%	1 0.0%	1872 19.4%	1 0.0%	97.3% 2.7%		
	5	19 0.2%	1 0.0%	51 0.5%	4 0.0%	1846 19.2%	96.1% 3.9%		
		93.8% 6.2%	98.3% 1.7%	95.4% 4.6%	96.2% 3.8%	96.2% 3.8%	96.0% 4.0%		

		Confusion Matrix							
		1	2	3	4	5			
Output Class	1	1719 17.9%	83 0.9%	30 0.3%	67 0.7%	44 0.5%	88.5% 11.5%		
	2	56 0.6%	1817 18.9%	0 0.0%	24 0.2%	3 0.0%	95.6% 4.4%		
	3	27 0.3%	0 0.0%	1789 18.6%	0 0.0%	102 1.1%	93.3% 6.7%		
	4	43 0.4%	3 0.0%	4 0.0%	1854 19.3%	15 0.2%	96.6% 3.4%		
	5	22 0.2%	3 0.0%	89 0.9%	13 0.1%	1823 18.9%	93.5% 6.5%		
		92.1% 7.9%	95.3% 4.7%	93.6% 6.4%	94.7% 5.3%	91.7% 8.3%	93.5% 6.5%		

Fig. 23.4 Temporal features confusion matrix: original (a) and reduced (b) space cases

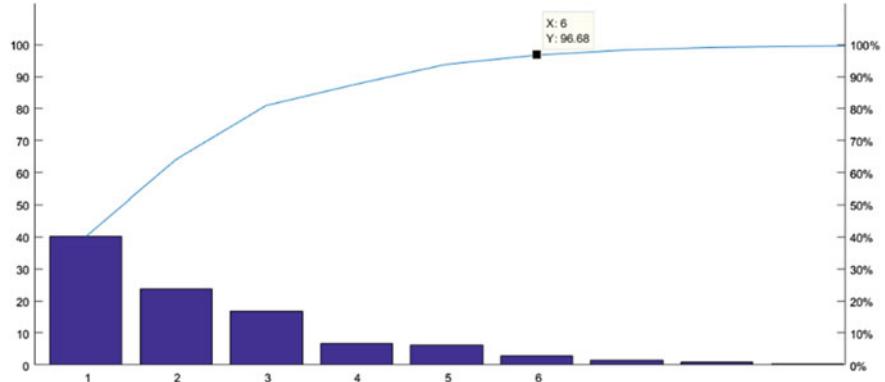
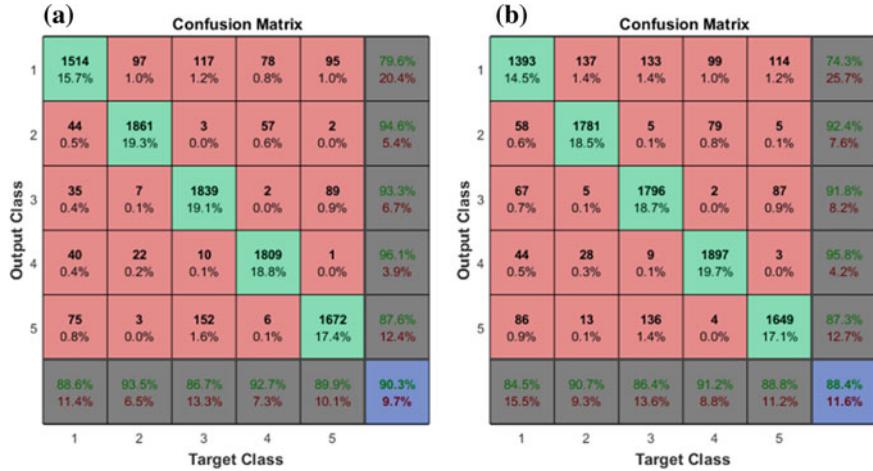


Fig. 23.5 Temporal features PCA analysis

Figure 23.5 shows the result of the PCA analysis: The intrinsic dimensionality is probably six (96.68% explained). In order to check the nonlinearity of the manifold, CCA is again performed ( $\lambda = 70$ , epochs = 10), by projecting to a six-dimensional space. The corresponding  $dy-dx$  diagram, see Fig. 23.8 middle, is less concentrated around the bisector. However, it is thicker for larger distances. The manifold is nonlinear, but locally linear (short distances are well preserved in the projection). If the six projected features are the inputs of an MLP with one hidden layer of 20 neurons, the overall test performance is decreased to 93.5% (see Table 23.1 and Fig. 23.4b), that is 2.6% loss of accuracy.



**Fig. 23.6** Eigenvector features confusion matrix: original (a) and reduced (b) space cases

### 23.3.3 Eigenvector Features

The third dataset, normalized with z-score, is made of eight features extracted from each record of the ECG raw data dataset using the MUSIC algorithm. Different subspace dimensions for the algorithm have been tried and compared in order to check how the classification performance varies versus this parameter. The best results have been obtained with a subspace of dimensionality equal to five. As before, the R-R time has been added as last feature of this set. Two-thirds of this set, i.e., the training set, has been fed to the MLP. Here, the input layer is composed of 9 neurons and the hidden layer of 40. The confusion matrix resulting from the testing is shown in Fig. 23.6a. An overall accuracy of 90.3% is reached (see Table 23.1). The classification is the worst, but still accurate. However, this method requires the smallest number of attributes: from 42 to 9, that is a nearly 78.6% reduction.

Figure 23.7 shows the result of the PCA analysis: The intrinsic dimensionality is probably six (99.12% explained). In order to check the nonlinearity of the manifold, CCA is again performed ( $\lambda = 30$ , epochs = 10), by projecting to a six-dimensional space. The corresponding  $dy-dx$  diagram, see Fig. 23.8 right, is similar to the corresponding temporal feature case. However, it is thicker for smaller distances. The manifold is still nonlinear, but locally less linear. If the six projected features are the inputs of an MLP with one hidden layer of 20 neurons, the overall test performance is decreased to 88.4% (see Table 23.1 and Fig. 23.6b), that is 2.1% loss of accuracy.

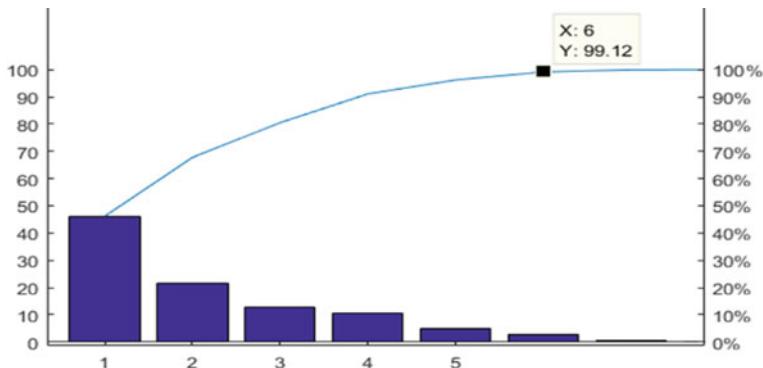


Fig. 23.7 Eigenvector features PCA analysis

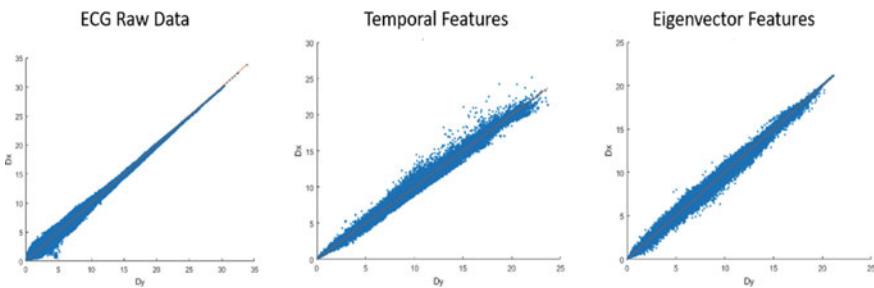


Fig. 23.8 CCA dy-dx diagrams

### 23.3.4 Discussion

All the experiments have shown a trade-off between smallest number of features and linearity. This is also more obvious in the case of dimensionality reduction. The raw data (no feature extraction) belong to a quasi-linear manifold in a four-dimensional space. Despite this simple geometry, the largest number of features (data in a time window) is required (forty-two values). Also, it is more evident when a nonlinear reduction to a space with the intrinsic dimensionality of data is performed: Indeed, the worst decrease in accuracy (9.78%) is observed.

The choice of feature extraction techniques, as the temporal and the eigenvector ones, implies an important economy in the number of attributes, but at the expense of a loss of linearity. The temporal features lie on a nonlinear manifold with local linearity. The MUSIC features also lie on the same kind of manifold, but the linearity exists only for smaller neighborhoods. The accuracy of the temporal method is close to the raw data one but requires only sixteen features (61.9% for a loss of only 2.9% of overall accuracy) and the minimum number of features (six) w.r.t. the classification performance (loss of 5.6%). The same observations can be repeated for the eigenvector technique. However, the classification is slightly worse.

This paper has shown a correlation between nonlinearity, number of features and accuracy. It can be concluded that the best representation of the QRS complexes is determined by the nonlinearity of the temporal features: 93.5% precision for only six features (extracted from sixteen attributes by means of CCA dimensionality reduction). On the other end, the large number of features of the raw data representation yields the best accuracy, but the simplicity of the manifold does not allow any good dimensionality reduction.

## References

1. Thakor, N.V., Zhu, Y.S.: Applications of adaptive filtering to ECG analysis: noise cancellation and arrhythmia detection. *IEEE Trans. Biomed. Eng.* **38**(8), 785–794 (1991)
2. Arif, M., Akram, M., Minhas, F.: Pruned fuzzy K-nearest neighbor classifier for beat classification. *J. Biomed. Sci. Eng.* **3**, 380–389 (2010)
3. El-Khafif, S.H., El-Brawany, M.A.: Artificial Neural Network-Based Automated ECG Signal Classifier. *ISRN Biomedical Engineering* (2013)
4. Vijaya, G., Kumar, V., Verma, H.K.: ANN-based QRS-complex analysis of ECG. *J. Med. Eng. Technol.* **22**(4), 160–167 (1998)
5. Randazzo, V., Pasero, E., Navaretti, S.: VITAL-ECG: a portable wearable hospital. In: 2018 IEEE Sensors Applications Symposium (SAS), pp. 1–6, Seoul, Korea (South) (2018)
6. Gambarotta, N., Aletti, F., Baselli, G., et al.: A review of methods for the signal quality assessment to improve reliability of heart rate and blood pressures derived parameters. *Med. Biol. Eng. Comput.* **1025**(7), 1025–1035 (2016)
7. Bevilacqua, V., Carnimeo, L., Mastronardi, G., Santarcangelo, V., Scaramuzzi, R.: On the comparison of NN-based architectures for diabetic damage detection in retinal images. *J. Circuits Syst. Comput.* **18**(08), 1369–1380 (2009)
8. Brunetti, A., Buongiorno, D., Trotta, G.F., Bevilacqua, V.: Computer vision and deep learning techniques for pedestrian detection and tracking: a survey. *Neurocomputing* **300**, 17–33 (2018)
9. Bevilacqua, V., Pietroleonardo, N., Triggiani, V., Brunetti, A., Di Palma, A.M., Rossini, M., Gesualdo, L.: An innovative neural network framework to classify blood vessels and tubules based on Haralick features evaluated in histological images of kidney biopsy. *Neurocomputing* **228**, 143–153 (2017)
10. Cavalcanti Roza, V.C., De Almeida, A.M., Postolache, O.A.: Design of an artificial neural network and feature extraction to identify arrhythmias from ECG. In: 2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA), pp. 391–396. Rochester, MN (2017)
11. Cuesta-Frau, D., Pérez-Cortés, J.C., Andreu-García, G.: Clustering of electrocardiograph signals in computer-aided Holter analysis. *Comput. Methods Programs Biomed.* **72**(3), 179–196 (2003)
12. Pisarenko, V.F.: The retrieval of harmonics from a covariance function. *Geophys. J. Int.* **33**(3), 347–366 (1973)
13. Schmidt, R.: Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas Propag.* **34**(3), 276–280 (1986)
14. Übeyli, E.D.: Combining recurrent neural networks with eigenvector methods for classification of ECG beats. *Digit. Signal Proc.* **19**(2), 320–329 (2009)
15. Übeyli, E.D., Cvetkovic, D., Cosic, I.: Analysis of human PPG, ECG and EEG signals by eigenvector methods. *Digit. Signal Proc.* **20**(3), 956–963 (2010)
16. Zgallai, W., Sabry-Rizk, M., Hardiman, P., O’Riordan, J.: MUSIC-based bispectrum detector: a novel non-invasive detection method for overlapping fetal and mother ECG signals. In:

- 19th Annual International Conference Proceedings of the IEEE Engineering in Medicine and Biology Society, pp. 72–75. IEEE, Chicago, IL (1997)
- 17. Demartines, P., Herault, J.: Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of data sets. *IEEE Trans. Neural Netw.* **8**(1), 148–154 (1997)
  - 18. Bishop, C.M.: *Neural Networks for Pattern Recognition*. Oxford University Press (1995)
  - 19. Moody, G.B., Mark, R.G.: The impact of the MIT-BIH Arrhythmia Database. *IEEE Eng. Med. Biol.* **20**(3), 45–50 (2001)
  - 20. Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, PCh., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.K., Stanley, H.E.: PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* **101**(23), 215–220 (2000)
  - 21. MIT-BIH Arrhythmia Database. <https://www.physionet.org/physiobank/database/mitdb/>. Last accessed 19 April 2018
  - 22. MIT-BIH Arrhythmia Database Directory. <https://www.physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm>

## Chapter 24

# A Multi-modal Tool Suite for Parkinson's Disease Evaluation and Grading



Giacomo Donato Cascarano, Antonio Brunetti, Domenico Buongiorno, Gianpaolo Francesco Trotta, Claudio Loconsole, Ilaria Bortone and Vitoantonio Bevilacqua

**Abstract** The traditional diagnosis of Parkinson's disease (PD) aims to assess several clinical manifestations, and it is commonly based on medical observations. However, an overall evaluation is extremely difficult due to the large variety of symptoms that affect PD patients. Furthermore, the traditional PD assessment is based on visual subjective observation of different motor tasks. For this reasons, an automatic system could be able to automatically assess and rate the PD and objectively evaluate the performed motor tasks. Such system could then support medical specialists in the assessment and rating of PD patients in a real clinical scenario. In this work, we developed multi-modal tool suite able to extract and process meaningful features from different motor tasks by means of two main experimental set-ups. In detail, we acquired and evaluated the motor performance acquired during the finger tapping, the foot tapping and the hand writing exercises. Several sets of features have been extracted from the acquired signals and used to both successfully classify a subject as PD patient or healthy subject, and rate the disease among PD patients.

### 24.1 Introduction

Parkinson's disease (PD) is one of the most spread neurodegenerative disorders. PD is a degenerative brain disorder characterized by a loss of midbrain dopamine DA neurons [1]. The main clinical PD symptoms involving body movements are: tremor, rigidity, bradykinesia, and gait abnormalities. Common physician evaluations

---

G. D. Cascarano · A. Brunetti · D. Buongiorno · C. Loconsole · V. Bevilacqua (✉)  
Department of Electrical and Information Engineering,  
Polytechnic University of Bari, Bari, Italy  
e-mail: [vitoantonio.bevilacqua@poliba.it](mailto:vitoantonio.bevilacqua@poliba.it)

G. F. Trotta  
Department of Mechanics, Mathematics and Management Engineering,  
Polytechnic University of Bari, Bari, Italy

I. Bortone  
Institute of Clinical Physiology, National Research Council, Pisa, Italy

are based on the motor function ability during the daily living activities and specific clinical observations using Unified PD Rating Scale (UPDRS), and the “Hoehn and Yahr” staging scale [2]. The UPDRS is a numeric scale widely used to asses PD severity. Even if the validity of the UPDRS has been proved in several scientific studies, the subjectivity and low efficiency are inevitable leading to a not quantified diagnostic basis. In particular, the main problems regard the evaluation of the severity of specific symptoms such as freezing of gait [3–6], dysarthria [7], tremor [8–12], bradykinesia [13–15] and dyskinesia [16–21]. Therefore, the development of computer-assisted diagnosis system could be useful to automatically evaluate and assess the disease.

Another interesting research field focuses on the analysis of different common life tasks such as handwriting. The handwriting is a highly over-learned fine and complex manual skill involving an intricate blend of cognitive, sensory and perceptual-motor components [22, 23]. For these reasons, the presence of abnormality in the handwriting process is a well-known and well-recognized manifestation of a wide variety of neuromotor diseases. There are two main difficulties related to the handwriting and affecting Parkinson’s disease (PD) patients: (i) the difficulty in controlling the amplitude of the movement, i.e. decreased letter size (micrography) and failing in maintaining stroke width of the characters as writing progresses [24], and (ii) the irregular and bradykinetic movements, i.e. increased movement time, decreased velocities and accelerations, and irregular velocity and acceleration trends over time [25]. For these reasons, in the literature, there are several works investigating the possibility of a differentiation between PD patients and healthy subjects by means of computer-aided handwriting analysis tools.

According to the last trends in machine learning applied in the medicine field [26–38], the main goal of this study is to provide a tool suite to support clinicians in the objective assessment of the typical PD motor issues and alterations. The assessment has been done by means of an overall integration of different information sources, i.e. the integration of several features acquired during the execution of different motor tasks. We focused on the independent analysis of three main motor tasks: finger tapping, foot tapping and handwriting. We then used machine learning techniques to detect and classify the status of the PD disease enabling the continuous monitoring of the disease progress. The main novel contributions with respect to the state of the art are the design, the development and the evaluation with both healthy subjects and PD patients of two systems: a vision-based system able to capture specific hand and foot movements, and a handwriting analysis tool. Furthermore, we developed and compared several classifiers able to assess and rate the movement impairment using a specific set of features.

## 24.2 Materials and Methods

### 24.2.1 Participants

The two experiments have been conducted on two different PD groups, and for each of them, an age-matched control group has been properly selected.

Regarding the UPDRS motor task analysis, i.e. the finger tapping and the foot tapping, 33 PD patients (mean age 71.6 years, SD 9.0, age range 54–87) and 29 healthy subjects (mean age 71.1 years, SD 9.2, age range 57–90) participated in the experiments after giving a written informed consent. The 33 PD patients were examined by a medical doctor and rated according to MDS-UPDRS Part IV for motor complications that considers a scoring with five level (normal, slight, mild, moderate and severe). In detail, 14 (mean age 67.2 years, SD 9.8, age range 54–81) and 19 (mean age 74.1 years, SD 7.1, age range 63–87) patients were classified as mild and moderate PD patients, respectively. None of the patients was classified as either slight or severe PD patient.

Concerning the handwriting analysis, 32 participants (21 males, 11 females, age:  $71.4 \pm 8.3$  years old) took part in the experimental tests. In detail, the age-matched control group was composed of 11 healthy subjects (4 males, 7 females, age:  $70.2 \pm 10.2$  years old), whereas the PD group was composed of 21 subjects (17 males, 4 females, age:  $72.1 \pm 8.3$ ). According to the degree of the disease, the PD group was divided into two subgroups: mild and moderate. The mild group was composed of 12 patients (9 males, 3 females, age:  $70.5 \pm 10.0$ ), whereas the moderate one was composed of 9 patients (8 males, 1 female, age:  $73.8 \pm 6.0$ ).

### 24.2.2 Experimental Set-up

We designed two different system set-ups able to acquire and extract the main features we are interested in.

Regarding the assessment of the MDS-UPDRS tasks, we chose to design and develop a system able to reproduce and record two tasks: finger tapping and foot tapping. We developed two separate vision-based systems able to acquire the movement of the thumb, the index finger and the toes. Both acquisition systems are based on passive markers made of reflective material and the Microsoft Kinect One RGBD camera. A brief and detailed description of two acquisition systems follows:

- Finger tapping exercise set-up: this test considers the examination of both hands separately. The tested subject is seated in front of the camera and is instructed to tap the index finger on the thumb ten times as quickly and as big as possible. During the task, the subject wears two thimbles made of a reflective material on both the index finger and thumb.

- Foot tapping exercise set-up: the feet are tested separately. The tested subject sits in a straight-backed chair in front of the camera and has both feet on the floor. He is then instructed to place the heel on the ground in a comfortable position and then tap the toes ten times as big and as fast as possible. A system of stripes with a reflective material is positioned on the toes.

For the handwriting analysis, instead, the system set-up includes two main sensors: (i) the Myo gesture control armband that allows us to synchronously acquire 8 different sEMG sources of the forearm, and (ii) the Wacom Cintiq 13" HD, a graphics tablet providing visual feedback for acquiring pen tip planar coordinates and pressure, and the tilt of the pen with respect to the writing surface. For the experiments, we used three writing patterns (WPs) leading to as many writing tasks; these are: a five-turn spiral drawn in anticlockwise direction (WP 1), a sequence of 8 Latin letter "l" with a size of 2.5 cm (WP 2) and with a size of 5 cm (WP 3). Since the last two WPs were size-constrained, a visual marker was provided as reference. In the experiment, we asked each subject to perform the three writing tasks four times each for a total of twelve tasks: first for familiarization purposes, whereas the other three were acquired and stored for the subsequent feature extraction and processing. The subject was asked to rest between two subsequent handwriting tasks for at least three seconds. The beginning of the task signal acquisition was triggered by a positive pen pressure applied on the graphic tablet. The processing of the acquired raw signals led to the extraction of several features.

### **24.2.3 Feature Extraction**

The two vision-based acquisition systems used in the first experimental set-up is based on passive reflective markers to track the position of the thumb, the index finger and the toes. After the movement acquisition, an image processing phase is needed to recognize the marker in each acquired video frame and compute the 3D position of a centroid point associated with the specific marker. This post-processing phase has been conducted using the OpenCV library running the following steps on each image frame: (i) grayscale image conversion; (ii) threshold operation to extract the pixels associated with the reflective passive markers; (iii) several blur and erosion phases. Following the blobs are extracted using an edge detection procedure and only the blobs having sizes comparable with markers size are kept for the next analysis. As final step, the centroid of each blob (only one blob for the foot tapping and two blobs for the finger tapping) is computed. Given the position of the centroid, its depth information and the intrinsic parameters of the used camera, we then computed the 3D position of the centroid associated with each tracked marker in the camera reference system. Such centroid has been then considered as the position of the specific finger or of the foots toes. As a result, the 3D positions of toes' marker (Foot Tapping) and of the two fingers' markers (Finger Tapping) have been produced. Given the position of each marker, we then extracted the following signals over time:

- $d1(t)$  - the distance between the two fingers' markers over time (finger tapping);
- $d2(t)$  - the distance between the position of the toes marker over time and the position of the same marker when the toes are completely on the ground (foot tapping).

Both signals have been normalized to make them range in [0, 1]. Given the entire acquired signal, all the single trials (ten finger tappings and ten foot tappings) have been extracted for each side. We then extracted the same set of features for both computed signals, i.e.  $d1(t)$  and  $d2(t)$ . The set of the extracted features contains features of the time domain, space domain and frequency domain. In particular, the features are:

- meanTime: averaged execution time of the single exercise trial;
- varTime: variance of the execution time of the single exercise trial;
- meanAmplitude: averaged space amplitude of the single exercise trial;
- varAmplitude: variance of the space amplitude of the single exercise trial;
- tremors: number of peaks detected during the entire acquisition;
- hesitations: number of amplitude peaks detected in the velocity signal during the entire acquisition;
- periodicity: periodicity of the exercise computed as reported in [39];
- AxF: (amplitude times frequency) the averaged value of the division between the amplitude peak reached in a single exercise trial and the time duration of the trial.

For the handwriting analysis, instead, starting from biometric signals acquired during the handwriting tasks, we extracted several features. In particular, it is possible to group the proposed features into two categories:

- sEMG-related and pen-tip-related features:
  - Root-mean-square (RMS) features extracted for each sEMG channel. RMS is computed as the square root of the mean of the sample squares.
  - Zero-crossing (ZC) features, an index related to the signal sign variation. To normalize the features among the subjects, its value is divided by the length of the signal.
- Pen-tip-related features—these features are extracted from the signals generated by a graphic tablet during the handwriting task:
  - Cartesian and XY features are referred to the pen tip position and are extracted starting from the XY axes position: Cartesian and XY (i) velocity, (ii) acceleration, and (iii) jerk. This leads to a total of nine signals.
  - Pen tip pressure feature—a scalar feature and corresponds to the pressure applied by the pen tip on the surface of the tablet.
  - Azimuth and altitude feature: the azimuth feature is the value of the angle between a reference direction (e.g. the Y axis of the tablet) and the pen direction projected on the horizontal plane. The altitude feature is the value of the angle between the pen direction and the horizontal plane.

- Pattern-specific features associated with a specific writing pattern (WP). For letter-based WPs, the features are mainly related to the writing size, whereas for the spiral-based WPs, the features are mainly related to the writing precision. For the features extracted from the letter-based WPs, the upper and the lower peaks of the  $Y$ -coordinate of the pen tip position are computed and, then, used as input data of a linear regressor. Finally, the angle  $\alpha$  between the  $R_{\text{up}}$  and  $R_{\text{low}}$  regression lines and the coefficient of determination ( $R^2$ ) are computed and selected as features. For spiral WPs, instead, the feature extracted is an index representative of the variability of the strokes. For each point  $P$  of the  $X$ - $Y$  pen tip position, the vector  $\vec{r}$  with respect to the spiral centroid point  $C$  having origin in  $P$  is computed. The angle  $\beta$  between  $\vec{r}$  and the direction vector  $\vec{d}$  tangent to the spiral in  $P$  is, then, calculated. The spiral precision index feature is the standard deviation of the  $\beta$  angles computed for each point  $P$ .

As a result, three different feature datasets were created:

- the first dataset including only the features extracted from writing pattern 1 (41 features);
- the second dataset including only the features extracted from writing pattern 2 (43 features);
- the third dataset including only the features extracted from writing pattern 3 (43 features).

#### **24.2.4 Classification**

As reported above, in this study, we investigated whether the set of feature extracted can be used both to classify a subject either as healthy or PD affected and to infer the rate of the disease. In particular, in this study, we focused on the discrimination between mild and moderate PD patients.

**Healthy subjects versus PD patients.** In the first part of the study, we decided to analyse the capability of discrimination by means of two different machine learning approaches, one for each study. For the motor tasks, we trained three different binary support vector machine (SVM) classifiers based on following three set of features: all the finger tapping features (Set 1), all the foot tapping features (Set 1) and both finger and foot tapping features (Set 3). The handwriting features, instead, have been used to train a classifier based on artificial neural network (ANN) and the optimal topology was found by exploiting a multi-objective genetic algorithm (MOGA) and by maximizing the average test accuracy on a certain number of training, validation and test iterations [40].

**Mild PD patients versus Moderate PD patients.** In the second part of the study, we focused on the “Mild PD patients versus Moderate PD patients” classification. As

for the “Healthy subjects versus PD patients” classification, we used two different classification approaches, one for each reported study.

### 24.3 Results and Discussion

In this section, we report the classification results obtained for each studies listed above and discuss the main findings. In particular, we reported and analysed the best results in terms of performance by means with the following indices:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (24.1)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \quad (24.2)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (24.3)$$

where TP, TN, FP and FN stand for true positive, true negative, false positive and false negative, respectively.

**Healthy subjects versus PD patients classification.** In this subsection, we reported and discussed the results about the “Healthy subjects versus PD patients” classification. In particular, we refers to the patients with positive and to the healthy subjects with negative. For the UPDRS-based tasks, we obtained:

- Finger tapping features set: the best classifier was the Gaussian SVM with an accuracy of 71.0%, a sensitivity of 75.7% and a specificity of 65.5%.
- Foot tapping features set: the best classifier was the Gaussian SVM with an accuracy of 85.5%, a sensitivity of 91.0% and a specificity of 79.0%.
- Both finger and foot tapping feature sets: the best classifier was the quadratic SVM with an accuracy of 87.1%, a sensitivity of 87.8% and a specificity of 86.0%.

For the handwriting datasets, we obtained:

- *Dataset with all 41 features extracted from writing pattern 1*: the optimal topology ANN featured 2 hidden layers (respectively composed of 186 and 15 neurons), and 2 neurons for the output layer. The activation function found by the MOGA was logsig for the hidden layers. For the output layer, the softmax function was preliminary selected as activation function. Accuracy: 90.76% (std = 0.0764), specificity: 0.8530 (std = 0.1553), sensitivity: 0.9389 (std = 0.0720).
- *Dataset with all 43 features extracted from writing pattern 2*: the optimal topology ANN featured 2 hidden layers (respectively composed of 44 and 10 neurons), and 2 neurons for the output layer. The activation function found by the MOGA was logsig for the hidden layers. For the output layer, the softmax function was

preliminary selected as activation function. Accuracy: 92.98% (std = 0.0523), specificity: 0.8970 (std = 0.1212), sensitivity: 0.9486 (std = 0.0587).

- *Dataset with all 43 features extracted from writing pattern 3:* the optimal topology ANN featured 3 hidden layers (respectively composed of 232, 82 and 7 neurons), and 2 neurons for the output layer. The activation function found by the MOGA was logsig for the hidden layers. For the output layer, the softmax function was preliminary selected as activation function. Accuracy: 95.95% (std = 0.0479), specificity: 0.9425 (std = 0.0831), sensitivity: 0.9691 (std = 0.0575).

**Mild PD patients versus Moderate PD patients classification.** In this subsection, we reported and discussed the results about the “Mild PD patients versus Moderate PD patients” classification. In particular, we refers to the moderate PD patients with positive and to the mild PD patients with negative.

For the UPDRS-based tasks, we obtained:

- Finger tapping features set: the best classifier was the Gaussian SVM with an accuracy of 57.0%, a sensitivity of 100% and a specificity of 0%.
- Foot tapping features set: the best classifier was the Gaussian SVM with an accuracy of 81.0%, a sensitivity of 84.0% and a specificity of 78.0%.
- Both finger and foot tapping feature sets: the best classifier was the Gaussian SVM with an accuracy of 78.0%, a sensitivity of 89.0% and a specificity of 64.0%.

For the handwriting datasets, we obtained:

- *Dataset with all 41 features extracted from writing pattern 1:* the optimal topology ANN featured 3 hidden layers (respectively composed of 59, 65 and 2 neurons), and 2 neurons for the output layer. The activation function found by the MOGA was logsig for the hidden layers. For the output layer, the softmax function was preliminary selected as activation function. Accuracy: 94.34% (std = 0.0626), specificity: 0.9595 (std = 0.0763), sensitivity: 0.9220 (std = 0.1158).
- *Dataset with all 43 features extracted from writing pattern 2:* the optimal topology ANN featured 3 hidden layers (respectively composed of 133, 18 and 1 neurons), and 2 neurons for the output layer. The activation function found by the MOGA was logsig for the hidden layers. For the output layer, the softmax function was preliminary selected as activation function. Accuracy: 87.26% (std = 0.0850), specificity: 0.8720 (std = 0.1206), sensitivity: 0.8733 (std = 0.1544).
- *Dataset with all 43 features extracted from writing pattern 3:* the optimal topology ANN featured 3 hidden layers (respectively composed of 65, 36 and 7 neurons), and 2 neurons for the output layer. The activation function found by the MOGA was logsig for the hidden layers. For the output layer, the softmax function was preliminary selected as activation function. Accuracy: 91.86% (std = 0.0830), specificity: 0.9169 (std = 0.1167), sensitivity: 0.9220 (std = 0.1286).

## 24.4 Conclusion

In this work, we conducted a parallel analysis of two different tools for Parkinson disease assessment. The proposed technique is based on two different set-ups: a first vision-based tool able to automatically track two of the main exercises evaluated by the UPDRS scale and a second one based on the extraction of different features from biometric signals acquired during the handwriting task. Both the approaches allowed us to compute different features, subsequently used to find a separation between healthy and PD subjects and between two different PD grades; the separation has been achieved by means of two different classification approaches: SVM for the UPDRS-based tasks and ANN for the handwriting analysis. Results showed that an SVM using the features extracted by both considered UPDRS exercises, i.e. finger tapping and foot tapping, is able to classify between healthy subjects and PD patients with great performances by reaching 87.1% of accuracy, 86.0% of specificity and 87.8% of sensitivity. The results of the classification between mild and moderate PD patients indicated that the foot tapping features are more representative than the finger tapping ones. In fact, the SVM based on the foot tapping features reached the best score in terms of accuracy (81.0%) and specificity (78.0%). The results obtained from the handwriting analysis, instead, showed that healthy subjects can be differentiated from PD patients and that mild PD patients can be differentiated from moderate PD patients both with a high classification accuracy (over 90%).

From our findings, we can conclude that both the presented approaches could lead to a comprehensive tool suite for evaluating and grading of Parkinson disease and, at the same time, able to synthesize different aspects of subject symptoms by means of the analysis of different source of information.

**Acknowledgements** This work has been supported by the Italian project RoboVir (within the BRIC INAIL-2017 programme).

## References

1. Twelves, D., Perkins, K.S.M., Uk, M., Counsell, C.: Systematic review of incidence studies of Parkinson's disease. *Mov. Disord.* **18**(1), 19–31 (2003)
2. Goetz, G., Poewe, W., Rascol, O., Sampaio, C., Stebbins, G.T., Counsell, C., Giladi, N., Holloway, R., Moore, C.G., Wenning, G.K., Yahr, M.D., Seidl, L.: Movement disorder society task force report on the Hoehn and Yahr staging scale: status and recommendations. *Mov. Disord.* **19**(9), 1020–1028 (2004)
3. Bortone, I., Buongiorno, D., Lelli, G., Di Candia, A., Casciarano, G.D., Trotta, G.F., Fiore, P., Bevilacqua, V.: Gait analysis and Parkinson's disease: recent trends on main applications in healthcare. In: Masia, L., Micera, S., Akay, M., Pons, J.L. (eds.) *Converging Clinical and Engineering Research on Neurorehabilitation III*, pp. 1121–1125. Springer, Cham (2019)
4. Djuric-Jovicic, M.D., Jovicic, N.S., Radovanovic, S.M., Stankovic, I.D., Popovic, M.B., Kostic, V.S.: Automatic identification and classification of freezing of gait episodes in Parkinson's disease patients. *IEEE Trans. Neural Syst. Rehabil. Eng.* **22**(3), 685–694 (2014)

5. Tripoliti, E.E., Tzallas, A.T., Tsipouras, M.G., Rigas, G., Bougia, P., Leontiou, M., Konitsiotis, S., Chondrogiorgi, M., Tsouli, S., Fotiadis, D.I.: Automatic detection of freezing of gait events in patients with Parkinson's disease. *Comput. Methods Prog. Biomed.* **110**(1), 12–26 (2013)
6. Bortone, I., Trotta, G.F., Brunetti, A., Cascarano, G.D., Loconsole, C., Agnello, N., Argentiero, A., Nicolardi, G., Frisoli, A., Bevilacqua, V.: A novel approach in combination of 3D gait analysis data for aiding clinical decision-making in patients with Parkinson's disease. In: *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10362, pp. 504–514. LNCS (2017)
7. Tsanas, A., Little, M., McSharry, P.E., Ramig, L.O.: Accurate telemonitoring of parkinsons disease progression by noninvasive speech tests. *IEEE Trans. Biomed. Eng.* **57**(4), 884–893 (2010)
8. Mellone, S., Palmerini, L., Cappello, A., Chiari, L.: Hilbert-huang-based tremor removal to assess postural properties from accelerometers. *IEEE Trans. Biomed. Eng.* **58**(6), 1752–1761 (2011)
9. Heldman, D.A., Espay, A.J., LeWitt, P.A., Giuffrida, J.P.: Clinician versus machine: reliability and responsiveness of motor endpoints in Parkinson's disease. *Parkinson. Relat. Dis.* **20**(6), 590–595 (2014)
10. Rigas, G., Tzallas, A.T., Tsipouras, M.G., Bougia, P., Tripoliti, E.E., Baga, D., Fotiadis, D.I., Tsouli, S.G., Konitsiotis, S.: Assessment of tremor activity in the parkinsons disease using a set of wearable sensors. *IEEE Trans. Biomed. Eng.* **16**(3), 478–487 (2012)
11. Bevilacqua, V., Trotta, G.F., Loconsole, C., Brunetti, A., Caporusso, N., Bellantuono, G.M., De Feudis, I., Patruno, D., De Marco, D., Venneri, A., Di Vetro, M.G., Losavio, G., Tatò, S.I.: A RGB-D sensor based tool for assessment and rating of movement disorders, vol. 590 (2018)
12. Buongiorno, D., Trotta, G.F., Bortone, I., Di Gioia, N., Avitto, F., Losavio, G., Bevilacqua, V.: Assessment and rating of movement impairment in Parkinson's disease using a low-cost vision-based system. In: Huang D.-S., Gromiha, M.M., Han, K., Hussain, A. (eds.) *Intelligent Computing Methodologies*, pp. 777–788. Springer, Cham (2018)
13. Salarian, A., Russmann, H., Wider, C., Burkhard, P.R., Vingerhoets, F.G., Aminian, K.: Quantification of tremor and bradykinesia in Parkinson's disease using a novel ambulatory monitoring system. *IEEE Trans. Biomed. Eng.* **54**(2), 313–322 (2007)
14. Houde, D., Haijun, L., Lueth, T.C.: Quantitative assessment of Parkinsonian bradykinesia based on an inertial measurement unit. *BioMed. Eng. Online* **14**(1) (2015)
15. Griffiths, R.I., Kotschet, K., Arfon, S., Xu, Z.M., Johnson, W., Drago, J., Evans, A., Kempster, P., Raghav, S., Horne, M.K.: Automated assessment of bradykinesia and dyskinesia in Parkinson's disease. *J. Parkinson's Dis.* **2**(1):47–55 (2012)
16. Keijser, N.L.W., Horstink, M.W.I.M., Gielen, S.C.A.M.: Automatic assessment of levodopa-induced dyskinésies in daily life by neural networks. *Mov. Disord.* **18**(1), 70–80 (2003)
17. Lopane, G., Mellone, S., Chiari, L., Cortelli, P., Calandra-Buonaura, G., Contin, M.: Dyskinesia detection and monitoring by a single sensor in patients with Parkinson's disease. *Mov. Disord.: Off. J. Mov. Disord. Soc.* **30**(9), 1267–1271 (2015)
18. Saunders-Pullman, R., Derby, C., Stanley, K., Floyd, A., Bressman, S., Lipton, R.B., Deligtisch, A., Severt, L., Yu, Q., Kurtis, M., Pullman, S.L.: Validity of spiral analysis in early Parkinson's disease. *Mov. Disord.* **23**(4), 531–537 (2008)
19. Westin, J., Ghiamati, S., Memedi, M., Nyholm, D., Johansson, A., Dougherty, M., Groth, T.: A new computer method for assessing drawing impairment in Parkinson's disease. *J. Neurosci. Methods* **190**(1), 143–148 (2010)
20. Liu, X., Carroll, C.B., Wang, S.Y., Zajicek, J., Bain, P.G.: Quantifying drug-induced dyskinésias in the arms using digitised spiral-drawing tasks. *J. Neurosci. Methods* **144**(1), 47–52 (2005)
21. Loconsole, C., Trotta, G.F., Brunetti, A., Trotta, J., Schiavone, A., Tatò, S.I., Losavio, G., Bevilacqua, V.: Computer vision and EMG-based handwriting analysis for classification in parkinson's disease, vol. 10362. LNCS (2017)
22. Carmeli, E., Patish, H., Coleman, R.: The aging hand. *J. Gerontol. Ser. A: Biol. Sci. Med. Sci.* **58**(2), M146–M152 (2003)

23. Loconsole, C., Cascarano, G.D., Lattarulo, A., Brunetti, A., Trotta, G.E., Buongiorno, D., Bortone, I., De Feudis, I., Losavio, G., Bevilacqua, V., Di Sciascio, E.: A comparison between ANN and SVM classifiers for Parkinson's disease by using a model-free computer-assisted handwriting analysis based on biometric signals. In: 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–8, July 2018
24. Van Gemmert, A.W.A., Teulings, H.-L., Contreras-Vidal, J.L., Stelmach, G.E.: Parkinsons disease and the control of size and speed in handwriting. *Neuropsychologia* **37**(6), 685–694 (1999)
25. Drotar, P., Mekyska, J., Smekal, Z., Rektorova, I., Masarova, L., Faundez-Zanuy, M.: Prediction potential of different handwriting tasks for diagnosis of Parkinson's. In: E-Health and Bioengineering Conference (EHB), pp. 1–4. IEEE (2013)
26. Bevilacqua, V., Salatino, A.A., Di Leo, C., Tattoli, G., Buongiorno, D., Signorile, D., Babiloni, C., Del Percio, C., Triggiani, A.I., Gesualdo, L.: Advanced classification of Alzheimer's disease and healthy subjects based on EEG markers. In: Proceedings of the International Joint Conference on Neural Networks, vol. 2015-Sept (2015)
27. Buongiorno, D., Barsotti, M., Sotgiu, E., Loconsole, C., Solazzi, M., Bevilacqua, V., Frisoli, A.: A neuromusculoskeletal model of the human upper limb for a myoelectric exoskeleton control using a reduced number of muscles. In: 2015 IEEE World Haptics Conference (WHC), pp. 273–279 (2015)
28. Bevilacqua, V., Mastronardi, G., Piscopo, G.: Evolutionary approach to inverse planning in coplanar radiotherapy. *Image Vis. Comput.* **25**(2), 196–203 (2007)
29. Menolascina, F., Bellomo, D., Maiwald, T., Bevilacqua, V., Ciminelli, C., Paradiso, A., Tommasi, S.: Developing optimal input design strategies in cancer systems biology with applications to microfluidic device engineering. *BMC Bioinf.* **10**(12), S4 (2009)
30. Menolascina, F., Tommasi, S., Paradiso, A., Cortellino, M., Bevilacqua, V., Mastronardi, G.: Novel data mining techniques in aCGH based breast cancer subtypes profiling: The biological perspective. In: 2007 IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology, CIBCB 2007, pp. 9–16 (2007)
31. Bevilacqua, V., Tattoli, G., Buongiorno, D., Loconsole, C., Leonardi, D., Barsotti, M., Frisoli, A., Bergamasco, M.: A novel BCI-SSVEP based approach for control of walking in virtual environment using a convolutional neural network. In: Proceedings of the International Joint Conference on Neural Networks (2014)
32. Manghisi, V.M., Uva, A.E., Fiorentino, M., Bevilacqua, V., Trotta, G.F., Monno, G.: Real time RULA assessment using Kinect v2 sensor. *Appl. Ergon.* **65**, 481–491 (2017)
33. Bevilacqua, V., Cariello, L., Columbo, D., Daleno, D., Fabiano, M.D., Giannini, M., Mastronardi, G., Castellano, M.: Retinal fundus biometric analysis for personal identifications. In: Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 5227 LNBI, pp. 1229–1237 (2008)
34. Bevilacqua, V., Buongiorno, D., Carlucci, P., Giglio, E., Tattoli, G., Guarini, A., Sgherza, N., De Tullio, G., Minola, C., Scattone, A., Simone, G., Girardi, F., Zito, A., Gesualdo, L.: A supervised CAD to support telemedicine in hematology. In: 2015 International Joint Conference on Neural Networks (IJCNN), pp. 1–7, July 2015
35. Buongiorno, D., Barone, F., Solazzi, M., Bevilacqua, V., Frisoli, A.: A linear optimization procedure for an EMG-driven neuromusculoskeletal model parameters adjusting: Validation through a myoelectric exoskeleton control. In: Bello, F., Kajimoto, H., Visell, Y. (eds.), *Haptics: Perception, Devices, Control, and Applications*, pp. 218–227. Springer, Cham (2016)
36. Buongiorno, D., Barone, F., Berger, D.J., Cesqui, B., Bevilacqua, V., d'Avella, A., Frisoli, A.: Evaluation of a pose-shared synergy-based isometric model for hand force estimation: Towards myocontrol. In: *Converging Clinical and Engineering Research on Neurorehabilitation II* (pp. 953–958). Springer, Cham (2017)
37. Bevilacqua, V., Tattoli, G., Buongiorno, D., Loconsole, C., Leonardi, D., Barsotti, M., Frisoli, A., Bergamasco, M.: A novel BCI-SSVEP based approach for control of walking in virtual environment using a convolutional neural network. In: 2014 International Joint Conference on Neural Networks (IJCNN), pp. 4121–4128, July 2014

38. Buongiorno, D., Barsotti, M., Barone, F., Bevilacqua, V., Frisoli, Antonio: A linear approach to optimize an emg-driven neuromusculoskeletal model for movement intention detection in myo-control: a case study on shoulder and elbow joints. *Front. Neurorobot.* **12**, 74 (2018)
39. Kanjilal, P.P., Palit, S., Saha, G.: Fetal ECG extraction from single-channel maternal ECG using singular value decomposition. *IEEE Tran. Biomed. Eng.* **44**(1), 51–59 (1997)
40. Bevilacqua, V., Brunetti, A., Triggiani, M., Magaletti, D., Telegrafo, M., Moschetta, M.: An optimized feed-forward artificial neural network topology to support radiologists in breast lesions classification. In: Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion—GECCO ’16 Companion, pp. 1385–1392. ACM, ACM Press, New York, New York, USA (2016)

## Chapter 25

# CNN-Based Prostate Zonal Segmentation on T2-Weighted MR Images: A Cross-Dataset Study



**Leonardo Rundo, Changhee Han, Jin Zhang, Ryuichiro Hataya, Yudai Nagano, Carmelo Militello, Claudio Ferretti, Marco S. Nobile, Andrea Tangherloni, Maria Carla Gilardi, Salvatore Vitabile, Hideki Nakayama and Giancarlo Mauri**

**Abstract** Prostate cancer is the most common cancer among US men. However, prostate imaging is still challenging despite the advances in multi-parametric magnetic resonance imaging (MRI), which provides both morphologic and functional information pertaining to the pathological regions. Along with whole prostate gland segmentation, distinguishing between the central gland (CG) and peripheral zone (PZ) can guide toward differential diagnosis, since the frequency and severity of tumors differ in these regions; however, their boundary is often weak and fuzzy. This work presents a preliminary study on deep learning to automatically delineate the CG and PZ, aiming at evaluating the generalization ability of convolutional neural networks (CNNs) on two multi-centric MRI prostate datasets. Especially, we compared three CNN-based architectures: SegNet, U-Net, and pix2pix. In such a context, the segmentation performances achieved with/without pre-training were compared in 4-fold cross-validation. In general, U-Net outperforms the other methods, especially when training and testing are performed on multiple datasets.

---

L. Rundo (✉) · C. Ferretti · M. S. Nobile · A. Tangherloni · G. Mauri  
Department of Informatics, Systems and Communication,  
University of Milano-Bicocca, Milan, Italy  
e-mail: [leonardo.rundo@disco.unimib.it](mailto:leonardo.rundo@disco.unimib.it)

L. Rundo · C. Militello · M. C. Gilardi  
Institute of Molecular Bioimaging and Physiology (IBFM),  
Italian National Research Council (CNR), Cefalù (PA), Italy

C. Han · J. Zhang · R. Hataya · Y. Nagano · H. Nakayama  
Graduate School of Information Science and Technology,  
The University of Tokyo, Tokyo, Japan

S. Vitabile  
Department of Biopathology and Medical Biotechnologies,  
University of Palermo, Palermo, Italy

## 25.1 Introduction

Prostate cancer (PCa) is expected to be the most common cancer among US men during 2018 [1]. Several imaging modalities can aid PCa diagnosis—such as transrectal ultrasound (TRUS), computed tomography (CT), and magnetic resonance imaging (MRI) [2]—according to the clinical context. Conventional structural T1-weighted (T1w) and T2-weighted (T2w) MRI sequences can play an important role along with functional MRI, such as dynamic contrast enhanced MRI (DCE-MRI), diffusion weighted imaging (DWI), and magnetic resonance spectroscopic imaging (MRSI) [3]. Therefore, MRI conveys more information for PCa diagnosis than CT, revealing the internal prostatic anatomy, prostatic margins, and the extent of prostatic tumors [4].

The manual delineation of both prostate whole gland (WG) and PCa on MR images is a time-consuming and operator-dependent task, which relies on experienced physicians [5]. Besides WG segmentation, distinguishing between the central gland (CG) and peripheral zone (PZ) of the prostate can guide toward differential diagnosis, since the frequency and severity of tumors differ in these regions [6, 7]; the PZ harbors 70–80% of PCa and is a target for prostate biopsy [8]. Regarding this anatomic division of the prostate, the zonal compartment scheme proposed by McNeal is widely accepted [9]. In this context, T2w MRI is the *de facto* standard in the clinical routine of prostate imaging thanks to its high resolution, which allows for differentiating the hyper-intense PZ and hypo-intense CG in young male subjects [10]. However, the conventional clinical protocol for PCa based on prostate-specific antigen (PSA) and systematic biopsy does not generally obtain reliable diagnostic outcomes; thus, the PZ volume ratio (i.e., the PZ volume divided by the WG volume) was recently integrated for PCa diagnostic refinement [11]; the CG volume ratio can be also useful for monitoring prostate hyperplasia [12]. Furthermore, for robust clinical applications, generalization—among different prostate MRI datasets from multiple institutions—is essential.

So, how can we extract the CG and PZ from the WG on different MRI datasets? In this work, we automatically segment the CG and PZ using deep learning to evaluate the generalization ability of convolutional neural networks (CNNs) on two different MRI prostate datasets. However, this is challenging since multi-centric datasets are generally characterized by different contrast, visual consistencies, and image characteristics. Therefore, prostate zones on T2w MR images were manually annotated for supervised learning and then automatically segmented using a mixed scheme by (*i*) training on either each individual dataset or both datasets and (*ii*) testing on both datasets, using CNN-based architectures: SegNet [13], U-Net [14], and pix2pix [15]. In such a context, we compared the segmentation performances achieved with/without pre-training [16].

The manuscript is structured as follows: Sect. 25.2 outlines the state of the art about MRI prostate zonal segmentation methods; Sect. 25.3 describes the MRI datasets as well as the proposed CNN-based segmentation approach; Sect. 25.4 shows our experimental results; finally, some conclusive remarks and possible future developments are given in Sect. 25.5.

## 25.2 Background

In prostate MR image analysis, WG segmentation is essential, especially on T2w MR images [17] or both T2w and the corresponding T1w images [5, 18]. Toward it, literature works used atlas-based methods [19], deformable models, or statistical priors [20]. More recently, deep learning [21] has been successfully applied to this domain, combining deep feature learning with shape models [22] or using CNN-based segmentation approaches [23].

Among the studies on prostate segmentation, the most representative works are outlined hereafter. The authors of [24] used active appearance models combined with multiple level sets for simultaneously segmenting prostatic zones. Qiu et al. [25] proposed a zonal segmentation approach introducing a continuous max-flow model based on a convex-relaxed optimization problem with region consistency constraints. Unlike these methods that analyzed T2w images alone, Makni et al. [26] exploited the evidential C-means algorithm to partition the voxels into their respective zones by integrating the information conveyed by T2w, DWI, and CE T1w MRI sequences.

However, these methods do not evaluate the generalization ability on different MRI datasets from multiple institutions, making their clinical applicability difficult [27]; thus, we verify the cross-dataset generalization using three CNN-based architectures with/without pre-training. Moreover, differently from a recent CNN-based work on DWI data [28], to the best of our knowledge, this is the first CNN-based prostate zonal segmentation approach on T2w MRI alone.

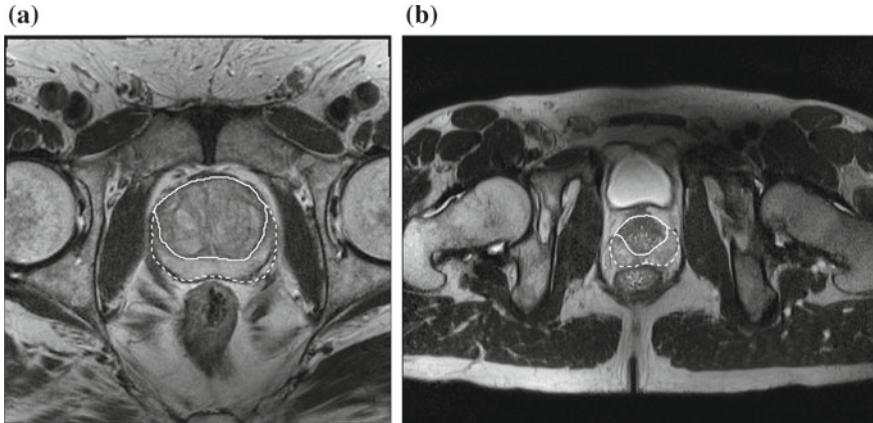
## 25.3 Materials and Methods

For clinical applications with better generalization ability, we evaluate prostate zonal segmentation performances of three CNN-based architectures: SegNet, U-Net, and pix2pix. We also compare the results in 4-fold cross-validation by (*i*) training on either each individual dataset or both datasets and (*ii*) testing on both datasets, with/without pre-training on a relatively large prostate dataset.

### 25.3.1 MRI Datasets

We segment the CG and PZ from the WG using two completely different multi-parametric prostate MRI datasets, namely

#1 dataset containing 21 patients/193 MR slices with prostate, acquired with a whole body Philips Achieva 3T MRI scanner using a phased-array pelvic coil at the Cannizzaro Hospital (Catania, Italy) [5]. The MRI parameters: matrix size =  $288 \times 288$  pixels; slice thickness = 3.0 mm; inter-slice spacing = 4 mm; pixel spacing = 0.625 mm; number of slices = 18;



**Fig. 25.1** Example input prostate T2w MR axial slices in their original image ratio: **a** dataset #1; **b** dataset #2. The CG and PZ are highlighted with solid and dashed white lines, respectively

#2 Initiative for Collaborative Computer Vision Benchmarking (I2CVB) dataset (19 patients/503 MR slices with prostate), acquired with a whole body Siemens TIM 3T MRI scanner using a body coil at the Hospital Center Regional University of Dijon-Bourgogne (France) [3]. The MRI parameters: matrix size  $\in \{308 \times 384, 336 \times 448, 360 \times 448, 368 \times 448\}$  pixels; slice thickness = 1.25 mm; inter-slice spacing = 1.0 mm; pixel spacing  $\in \{0.676, 0.721, 0.881, 0.789\}$  mm; number of slices = 64.

To make the proposed approach clinically feasible [10], we analyzed only T2w images—the most commonly used sequence for prostate zonal segmentation—among available sequences. Figure 25.1 shows two example T2w MR images of the analyzed two datasets. We conducted the following three experiments using a 4-fold cross-validation scheme to confirm the generalization effect under different training/testing conditions in our multi-centric study:

- Individual dataset #1: training on dataset #1 alone, and testing on the whole dataset #2 and the rest of dataset #1 separately for each round;
- Individual dataset #2: training on dataset #2 alone, and testing on the whole dataset #1 and the rest of dataset #2 separately for each round;
- Mixed dataset: training on both datasets #1 and #2, and testing on the rest of datasets #1 and #2 separately for each round.

For 4-fold cross-validation, we partitioned the datasets #1 and #2 using patient indices  $\{[1, \dots, 5], [6, \dots, 10], [11, \dots, 15], [16, \dots, 21]\}$  and  $\{[1, \dots, 5], [6, \dots, 10], [11, \dots, 15], [16, \dots, 19]\}$ , respectively. Finally, the results from the different cross-validation rounds were averaged.

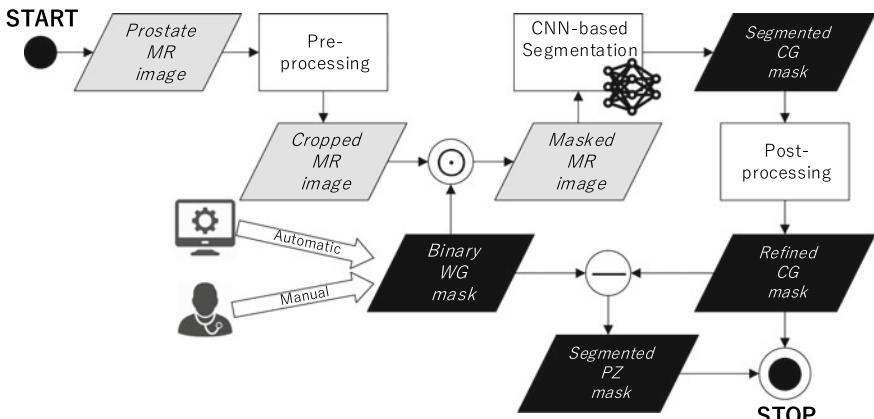
### 25.3.2 CNN-Based Prostate Zonal Segmentation

This work adopts a selective two-step delineation approach to focus on pathological regions in the CG and PZ denoted with  $\mathcal{R}_{CG}$  and  $\mathcal{R}_{PZ}$ , respectively. Relying on [4, 25], the PZ was obtained by subtracting the CG from the WG ( $\mathcal{R}_{WG}$ ) meeting the constraints:  $\mathcal{R}_{WG} = \mathcal{R}_{CG} \cup \mathcal{R}_{PZ}$  and  $\mathcal{R}_{CG} \cap \mathcal{R}_{PZ} = \emptyset$ . The overall prostate zonal segmentation method is outlined in Fig. 25.2.

Starting from the whole prostate MR image, a pre-processing phase, comprising image cropping and resizing to deal with the different characteristics of the two datasets, is performed (see Sect. 25.3.1). Afterward, the resulting image is masked with the binary WG and fed onto the investigated CNN-based models for CG segmentation, which is then refined. Finally, the PZ delineation is obtained by subtracting the CG from the WG according to [25].

#### 25.3.2.1 Pre-processing

To fit the image resolution of the dataset #1, we center-cropped the images of the dataset #2 and resized them to  $288 \times 288$  pixels. Furthermore, the images of these datasets were masked using the corresponding prostate binary masks to omit the background and only focus on extracting the CG and PZ from the WG. This operation can be performed either by an automated method [5, 18] or previously provided manual WG segmentation [3]. For better training, we randomly cropped the input images from  $288 \times 288$  to  $256 \times 256$  pixels and horizontally flipped them.



**Fig. 25.2** Work-flow of our CNN-based prostate zonal segmentation approach. The gray and black data blocks denote grayscale images and binary masks, respectively

### 25.3.2.2 Investigated CNN-Based Architectures

The following architectures were chosen in our comparative analysis since they cover several aspects regarding the CNN-based segmentation: SegNet [13] addresses semantic segmentation (i.e., assigning each pixel of the scene to an object class), U-Net [14] was successfully applied in biomedical image segmentation, while pix2pix [15] exploits an adversarial generative model to perform image-to-image translations.

During the training of all architectures, the  $\mathcal{L}_{\text{DSC}}$  loss function (i.e., a continuous version of the Dice similarity coefficient) was used [23]:

$$\mathcal{L}_{\text{DSC}} = -\frac{2 \sum_{i=1}^N s_i \cdot r_i}{\sum_{i=1}^N s_i + \sum_{i=1}^N r_i}, \quad (25.1)$$

where  $s_i$  and  $r_i$  represent the continuous values of the prediction map (i.e., the result of the final layer of the CNN) and the ground truth at the  $i$ th pixel ( $N$  is the total number of pixels to be classified), respectively.

*SegNet* is a CNN architecture for semantic pixel-wise segmentation [13]. More specifically, it was designed for semantic segmentation of road and traffic scenes, wherein classes represent macro-objects, aiming at smooth segmentation results by preserving boundary information. This non-fully connected architecture, which allows for parameter-efficient implementations suitable for embedded systems, consists of an encoder–decoder network followed by a pixel-wise classification layer. Since our classification task involves only one class, the soft-max operation and rectified linear unit (ReLU) activation function at the final layer were removed for stable training.

We implemented SegNet using PyTorch. During the training phase, we used the Stochastic Gradient Descent (SGD) [29] with a learning rate of 0.01, momentum of 0.9, weight decay of  $5 \times 10^{-4}$ , and batch size of 8. It was trained for 50 epochs, and the learning rate was multiplied by 0.2 at the 20th and 40th epochs.

*U-Net* is a fully CNN capable of stable training with a reduced number of samples [14], combining pooling operators with up-sampling operations. The general architecture is an encoder–decoder with skip connections between mirrored layers in the encoder–decoder stacks. By so doing, high-resolution features from the contracting path are combined with the up-sampled output for better localization. We utilized four scaling operations. U-Net achieved outstanding performance in biomedical benchmark problems [30] and has been also serving as an inspiration for novel deep learning models for image segmentation.

U-Net was implemented using Keras on top of TensorFlow. We used SGD with a learning rate of 0.01, momentum of 0.9, weight decay of  $5 \times 10^{-4}$ , and batch size of 4. Training was executed for 50 epochs, multiplying the learning rate by 0.2 at the 20th, and 40th epochs.

*pix2pix* is an image-to-image translation method coupled with conditional adversarial networks [15]. As a generator, U-Net is used to translate the original image into the segmented one [14], preserving the highest level of abstraction. The generator and discriminator include 8 and 5 scaling operations, respectively.

We implemented *pix2pix* on PyTorch. Kingma et al. [31] was used as an optimizer with a learning rate of 0.0002 and 0.01 for the discriminator and generator, respectively. The learning rate for generator was multiplied by 0.1 every 20 epochs. It was trained for 50 epochs with a batch size of 12.

### 25.3.2.3 Post-processing

Two simple morphological steps were applied on the obtained CG binary masks to smooth boundaries and avoid disconnected regions:

- a hole-filling algorithm on the segmented  $\mathcal{R}_{CG}$  to remove possible holes in the predicted map;
- a small area removal operation to delete connected components with area less than  $\lfloor |\mathcal{R}_{WG}|/8 \rfloor$  pixels, where  $|\mathcal{R}_{WG}|$  denotes the number of the pixels contained in the WG segmentation. This criterion effectively adapts according to the different dimensions of the  $\mathcal{R}_{WG}$ .

### 25.3.2.4 Evaluation

The accuracy of the achieved segmentation results  $\mathcal{S}$  was quantitatively evaluated with respect to the real measurement (i.e., the gold standard  $\mathcal{G}$  obtained manually by experienced radiologists) using the DSC:

$$\text{DSC} = \frac{2|\mathcal{S} \cap \mathcal{G}|}{|\mathcal{S}| + |\mathcal{G}|} \times 100 (\%). \quad (25.2)$$

### 25.3.3 Influence of Pre-training

In medical imaging, due to the lack of training data, ensuring CNN's proper training convergence is difficult from scratch. Therefore, pre-training models on a different application and then fine-tuning is common [16].

To evaluate cross-dataset generalization abilities *via* pre-training, we compared the performances of the three CNN-based architectures with/without pre-training on a similar application. We used a relatively large dataset of 50 manually segmented examples from the Prostate MR Image Segmentation 2012 (PROMISE12) challenge [32]. Since this competition focuses only on WG segmentation without providing prostate zonal labeling, we pre-trained the architectures on WG segmentation. To

adjust this dataset to our experimental setup, the images of this dataset were resized from  $512 \times 512$  to  $288 \times 288$  pixels and randomly cropped to  $256 \times 256$  pixels; because our task only focuses on slices with prostate, we also omitted initial/final slices without prostate, so the number of slices for each sample was fixed to 25.

## 25.4 Results

This section explains how the three CNN-based architectures segmented the prostate zones, evaluating their cross-dataset generalization ability.

Table 25.1 shows the 4-fold cross-validation results obtained in the different experimental conditions. When training and testing are both performed on the dataset #1, U-Net outperforms the other architectures on both CG and PZ segmentation; however, it experiences problems with testing on the dataset #2 due to the limited number of training images in the dataset #1. In such a case, pix2pix generalizes better thanks to its internal generative model. When trained on the dataset #2 alone, U-Net yields the most accurate results both in intra- and cross-dataset testing. This probably derives from the dataset #2's relatively larger training data as well as U-Net's good generalization ability when sufficient data are available. Moreover, SegNet reveals rather unstable results, especially when trained on a limited amount of data.

Finally, when trained on the mixed dataset, all three architectures—especially U-Net—achieve good results on both datasets without losing accuracy compared to training on the same dataset alone. Therefore, using mixed MRI datasets during training can considerably improve the performance in cross-dataset generalization toward other clinical applications. Comparing the CG and PZ segmentation, when tested on the dataset #1, the results on the PZ are generally more accurate, except when trained on the dataset #2 alone; however, for the dataset #2, segmentations on the CG are generally more accurate.

Fine-tuning after pre-training sometimes leads to slightly better results than training from scratch, when trained only on a single dataset. However, its influence is generally negligible or rather negative, when trained on the mixed dataset. This modest impact is probably due to the ineffective data size for pre-training.

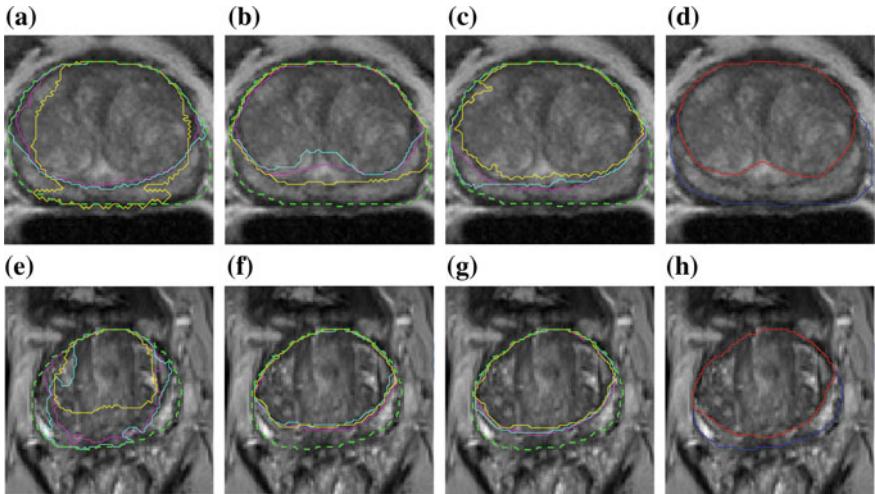
For a visual assessment, two examples (one for each dataset) are shown in Fig. 25.3. Relying on the gold standards in Fig. 25.3d, h, it can be seen that U-Net generally achieves more accurate results compared to SegNet and pix2pix. This finding confirms the trend revealed by the DSC values in Table 25.1.

## 25.5 Discussion and Conclusions

Our preliminary results show that CNN-based architectures can segment prostate zones on two different MRI datasets to some extent, leading to valuable clinical insights; CNNs suffer when training and testing are performed on different MRI

**Table 25.1** Prostate zonal segmentation results of the three CNN-based architectures in 4-fold cross-validation assessed by the DSC (presented as the average and standard deviation). The experimental results are calculated on the different setups of (*i*) training on either each individual dataset or both datasets and (*ii*) testing on both datasets. Numbers in bold indicate the highest DSC values for each prostate region (i.e., CG and PZ) among all architectures with/without pre-training (PT)

	Architecture	Zone	Testing on Dataset #1		Testing on Dataset #2	
			Average	Std. dev.	Average	Std. dev.
Training on Dataset #1	SegNet (w/o PT)	CG	80.20	3.28	74.48	5.82
		PZ	80.66	11.51	59.57	12.68
	SegNet (w/ PT)	CG	83.38	3.22	72.75	2.80
		PZ	87.39	3.90	66.20	5.64
	U-Net (w/o PT)	CG	84.33	2.37	74.18	3.77
		PZ	88.98	2.98	66.63	1.93
	U-Net (w/ PT)	CG	<b>86.88</b>	1.60	70.11	5.31
		PZ	<b>90.38</b>	3.38	58.89	7.06
Training on Dataset #2	pix2pix (w/o PT)	CG	82.35	2.09	<b>76.61</b>	2.17
		PZ	87.09	2.72	73.20	2.62
	pix2pix (w/ PT)	CG	80.38	2.81	76.19	5.77
		PZ	83.53	5.65	<b>73.73</b>	2.40
	SegNet (w/o PT)	CG	76.04	2.05	87.07	2.41
		PZ	77.25	3.09	82.45	1.77
	SegNet (w/ PT)	CG	77.99	2.15	87.75	2.83
		PZ	76.51	2.70	82.26	2.09
Training on mixed dataset	U-Net (w/o PT)	CG	78.88	0.88	88.21	2.10
		PZ	74.52	1.85	<b>83.03</b>	2.46
	U-Net (w/ PT)	CG	<b>79.82</b>	1.11	<b>88.66</b>	2.28
		PZ	<b>74.56</b>	5.12	82.48	2.47
	pix2pix (w/o PT)	CG	77.90	0.73	86.95	2.93
		PZ	66.09	3.07	81.33	0.90
	pix2pix (w/ PT)	CG	77.21	1.02	85.94	4.31
		PZ	67.39	5.04	80.07	0.84



**Fig. 25.3** Examples of prostate zonal segmentation in pre-training/fine-tuning. The first row concerns testing on dataset #1, trained on: **a** dataset #1; **b** dataset #2; **c** mixed dataset. The second row concerns testing on dataset #2, trained on: **e** dataset #1; **f** dataset #2; **g** mixed dataset. The  $\mathcal{R}_{CG}$  segmentation results are represented with magenta, cyan, and yellow solid contours for SegNet, U-Net, and pix2pix, respectively. The dashed green line denotes the  $\mathcal{R}_{WG}$  boundary. The last column (sub-figures **d** and **h**) shows the gold standard for  $\mathcal{R}_{CG}$  and  $\mathcal{R}_{PZ}$  with red and blue lines, respectively. The images are zoomed with a  $4 \times$  factor

datasets acquired by different devices and protocols, but this can be mitigated by training the CNNs on multiple datasets, even without pre-training. Generally, considering different experimental training and testing conditions, U-Net outperforms SegNet and pix2pix thanks to its good generalization ability. Furthermore, this study suggests that significant performance improvement *via* fine-tuning may require a remarkably large dataset for pre-training.

As future developments, we plan to improve the results by refining the predicted binary masks for better smoothness and continuity, avoiding disconnected segments; furthermore, we should enhance the output delineations considering the three-dimensional spatial information among slices. Furthermore, relying on the encouraging cross-dataset capability of U-Net, it is worth to devise and test new solutions aiming at improving the performance of the standard U-Net architecture [30]. Finally, for better cross-dataset generalization, additional prostate zonal datasets and domain adaptation using transfer learning with generative adversarial networks (GANs) [33, 34] and variational auto-encoders (VAEs) [35] could be useful.

**Acknowledgements** This work was partially supported by the Graduate Program for Social ICT Global Creative Leaders of The University of Tokyo by JSPS. We thank the Cannizzaro Hospital, Catania, Italy, for providing one of the imaging datasets analyzed in this study.

## References

1. Siegel, R.L., Miller, K.D., Jemal, A.: Cancer statistics, 2018. *CA Cancer J. Clin.* **68**(1), 7–30 (2018)
2. Rundo, L., Tangherloni, A., Nobile, M.S., Militello, C., Besozzi, D., Mauri, G., Cazzaniga, P.: MedGA: a novel evolutionary method for image enhancement in medical imaging systems. *Expert Syst. Appl.* **119**, 387–399 (2019)
3. Lemaitre, G., Martí, R., Freixenet, J., Vilanova, J.C., Walker, P.M., Meriaudeau, F.: Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: a review. *Comput. Biol. Med.* **60**, 8–31 (2015)
4. Villeirs, G.M., De Meerleer, G.O.: Magnetic resonance imaging (MRI) anatomy of the prostate and application of MRI in radiotherapy planning. *Eur. J. Radiol.* **63**(3), 361–368 (2007)
5. Rundo, L., Militello, C., Russo, G., Garufi, A., Vitabile, S., Gilardi, M.C., Mauri, G.: Automated prostate gland segmentation based on an unsupervised fuzzy c-means clustering technique using multispectral T1w and T2w MR imaging. *Information* **8**(2), 49 (2017)
6. Choi, Y.J., Kim, J.K., Kim, N., Kim, K.W., Choi, E.K., Cho, K.S.: Functional MR imaging of prostate cancer. *Radiographics* **27**(1), 63–75 (2007)
7. Niaf, E., Rouvière, O., Mège-Lechevallier, F., Bratan, F., Lartizien, C.: Computer-aided diagnosis of prostate cancer in the peripheral zone using multiparametric MRI. *Phys. Med. Biol.* **57**(12), 3833 (2012)
8. Haffner, J., Potiron, E., Bouyé, S., Puech, P., Leroy, X., Lemaitre, L., Villers, A.: Peripheral zone prostate cancers: location and intraprostatic patterns of spread at histopathology. *Prostate* **69**(3), 276–282 (2009)
9. Selman, S.H.: The McNeal prostate: a review. *Urology* **78**(6), 1224–1228 (2011)
10. Hoeks, C.M., Barentsz, J.O., Hambrock, T., Yakar, D., Somford, D.M., Heijmink, S.W., et al.: Prostate cancer: multiparametric MR imaging for detection, localization, and staging. *Radiology* **261**(1), 46–66 (2011)
11. Chang, Y., Chen, R., Yang, Q., Gao, X., Xu, C., Lu, J., Sun, Y.: Peripheral zone volume ratio (PZ-ratio) is relevant with biopsy results and can increase the accuracy of current diagnostic modality. *Oncotarget* **8**(21), 34836 (2017)
12. Kirby, R., Gilling, R.: *Fast Facts: Benign Prostatic Hyperplasia*, 7th edn. Health Press Limited, Abingdon, UK (2011)
13. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**(12), 2481–2495 (2017)
14. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, vol. 9351 of *LNCS*, pp. 234–241. Springer (2015)
15. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv:1611.07004* (2016)
16. Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J.: Convolutional neural networks for medical image analysis: full training or fine tuning? *IEEE Trans. Med. Imaging* **35**(5), 1299–1312 (2016)
17. Ghose, S., Oliver, A., Martí, R., Lladó, X., Vilanova, J.C., Freixenet, J., et al.: A survey of prostate segmentation methodologies in ultrasound, magnetic resonance and computed tomography images. *Comput. Methods Prog. Biomed.* **108**(1), 262–287 (2012)
18. Rundo, L., Militello, C., Russo, G., D’Urso, D., Valastro, L.M., Garufi, A., et al.: Fully automatic multispectral MR image segmentation of prostate gland based on the fuzzy c-means clustering algorithm. In: *Multidisciplinary Approaches to Neural Computing*, vol. 69 of *Smart Innovation, Systems and Technologies*, pp. 23–37. Springer (2018)
19. Klein, S., Van Der Heide, U.A., Lips, I.M., Van Vulpen, M., Staring, M., Pluim, J.P.: Automatic segmentation of the prostate in 3D MR images by atlas matching using localized mutual information. *Med. Phys.* **35**(4), 1407–1417 (2008)

20. Martin, S., Troccaz, J., Daanen, V.: Automated segmentation of the prostate in 3D MR images using a probabilistic atlas and a spatially constrained deformable model. *Med. Phys.* **37**(4), 1579–1590 (2010)
21. Bevilacqua, V., Brunetti, A., Guerriero, A., Trotta, G.F., Telegrafo, M., Moschetta, M.: A performance comparison between shallow and deeper neural networks supervised classification of tomosynthesis breast lesions images. *Cogn. Syst. Res.* **53**, 3–19 (2019)
22. Guo, Y., Gao, Y., Shen, D.: Deformable MR prostate segmentation via deep feature learning and sparse patch matching. *IEEE Trans. Med. Imaging* **35**(4), 1077–1089 (2016)
23. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: Proceedings of the 4th International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)
24. Toth, R., Ribault, J., Gentile, J., Sperling, D., Madabhushi, A.: Simultaneous segmentation of prostatic zones using active appearance models with multiple coupled levelsets. *Comput. Vis. Image Underst.* **117**(9), 1051–1060 (2013)
25. Qiu, W., Yuan, J., Ukwatta, E., Sun, Y., Rajchl, M., Fenster, A.: Dual optimization based prostate zonal segmentation in 3D MR images. *Med. Image Anal.* **18**(4), 660–673 (2014)
26. Makni, N., Iancu, A., Colot, O., Puech, P., Mordon, S., Betrouni, N.: Zonal segmentation of prostate using multispectral magnetic resonance images. *Med. Phys.* **38**(11), 6093–6105 (2011)
27. AlBadawy, E.A., Saha, A., Mazurowski, M.A.: Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing. *Med. Phys.* (2018)
28. Clark, T., Zhang, J., Baig, S., Wong, A., Haider, M.A., Khalvati, F.: Fully automated segmentation of prostate whole gland and transition zone in diffusion-weighted MRI using convolutional neural networks. *J. Med. Imaging* **4**(4), 041307 (2017)
29. Bottou, L.: Large-scale machine learning with stochastic gradient descent. In: Proceedings of COMPSTAT'2010, pp. 177–186. Springer (2010)
30. Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., et al.: U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**(1), 67 (2019)
31. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
32. Litjens, G., Toth, R., van de Ven, W., Hoeks, C., Kerkstra, S., van Ginneken, B., et al.: Evaluation of prostate segmentation algorithms for MRI: the PROMISE12 challenge. *Med. Image Anal.* **18**(2), 359–373 (2014)
33. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al.: Generative adversarial nets. In: Proceedings of Advances in Neural Information Processing Systems (NIPS), pp. 2672–2680 (2014)
34. Han, C., Hayashi, H., Rundo, L., Araki, R., Shimoda, W., Muramatsu, S., et al.: GAN-based synthetic brain MR image generation. In: Proceedings of International Symposium on Biomedical Imaging (ISBI), pp. 734–738. IEEE (2018)
35. Kingma, D., Welling, M.: Auto-encoding variational Bayes. In: Proc. International Conference on Learning Representations (ICLR). arXiv preprint [arXiv:1312.6114](https://arxiv.org/abs/1312.6114) (2014)

# Chapter 26

## Understanding Cancer Phenomenon at Gene Expression Level by using a Shallow Neural Network Chain



Pietro Barbiero, Andrea Bertotti, Gabriele Ciravegna, Giansalvo Cirrincione,  
Elio Piccolo and Alberto Tonda

**Abstract** Exploiting the availability of the largest collection of patient-derived xenografts from metastatic colorectal cancer annotated for a response to therapies, this manuscript aims to characterize the biological phenomenon from a mathematical point of view. In particular, we design an experiment in order to investigate how genes interact with each other. By using a shallow neural network model, we find reduced feature subspaces where the resistance phenomenon may be much easier to understand and analyze.

---

P. Barbiero (✉)

Department of Mathematical Sciences, Politecnico di Torino, Turin, Italy  
e-mail: [pietro.barbiero@studenti.polito.it](mailto:pietro.barbiero@studenti.polito.it)

A. Bertotti

Dipartimento di Oncologia, Candiolo Cancer Institute - FPO, Università degli studi di Torino,  
Torino, Italy  
e-mail: [andrea.bertotti@ircc.it](mailto:andrea.bertotti@ircc.it)

G. Ciravegna

Università degli Studi di Siena, DIISM, Siena, Italy  
e-mail: [elio.piccolo@polito.it](mailto:elio.piccolo@polito.it)

G. Cirrincione

Lab. LTI, University of Picardie Jules Verne, Amiens, France  
e-mail: [exin@u-picardie.fr](mailto:exin@u-picardie.fr)

G. Cirrincione

University of South Pacific, Suva, Fiji

E. Piccolo

Politecnico di Torino, DAUIN, Turin, Italy  
e-mail: [elio.piccolo@polito.it](mailto:elio.piccolo@polito.it)

A. Tonda

UMR 782, Université Paris-Saclay, INRA, Thiverval-Grignon, France  
e-mail: [alberto.tonda@inra.fr](mailto:alberto.tonda@inra.fr)

## 26.1 Biological Introduction

The cancer phenomenon seems to be the result of a different sequence of genetic alterations. In this difficult setting, clinical treatments add an external complexity to the tumor behavior. In recent years, patient-derived xenografts (PDXs) have emerged as powerful tools for biomarker discovery and drug development in oncology [1–3]. PDXs are obtained by propagating surgically derived tumor specimens in immunocompromised mice. Through this procedure, cancer cells remain viable ex-vivo and retain the typical characteristics of different tumors from different patients. Hence, they can effectively recapitulate the intra- and inter-tumor heterogeneity that is found in real patients. Based on this idea, the PDX technology has been leveraged to conduct large-scale preclinical analyses to identify reliable correlations between genetic or functional traits and sensitivity to anticancer drugs. In this context, during the last decade we have been assembling the largest collection of PDXs from metastatic colorectal cancer (mCRC) available worldwide in an academic environment. Such resource has been widely characterized at the molecular level and has been annotated for a response to therapies, including cetuximab, an anti-EGFR antibody approved for clinical use [4–6]. Such multilayered information has been already leveraged to reliably anticipate clinical findings [7] with major therapeutic implications. Here, we propose to exploit and combine available transcriptional data obtained from mCRC PDXs of our collection through the Illumina bead array technology [8].

## 26.2 Data Set

The data consists of a DNA microarray, with the expression of 20,023 genes in 403 CRC murine tissues. Each cancerous tissue is associated with a Boolean variable describing the tumor response to cetuximab (responsive or not responsive to treatments), as described in previous works [9]. For the purpose of the analyses, genes are considered as features, while each patient corresponds to a sample; the response to the treatment identifies two classes. The data set is normalized using z-score normalization on features and on samples.

## 26.3 Objective

Previous works concerning patient classification on this data set have shown remarkable results, reaching high accuracy. In [10], the authors have dealt with the high-dimensional space of genes by using dimensionality reduction algorithms and manifold analyses. In [11], they have shown how several machine learning classifiers are able to assess patient response to drugs reaching similar performances in cross-validated test sets. In this work, going beyond the previous analyses, we try to

investigate the structure of the biological phenomenon at the gene expression level from a mathematical point of view. In particular, we are interested in understanding how the cancer resistance to treatments can be explained by using only the information contained in a DNA microarray. At the same time, we take advantage of the availability of real-world data to make some considerations about the behavior of shallow neural networks dealing with high-dimensional spaces.

## 26.4 Shallow Neural Networks

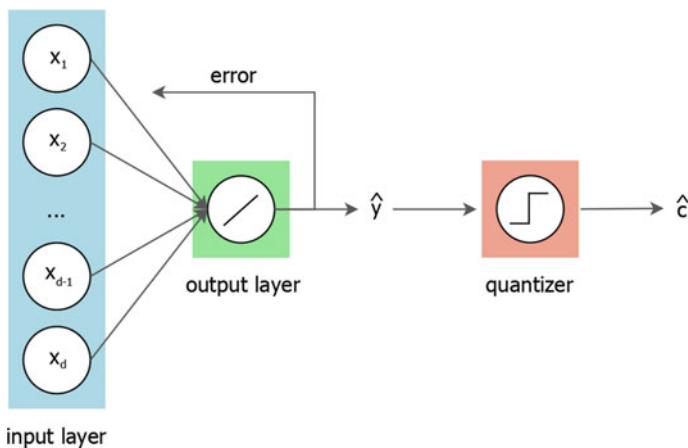
### 26.4.1 Mathematical Model of the Network Architecture

The neural network architecture we used in the following experiments is known as *ADALINE* [12–14]. It has 20,023 inputs corresponding to the input features (genes) and one output neuron equipped with a linear output function. The network does not have hidden layers. During forward propagation, the network computes the dot product between the weight vector  $w$  and the  $i$ th sample  $x^{(i)}$  plus the bias  $b$ . This corresponds to a weighted sum of the inputs with bias correction (as in a linear regression model):

$$z^{(i)} = w^T x^{(i)} + b \quad (26.1)$$

$$\hat{y}^{(i)} = f(z^{(i)}) = z^{(i)} \quad (26.2)$$

where  $w$  is the weight vector,  $b$  the bias,  $f$  the activation function and  $\hat{y}^{(i)}$  the network output (Fig. 26.1).



**Fig. 26.1** Shallow neural network architecture

### 26.4.2 Objective Function

The squared error function evaluates the performance of the algorithm on an individual sample:

$$\mathcal{L}(\hat{y}^{(i)}, y^{(i)}) = (y^{(i)} - \hat{y}^{(i)})^2 \quad (26.3)$$

where  $y^{(i)}$  is 1 if the  $i$ th sample belongs to class 1 and 0 if it belongs to class 0. In order to evaluate the global performance of the classifier, we use a cost function with L2 regularization of the weights [15, 16]. Appending a term to the cost function that penalizes large weights leads to a reduction of the search space, shrinking useless weights toward zero, thus providing simpler models:

$$J(w, b) = \frac{1}{m} \sum_{i=1}^m \mathcal{L}(y^{(i)}, \hat{y}^{(i)}) + \frac{\lambda}{2m} \|w\|_w^2 \quad (26.4)$$

where  $\lambda$  is the regularization parameter and  $\|w\|_w^2$  is the L2 norm of the weight vector. For big values of  $\lambda$ , the regularization is stronger, increasing the penalization related to weights. As a result, the weights which are not useful for the purpose of minimizing the MSE (i.e., the first part of the objective function) are shrunk toward zero. On the contrary, for low values of  $\lambda$ , the regularization effect is weaker.<sup>1</sup> In order to provide a quantitative measure of the network performance, we transform the regression outcomes into class labels by using a Heaviside step function:

$$\hat{c}^{(i)} = \frac{d}{d\hat{y}} \max\{0, \hat{y}^{(i)}\} \quad (26.5)$$

and we compute the accuracy as if it were a classification task.

### 26.4.3 Parameter Optimization

Since the cost function measures the errors in the current predictions, the problem of the learning process is equivalent to the minimization of the cost function. Whereas the training samples are fixed, the cost function depends only on the network parameters (weights and bias). So, the cost function minimization is equivalent to the optimization of the network parameters. For the following analyses, we use the adaptive moment estimation optimizer (Adam). Adam is an algorithm for first-order gradient-based optimization of stochastic objective functions, based on adaptive estimates of lower-order moments [17]. It is a variant of the classical gradient descent algorithm, designed to combine the advantages of two popular methods: Adagrad

---

<sup>1</sup>The described shallow neural network model is equivalent to a linear regression model with an L2 regularization of the parameters also known as ridge regression [16].

and RMSProp. According to [17], Adam's advantages are that its step sizes are approximately bounded by the learning rate, it does not require a stationary objective, it works with sparse gradients, and it naturally performs a form of step size annealing. In the context of feed-forward neural networks, the objective function to be minimized is the cost function  $J_t(\theta)$ , where  $t$  denotes the  $t$ th epoch and  $\theta$  is a label for  $w$  and  $b$ . The authors identify with  $g_t$  the gradient, i.e., the vector of partial derivatives of  $J_t$ , w.r.t  $w$  and  $b$  evaluated at epoch  $t$  (26.6). This estimate is then used to update two exponential moving averages of the gradient ( $m_t$ , (26.7)) and the squared gradient ( $v_t$ , (26.8)). The two hyper-parameters  $\beta_1, \beta_2 \in [0, 1]$  control the exponential decay rates of these moving averages. High values for  $\beta_1, \beta_2$  reduce the time window size of the moving averages, resulting in low inertial effects and greater oscillations. On the contrary, low values of  $\beta_1, \beta_2$  increase the time window size, providing a stronger smoothing effect. The first moving average  $m_t$  is an estimate of the first-order moment (the mean) of the gradient. The second one instead is an estimate of the second-order moment (the uncentered variance) of the gradient. Since these moving averages are initialized as vectors of zeros, the moment estimates are biased toward zero during the initial time steps (especially when the decay rates are small, i.e., the  $\beta$ s are close to 1). This issue can be alleviated by the bias correction shown in (26.9) and (26.10). The ratio of the two moving averages corresponds to a standardization of the first-order moment of the gradient. The network parameters are finally updated by using the classical formula of gradient descent in (26.11). The term  $\epsilon$  (typically  $10^{-8}$ ) ensures that the denominator is always nonzero, avoiding numerical issues.

$$g_t = \nabla_{\theta} J_t(\theta_{t-1}) \quad (26.6)$$

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \quad (26.7)$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t \odot g_t \quad (26.8)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (26.9)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (26.10)$$

$$\theta = \theta - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (26.11)$$

The initial conditions are:  $m_0 = 0$ ,  $v_0 = 0$  and  $t = 0$ . Typical values for  $\beta$ s are  $\beta_1 \approx 0.9$  and  $\beta_2 = 0.999$ . Overall, Adam is a very efficient algorithm, requiring very few computations and memory space, which is crucial in our case, given the size of the data set.

## 26.5 Experiments and Discussion

In order to assess the goodness of the proposed approach, we perform a series of cross-validated training of the neural model previously described. In particular, at each iteration the neural network is trained 30 times, each of which using a tenfold cross-validation with random folds. The neural network hyper-parameters are heuristically fixed to:

$$\lambda = \frac{1}{\#samples} \quad (26.12)$$

$$\alpha = \frac{1}{\lambda + L + 1} \quad (26.13)$$

where  $L$  is the maximum sum of the squares over all samples [18, 19]. Since the objective function to minimize contains the L2 norm of the weights (see Eq. 26.4), weights which are not fundamental for the classification task are shrunk toward zero by the optimizer [16]. Figures 26.2 and 26.3 show, respectively, the histogram and the notched box plot of the weights after the training process in the first iteration. It is important to notice that most of the weights are set to zero or are very close to zero. This means that their contribution to the weighted sum in Eq. (26.1) is almost negligible. Exploiting this result, for each fold we take note of the input features (i.e., the genes) which correspond to weights having an absolute value  $w_j$  after a training process:

$$|w_j| > 2\sigma_w \quad (26.14)$$

where  $\sigma_w$  is the variance of the weight distribution (see Fig. 26.3). At the end of the 30 iterations, we found that some input features are chosen more frequently than others.

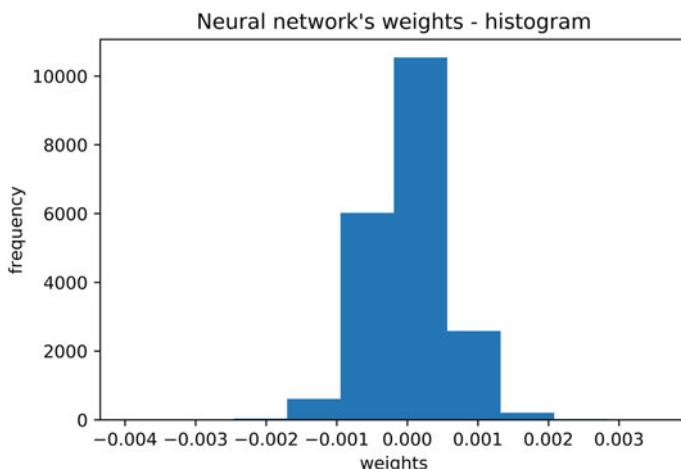
Biologically speaking, this result suggests that the information contained in the DNA microarray related to these genes may be relevant in understanding the cancer resistance to drugs. In order to investigate more deeply the biological phenomenon, we repeat the same experiments, modifying the database by keeping only the most frequently selected features, i.e., those which are selected at least half of the times after 300 training processes. This technique corresponds to a supervised feature selection performed using a chain of shallow neural networks. Figure 26.4 shows for each iteration the number of features used to train the neural network and the corresponding tenfold cross-validation accuracy. The blue bars correspond to the feature selection technique described above, while the violet ones to the ANOVA-F statistic method.<sup>2</sup> Notice that, initially, by using the original data set, the cross-validation accuracy is around 0.7. Such result may have two main explanations. First, the classes are not perfectly balanced, since 66% of samples belong to class

---

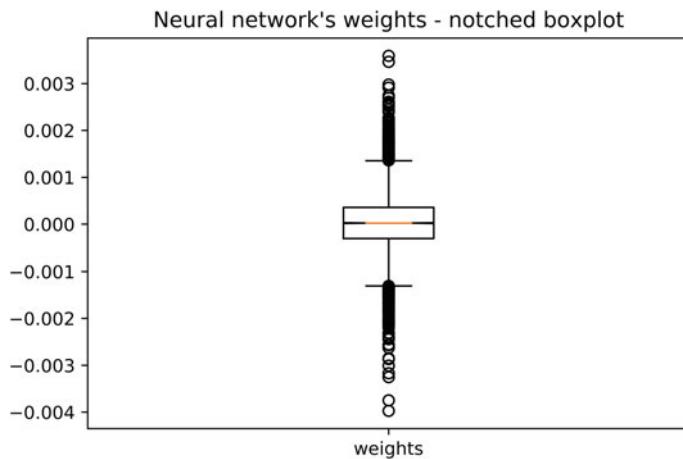
<sup>2</sup>The corresponding standard deviation is always in the order of few percentage decimals, and it is not directly displayed since it is not relevant to the purpose of the discussion. However, you can reproduce the experiment by using our code if you need more precision.

0. Secondly, the high dimensionality of the data may generate a slight overfitting. Finally, the high dimensionality of the search space of the network weights and biases might increase the number of the local optima and lead Adam to sub-optimal solutions. However, by reducing the input features using the method previously, the cross-validation accuracy raises above 0.9, decreasing progressively as the number of features is further diminished.

It is important to notice that the neural network used as classifier is linear; i.e., geometrically speaking, it delimits the input space with an hyperplane in order to classify data. This means that the proposed approach provides better results if the underlying phenomenon represented by the input data set is also linear. The results in Fig. 26.4 show how the shallow neural network classifier delivers better results in the 737-dimensional space identified by the proposed feature extraction technique, than in the original 20,023-dimensional data set. This may suggest that the underlying biological phenomenon at the DNA microarray level is more linear in the reduced space than in the original one. Practically speaking, a linear problem is much easier to understand and tackle because the superposition principle holds; i.e., the net response caused by two or more stimuli is the sum of the responses that would have been caused by each stimulus individually. Therefore, from a biological point of view, these results may suggest that the above experiment generates subspaces of the input features where the cancer resistance to treatments can be studied more easily. In particular, in the 737-dimensional space the biological phenomenon is easier in the sense that it is more linear than in the original space, while in the 90- or 20-dimensional spaces it is easier because, while the classification accuracy decreases, the limited number of genes involved can be more thoroughly analyzed by human experts.

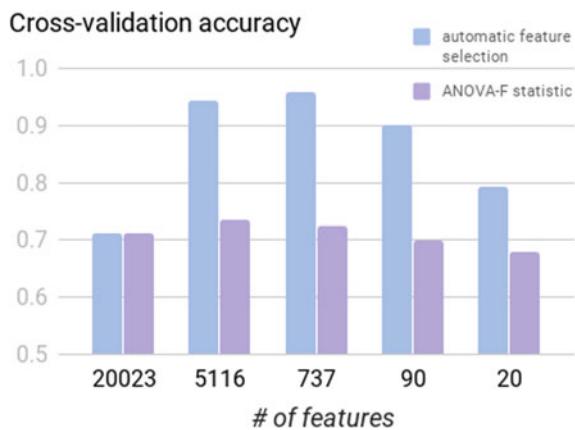


**Fig. 26.2** An example of histogram of the neural network weights after the first training iteration



**Fig. 26.3** An example of notched box plot of the neural network weights after the first training iteration

**Fig. 26.4** Histogram displaying the tenfold cross-validation accuracy at each iteration of the experiment



## 26.6 Conclusion

In this manuscript, we tried to investigate the biological phenomenon (i.e., cancer resistance to treatments) by using gene expression information through a linear mathematical model for classification. Taking advantage of the weight penalization, the proposed shallow neural network has found a way to tackle the curse of dimensionality by reducing its optimization space. We exploited this property in order to design an experiment to understand more deeply how genes interact with each other. The results are critically interpreted in order to retrieve insights about the underlying phenomena. The linear nature of the classifier allowed us to make some considerations concerning the kind of interactions among genes. In particular, we identified sub-

spaces of the data set where the biological phenomenon is much easier to understand and analyze, because the superposition principle holds or the number of features is considerably restricted.

## References

1. Hidalgo, M., et al.: Patient-derived xenograft models: an emerging platform for translational cancer research. *Cancer Discov.* **4**, 998–1013 (2014)
2. Tentler, J.J., et al.: Patient-derived tumour xenografts as models for oncology drug development. *Nat. Rev. Clin. Oncol.* **9**, 338–350 (2012)
3. Byrne A.T., et al.: Interrogating open issues in cancer precision medicine with patient derived xenografts. In: *Nat. Rev. Cancer* (2017). <https://doi.org/10.1038/nrc.2016.140>
4. Bertotti, A., et al.: A molecularly annotated platform of patient-derived xenografts ('xenopatients') identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer. *Cancer Discov.* **1**, 508–523 (2011)
5. Zanella, E.R., et al.: IGF2 is an actionable target that identifies a distinct subpopulation of colorectal cancer patients with marginal response to anti-EGFR therapies. *Sci. Transl. Med.* (2015). <https://doi.org/10.1126/scitranslmed.3010445>
6. Bertotti, A., et al.: The genomic landscape of response to EGFR blockade in colorectal cancer. *Nature* **526**, 263–267 (2015)
7. Sartore Bianchi, A., et al.: Dual-targeted therapy with trastuzumab and lapatinib in treatment-refractory, KRAS codon 12/13 wild-type, HER2-positive metastatic colorectal cancer (HERA-CLES): a proof-of-concept, multicentre, open-label, phase 2 trial. *Lancet Oncol.* **17**, 738–746 (2016)
8. Illumina: Array-based gene expression analysis. Data Sheet Gene Expr (2011). [http://res.illumina.com/documents/products/datasheets/datasheet\\_gene\\_exp\\_analysis.pdf](http://res.illumina.com/documents/products/datasheets/datasheet_gene_exp_analysis.pdf)
9. Isella, C., et al.: Selective analysis of cancer-cell intrinsic transcriptional traits defines novel clinically relevant subtypes of colorectal cancer. *Nat. Gen.* **8**, (2017). <https://doi.org/10.1038/ncomms15107>
10. Barbiero P., Bertotti A., Ciravegna G., Cirrincione G., Pasero E., Piccolo E. : Supervised gene identification in colorectal cancer. In: Quantifying and Processing Biomedical and Behavioral Signals. Springer (2018). ISBN 9783319950945. [https://doi.org/10.1007/978-3-319-95095-2\\_21](https://doi.org/10.1007/978-3-319-95095-2_21)
11. Barbiero, P., Bertotti, A., Ciravegna, G., Cirrincione, G., Piccolo, E.: DNA microarray classification: evolutionary optimization of neural network hyperparameters. In: Italian Workshop on Neural Networks (WIRN 2018). Vietri Sul Mare, Italy, June 2018
12. Widrow, B., Lehr, M.A.: Artificial neural networks of the perceptron, madaline, and backpropagation family. *Neurobionics* (1993). <https://doi.org/10.1016/B978-0-444-89958-3.50013-9>
13. Haykin, S.: Neural Networks: A Comprehensive Foundation. Prentice Hall (1998). ISBN 0132733501
14. Chollet, F., et al.: Keras (2015). <https://keras.io>
15. Ng, A.Y.: Feature selection, L1 versus L2 regularization, and rotational in-variance. In: International Conference on Machine Learning (2004). <https://doi.org/10.1145/1015330.1015435>
16. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Series in Statistics (2009). ISBN 0387848576

17. Kingma D.P., Ba, J.: Adam: a method for stochastic optimization. In: International Conference for Learning Representations (2017). [arXiv:1412.6980v9](https://arxiv.org/abs/1412.6980v9)
18. Schmidt, M., Le Roux, N., Bach, F.: Minimizing finite sums with the stochastic average gradient. *Math. Prog.* (2013). <https://doi.org/10.1007/s10107-016-1030-6>
19. Defazio, A., Bach, F., Lacoste-Julien, S.: SAGA: a fast incremental gradient method with support for non-strongly convex composite objectives In: Advances in Neural Information Processing Systems (2014). [arXiv:1407.0202](https://arxiv.org/abs/1407.0202)

# Chapter 27

## Infinite Brain MR Images: PGGAN-Based Data Augmentation for Tumor Detection



**Changhee Han, Leonardo Rundo, Ryosuke Araki, Yujiro Furukawa,  
Giancarlo Mauri, Hideki Nakayama and Hideaki Hayashi**

**Abstract** Due to the lack of available annotated medical images, accurate computer-assisted diagnosis requires intensive data augmentation (DA) techniques, such as geometric/intensity transformations of original images; however, those transformed images intrinsically have a similar distribution to the original ones, leading to limited performance improvement. To fill the data lack in the real image distribution, we synthesize brain contrast-enhanced magnetic resonance (MR) images—realistic but completely different from the original ones—using generative adversarial networks (GANs). This study exploits progressive growing of GANs (PGGANs), a multistage generative training method, to generate original-sized  $256 \times 256$  MR images for convolutional neural network-based brain tumor detection, which is challenging *via* conventional GANs; difficulties arise due to unstable GAN training with high resolution and a variety of tumors in size, location, shape, and contrast. Our preliminary results show that this novel PGGAN-based DA method can achieve a promising performance improvement, when combined with classical DA, in tumor detection and also in other medical imaging tasks.

---

C. Han (✉) · H. Nakayama  
Graduate School of Information Science and Technology,  
The University of Tokyo, Tokyo, Japan  
e-mail: [han@nlab.ci.i.u-tokyo.ac.jp](mailto:han@nlab.ci.i.u-tokyo.ac.jp)

L. Rundo · G. Mauri  
Department of Informatics, Systems and Communication,  
University of Milano-Bicocca, Milan, Italy

L. Rundo  
Institute of Molecular Bioimaging and Physiology (IBFM),  
Italian National Research Council (CNR), Cefalù (PA), Italy

R. Araki  
Graduate School of Engineering, Chubu University, Aichi, Japan

Y. Furukawa  
Kanto Rosai Hospital, Kanagawa, Japan

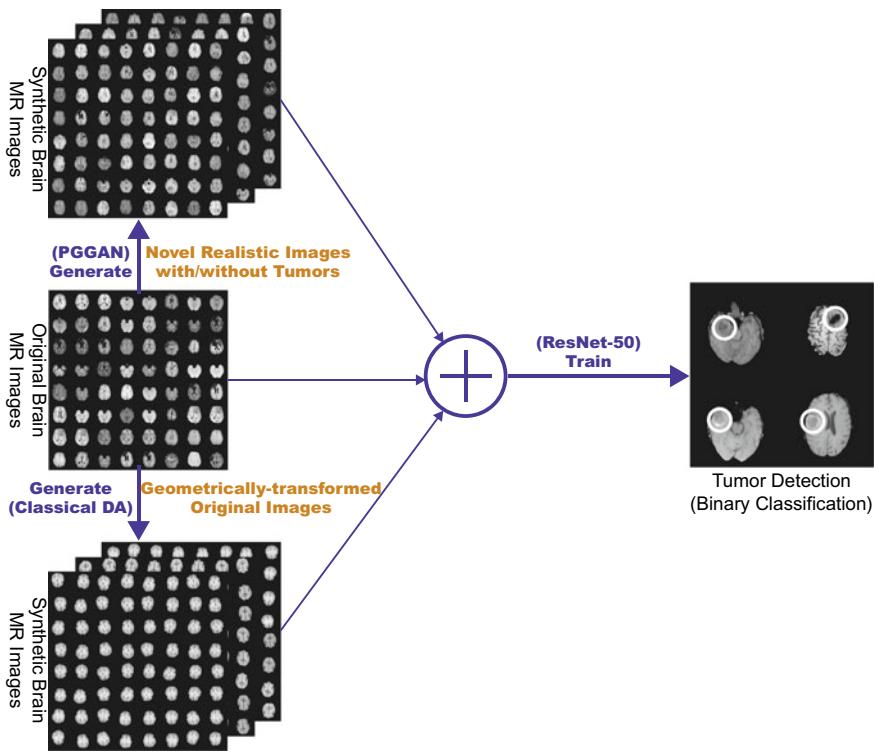
H. Hayashi  
Department of Advanced Information Technology, Kyushu University, Fukuoka, Japan

## 27.1 Introduction

Along with classical methods [1, 2], convolutional neural networks (CNNs) have dramatically improved medical image analysis [3, 4], such as brain magnetic resonance imaging (MRI) segmentation [5, 6], primarily thanks to large-scale annotated training data. Unfortunately, obtaining such massive medical data is challenging; consequently, better training requires intensive data augmentation (DA) techniques, such as geometric/intensity transformations of original images [7, 8]. However, those transformed images intrinsically have a similar distribution with respect to the original ones, leading to limited performance improvement; thus, generating realistic (i.e., similar to the real image distribution) but completely new samples is essential to fill the real image distribution uncovered by the original dataset. In this context, generative adversarial network (GAN)-based DA is promising, as it has shown excellent performance in computer vision, revealing good generalization ability. Especially, SimGAN outperformed the state of the art with 21% improvement in eye gaze estimation [9].

Also in medical imaging, realistic retinal image and computed tomography (CT) image generation have been tackled using adversarial learning [10, 11]; a very recent study reported performance improvement with synthetic training data in CNN-based liver lesion classification, using a small number of  $64 \times 64$  CT images for GAN training [12]. However, GAN-based image generation using MRI, the most effective modality for soft-tissue acquisition, has not yet been reported due to the difficulties from low-contrast MR images, strong anatomical consistency, and intra-sequence variability; in our previous work [13], we generated  $64 \times 64/128 \times 128$  MR images using conventional GANs and even an expert physician failed to accurately distinguish between the real/synthetic images.

So, how can we generate highly realistic and original-sized  $256 \times 256$  images, while maintaining clear tumor/non-tumor features using GANs? Our aim is to generate GAN-based synthetic contrast-enhanced T1-weighted (T1c) brain MR images—the most commonly used sequence in tumor detection thanks to its high contrast [14, 15]—for CNN-based tumor detection. This computer-assisted brain tumor MRI analysis task is clinically valuable for better diagnosis, prognosis, and treatment [5, 6]. Generating  $256 \times 256$  images is extremely challenging: (*i*) GAN training is unstable with high-resolution inputs, and severe artifacts appear due to strong consistency in brain anatomy; (*ii*) brain tumors vary in size, location, shape, and contrast. However, it is beneficial, because most CNN architectures adopt around  $256 \times 256$  input sizes (e.g., Inception-ResNet-V2 [16]:  $299 \times 299$ , ResNet-50 [17]:  $224 \times 224$ ) and we can achieve better results with original-sized image augmentation—toward this, we use progressive growing of GANs (PGGANs), a multistage generative training method [18]. Moreover, an expert physician evaluates the generated images’ realism and tumor/non-tumor features *via* the visual Turing test [19]. Using the synthetic images, our novel PGGAN-based DA approach achieves better performance in CNN-based tumor detection, when combined with classical DA (Fig. 27.1).



**Fig. 27.1** PGGAN-based DA for better tumor detection: The PGGANs method generates a number of realistic brain tumor/non-tumor MR images, and the binary classifier uses them as additional training data

*Contributions.* Our main contributions are as follows:

- **MR Image Generation:** This research explains how to exploit MRI data to generate realistic and original-sized  $256 \times 256$  whole-brain MR images using PGGANs, while maintaining clear tumor/non-tumor features.
- **MR Image Augmentation:** This study shows encouraging results on PGGAN-based DA, when combined with classical DA, for better tumor detection and other medical imaging tasks.

The rest of the manuscript is organized as follows: Sect. 27.2 introduces background on GANs; Sect. 27.3 describes our MRI dataset and PGGAN-based DA approach for tumor detection with its validations; experimental results are shown and analyzed in Sect. 27.4; Sect. 27.5 presents conclusion and future work.

## 27.2 Generative Adversarial Networks

Originally proposed by Goodfellow et al. in 2014 [20], GANs have shown remarkable results in image generation [21] relying on a two-player minimax game: A generator network aims at generating realistic images to fool a discriminator network that aims at distinguishing between the real/synthetic images. However, the two-player objective function leads to difficult training accompanying artificiality and mode collapse [22], especially with high resolution. Deep convolutional GAN (DCGAN) [23], the most standard GAN, results in stable training on  $64 \times 64$  images. In this context, several multistage generative training methods have been proposed: Composite GAN exploits multiple generators to separately generate different parts of an image [24]; the PGGANs method adopts multiple training procedures from low resolution to high to incrementally generate a realistic image [18].

Recently, researchers applied GANs to medical imaging, mainly for image-to-image translation, such as segmentation [25], super-resolution [26], and cross-modality translation [27]. Since GANs allow for adding conditional dependency on the input information (e.g., category, image, and text), they used such conditional GANs to produce the desired corresponding images. However, GAN-based research on generating large-scale synthetic training images is limited, while the biggest challenge in this field is handling small datasets.

Differently from a very recent DA work for  $64 \times 64$  CT liver lesion region of interest (ROI) classification [12], to the best of our knowledge, this is the first GAN-based whole MR image augmentation approach. This work also firstly uses PGGANs to generate  $256 \times 256$  medical images. Along with classical transformations of real images, a completely different approach—generating novel realistic images using PGGANs—may become a clinical breakthrough.

## 27.3 Materials and Methods

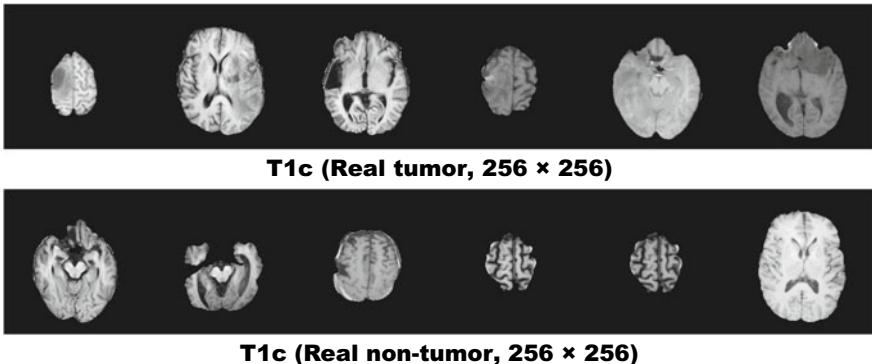
### 27.3.1 *BRATS 2016 Training Dataset*

This paper exploits a dataset of  $240 \times 240$  T1c brain axial MR images containing 220 high-grade glioma cases to train PGGANs with sufficient data and image resolution. These MR images are extracted from the Multimodal Brain Tumor Image Segmentation Benchmark (BRATS) 2016 [28].

### 27.3.2 *PGGAN-Based Image Generation*

#### 27.3.2.1 *Data Preparation*

We select the slices from #30 to #130 among the whole 155 slices to omit initial/final slices, since they convey a negligible amount of useful information and negatively



**Fig. 27.2** Example of real  $256 \times 256$  MR images used for PGGAN training

affect the training of both PGGANs and ResNet-50. For tumor detection, our whole dataset (220 patients) is divided into: (i) a training set (154 patients); (ii) a validation set (44 patients); and (iii) a test set (22 patients). Only the training set is used for the PGGAN training to be fair. Since tumor/non-tumor annotations are based on 3D volumes, these labels are often incorrect/ambiguous on 2D slices; so, we discard (i) tumor images tagged as non-tumor, (ii) non-tumor images tagged as tumor, (iii) unclear boundary images, and (iv) too small/big images; after all, our datasets consist of:

- Training set (5,036 tumor/3,853 non-tumor images);
- Validation set (793 tumor/640 non-tumor images);
- Test set (1,575 tumor/1,082 non-tumor images).

The images from the training set are zero-padded to reach a power of 2,  $256 \times 256$  from  $240 \times 240$  pixels for better PGGAN training. Figure 27.2 shows examples of real MR images.

### 27.3.2.2 PGGANs

PGGAN is a novel training method for GANs with a progressively growing generator and discriminator [18]: Starting from low resolution, newly added layers model fine-grained details as training progresses. As Fig. 27.3 shows, we adopt PGGANs to generate highly realistic and original-sized  $256 \times 256$  brain MR images; tumor/non-tumor images are separately trained and generated.

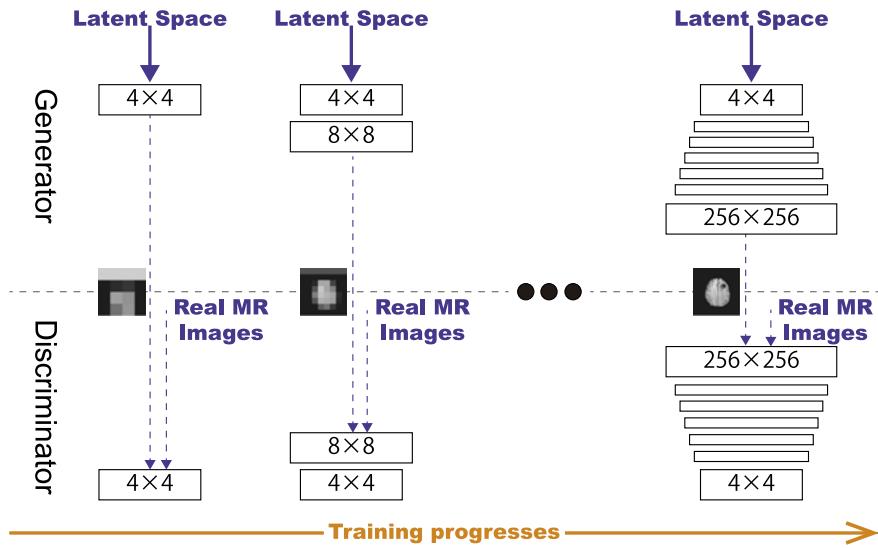


Fig. 27.3 PGGANs architecture for synthetic  $256 \times 256$  MR image generation

### 27.3.2.3 PGGAN Implementation Details

We use the PGGAN architecture with the Wasserstein loss using gradient penalty [22]. Training lasts for 100 epochs with a batch size of 16 and  $1.0 \times 10^{-3}$  learning rate for Adam optimizer.

### 27.3.3 Tumor Detection Using ResNet-50

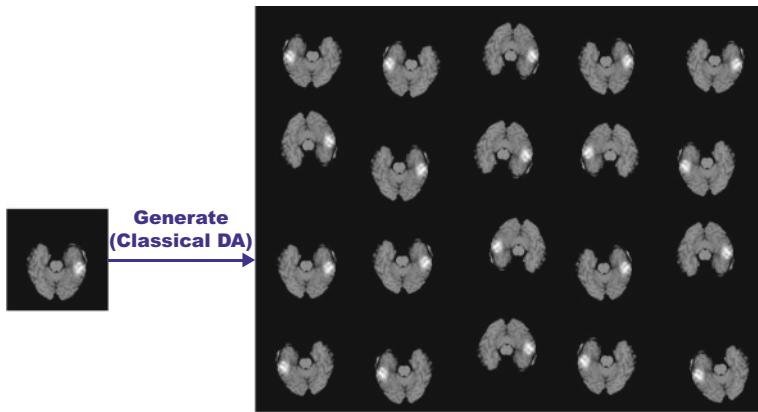
#### 27.3.3.1 Preprocessing

To fit ResNet-50's input size, we center-crop the whole images from  $240 \times 240$  to  $224 \times 224$  pixels.

#### 27.3.3.2 ResNet-50

ResNet-50 is a residual learning-based CNN with 50 layers [17]: Unlike conventional learning unreference functions, it reformulates the layers as learning residual functions for sustainable and easy training. We adopt ResNet-50 to detect tumors in brain MR images, i.e., the binary classification of images with/without tumors.

To confirm the effect of PGGAN-based DA, the following classification results are compared: (i) without DA, (ii) with 200,000 classical DA (100,000 for each



**Fig. 27.4** Example of real MR image and its geometrically transformed synthetic images

class), (iii) with 200,000 PGGAN-based DA, and (iv) with both 200,000 classical DA and 200,000 PGGAN-based DA; the classical DA adopts a random combination of horizontal/vertical flipping, rotation up to 10 degrees, width/height shift up to 8%, shearing up to 8%, zooming up to 8%, and constant filling of points outside the input boundaries (Fig. 27.4). For better DA, highly unrealistic PGGAN-generated images are manually discarded.

### 27.3.3.3 ResNet-50 Implementation Details

We use the ResNet-50 architecture pre-trained on ImageNet with a dropout of 0.5 before the final softmax layer, along with a batch size of 192,  $1.0 \times 10^{-3}$  learning rate for Adam optimizer, and early stopping of 10 epochs.

### 27.3.4 Clinical Validation Using the Visual Turing Test

To quantitatively evaluate (i) how realistic the PGGAN-based synthetic images are and (ii) how obvious the synthetic images' tumor/non-tumor features are, we supply, in a random order, to an expert physician a random selection of:

- 50 real tumor images;
- 50 real non-tumor images;
- 50 synthetic tumor images;
- 50 synthetic non-tumor images.

Then, the physician is asked to constantly classify them as both (i) real/synthetic and (ii) tumor/non-tumor, without previous training stages revealing which is

real/synthetic and tumor/non-tumor; here, we only show successful cases of synthetic images, as we can discard failed cases for better data augmentation. The so-called visual Turing test [19] is used to probe the human ability to identify attributes and relationships in images, also in evaluating the visual quality of GAN-generated images [9]. Similarly, this applies to medical images in clinical environments [11, 12], wherein physicians' expertise is critical.

### 27.3.5 Visualization Using t-SNE

To visually analyze the distribution of both (*i*) real/synthetic and (*ii*) tumor/non-tumor images, we use t-distributed stochastic neighbor embedding (t-SNE) [29] on a random selection of:

- 300 real non-tumor images;
- 300 geometrically transformed non-tumor images;
- 300 PGGAN-generated non-tumor images;
- 300 real tumor images;
- 300 geometrically transformed tumor images;
- 300 PGGAN-generated tumor images.

Only 300 images per each category are selected for better visualization. t-SNE is a machine learning algorithm for dimensionality reduction to represent high-dimensional data into a lower-dimensional (2D/3D) space. It nonlinearly adapts to input data using perplexity, which balances between the data's local and global aspects.

#### 27.3.5.1 t-SNE Implementation Details

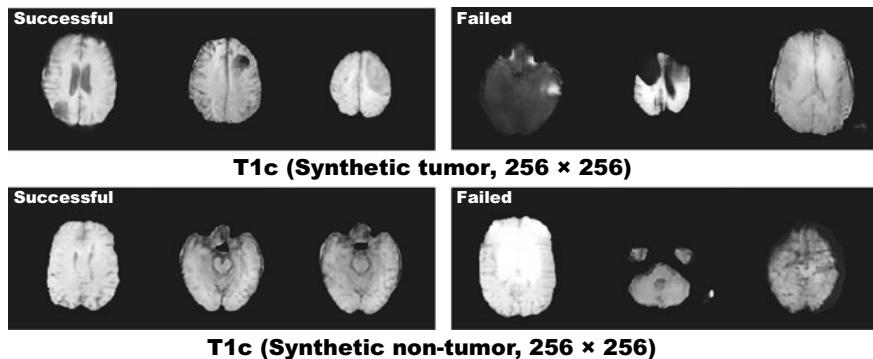
We use t-SNE with a perplexity of 100 for 1,000 iterations to obtain a 2D visual representation.

## 27.4 Results

This section shows how PGGANs generate synthetic brain MR images. The results include instances of synthetic images, their quantitative evaluation by an expert physician, and their influence on tumor detection.

### 27.4.1 MR Images Generated by PGGANs

Figure 27.5 illustrates examples of synthetic tumor/non-tumor images by PGGANs. In our visual confirmation, for about 75% of cases, PGGANs successfully capture



**Fig. 27.5** Example of synthetic MR images yielded by PGGANs: **a** successful cases and **b** failed cases

the T1c-specific texture and tumor appearance while maintaining the realism of the original brain MR images; however, for about 25% of cases, the generated images lack clear tumor/non-tumor features or contain unrealistic features, such as hyperintensity, gray contours, and odd artifacts.

#### 27.4.2 Tumor Detection Results

Table 27.1 shows the classification results for detecting brain tumors with/without DA techniques. As expected, the test accuracy improves by 0.64% with the additional 200, 000 geometrically transformed images for training. When only the PGGAN-based DA is applied, the test accuracy decreases drastically with almost 100% of sensitivity and 6.84% of specificity, because the classifier recognizes the synthetic images' prevailed unrealistic features as tumors, similarly to anomaly detection.

**Table 27.1** Binary classification results for detecting brain tumors with/without DA

Experimental condition	Accuracy (%)	Sensitivity (%)	Specificity (%)
ResNet-50 (w/o DA)	90.06	85.27	97.04
ResNet-50 (w/200k classical DA)	90.70	88.70	93.62
ResNet-50 (w/200k PGGAN-based DA)	62.02	<b>99.94</b>	6.84
ResNet-50 (w/200k classical DA + 200k PGGAN-based DA)	<b>91.08</b>	86.60	<b>97.60</b>

**Table 27.2** Visual Turing test results by a physician for classifying real ( $R$ ) versus synthetic ( $S$ ) images and tumor ( $T$ ) vs non-tumor ( $N$ ) images

Real/synthetic classification	$R$ as $R$	$R$ as $S$	$S$ as $R$	$S$ as $S$
78.5%	58	42	1	99
Tumor/non-tumor classification	$T$ as $T$	$T$ as $N$	$N$ as $T$	$N$ as $N$
90.5%	82	18 ( $R$ : 5, $S$ : 13)	1 ( $S$ : 1)	99

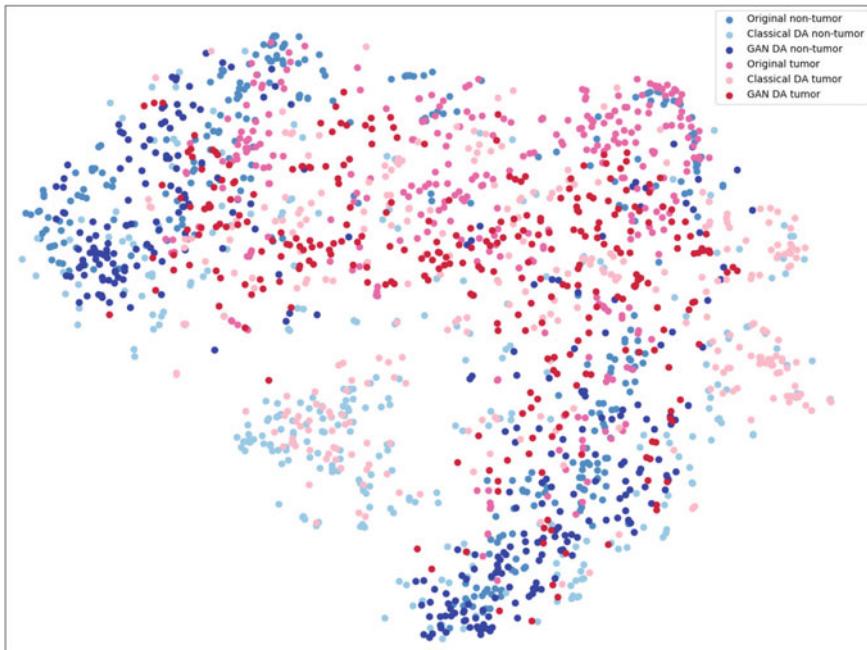
However, surprisingly, when it is combined with the classical DA, the accuracy increases by 1.02% with higher sensitivity and specificity; this could occur because the PGGAN-based DA fills the real image distribution uncovered by the original dataset, while the classical DA provides the robustness on training for most cases.

#### 27.4.3 Visual Turing Test Results

Table 27.2 shows the confusion matrix for the visual Turing test. Differently from our previous work on GAN-based  $64 \times 64/128 \times 128$  MR image generation, the expert physician easily recognizes  $256 \times 256$  synthetic images [13], while tending also to classify real images as synthetic; this can be attributed to high resolution associated with more difficult training and detailed appearance, making artifacts stand out, which is coherent to the ResNet-50's low tumor detection accuracy with only the PGGAN-based DA. Generally, the physician's tumor/non-tumor classification accuracy is high and the synthetic images successfully capture tumor/non-tumor features. However, unlike non-tumor images, the expert recognizes a considerable number of tumor images as non-tumor, especially on the synthetic images; this results from the remaining real images' ambiguous annotation, which is amplified in the synthetic images trained on them.

#### 27.4.4 t-SNE Result

As presented in Fig. 27.6, tumor/non-tumor images' distribution shows a tendency that non-tumor images locate from top left to bottom right and tumor images locate from top right to center, while the distinction is unclear with partial overlaps. Classical DA covers a wide range, including zones without any real/GAN-generated images, but tumor/non-tumor images often overlap there. Meanwhile, PGGAN-generated images concentrate differently from real images, while showing more frequent overlaps than the real ones; this probably derives from those synthetic images with unsatisfactory realism and tumor/non-tumor features.



**Fig. 27.6** t-SNE result on six categories, with 300 images per each category: **a** real tumor/non-tumor images; **b** geometrically transformed tumor/non-tumor images; and **c** PGGAN-generated tumor/non-tumor images

## 27.5 Conclusion

Our preliminary results show that PGGANs can generate original-sized  $256 \times 256$  realistic brain MR images and achieve higher performance in tumor detection, when combined with classical DA. This occurs because PGGANs' multistage image generation obtains good generalization and synthesizes images with the real image distribution unfilled by the original dataset. However, considering the visual Turing test and t-SNE results, yet unsatisfactory realism with high resolution strongly limits DA performance, so we plan to (*i*) generate only realistic images and then (*ii*) refine synthetic images more similar to the real image distribution.

For (*i*), we can map an input random vector onto each training image [30] and generate images with suitable vectors, to control the divergence of generated images; virtual adversarial training could be also integrated to control the output distribution. Moreover, (*ii*) can be achieved by GAN/VAE-based image-to-image translation, such as unsupervised image-to-image translation networks [31], considering SimGAN's remarkable performance improvement after refinement [9]. Moreover, we should further avoid real images with ambiguous/inaccurate annotation for better tumor detection.

Overall, our novel PGGAN-based DA approach sheds light on diagnostic and prognostic medical applications, not limited to tumor detection; future studies are needed to extend our encouraging results.

**Acknowledgements** This work was partially supported by the Graduate Program for Social ICT Global Creative Leaders of the University of Tokyo by JSPS.

## References

1. Rundo, L., Militello, C., Russo, G., Vitabile, S., Gilardi, M.C., Mauri, G.: GTVcut for neuro-radiosurgery treatment planning: an MRI brain cancer seeded image segmentation method based on a cellular automata model. *Nat. Comput.* **17**(3), 521–536 (2018)
2. Rundo, L., Militello, C., Vitabile, S., Russo, G., Pisciotta, P., Marletta, F., Ippolito, M., DArrigo, C., Midiri, M., Gilardi, M.C.: Semi-automatic brain lesion segmentation in Gamma Knife treatments using an unsupervised fuzzy c-means clustering technique. In: Advances in Neural Networks: Computational Intelligence for ICT. Volume 54 of Smart Innovation, Systems and Technologies, pp. 15–26. Springer (2016)
3. Bevilacqua, V., Brunetti, A., Cascarano, G.D., Palmieri, F., Guerriero, A., Moschetta, M.: A deep learning approach for the automatic detection and segmentation in autosomal dominant polycystic kidney disease based on Magnetic Resonance images. In: Proceedings of International Conference on Intelligent Computing (ICIP), pp. 643–649. Springer (2018)
4. Brunetti, A., Carnimeo, L., Trotta, G.F., Bevilacqua, V.: Computer-assisted frameworks for classification of liver, breast and blood neoplasias via neural networks: a survey based on medical images. *Neurocomputing* **335**, 274–298 (2018)
5. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., et al.: Brain tumor segmentation with deep neural networks. *Med. Image Anal.* **35**, 18–31 (2017)
6. Kamnitsas, K., Ledig, C., Newcombe, V.F.J., Simpson, J.P., Kane, A.D., Menon, D.K., et al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **36**, 61–78 (2017)
7. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 234–241 (2015)
8. Milletari, F., Navab, N., Ahmadi, S.: V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: Proceedings of International Conference on 3D Vision (3DV), pp. 565–571. IEEE (2016)
9. Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W., Webb, R.: Learning from simulated and unsupervised images through adversarial training. In: Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2107–2116. IEEE (2017)
10. Costa, P., Galdran, A., Meyer, M.I., Niemeijer, M., Abràmoff, M., Mendona, A.M., Campilho, A.: End-to-end adversarial retinal image synthesis. *IEEE Trans. Med. Imaging* **37**(3), 781–791 (2018)
11. Chuquicusma, M.J.M., Hussein, S., Burt, J., Bagci, U.: How to fool radiologists with generative adversarial networks? a visual Turing test for lung cancer diagnosis. In: Proceedings of International Symposium on Biomedical Imaging (ISBI), pp. 240–244. IEEE (2018)
12. Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., Greenspan, H.: GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* **321**, 321–331 (2018)
13. Han, C., Hayashi, H., Rundo, L., Araki, R., Shimoda, W., Muramatsu, S., et al.: GAN-based synthetic brain MR image generation. In: Proceedings of International Symposium on Biomedical Imaging (ISBI), pp. 734–738. IEEE (2018)

14. Militello, C., Rundo, L., Vitabile, S., et al.: Gamma knife treatment planning: MR brain tumor segmentation and volume measurement based on unsupervised fuzzy c-means clustering. *Int. J. Imaging Syst. Technol.* **25**(3), 213–225 (2015)
15. Rundo, L., Stefano, A., Militello, C., Russo, G., Sabini, M.G., D’Arrigo, C., Marletta, F., Ippolito, M., Mauri, G., Vitabile, S., Gilardi, M.C.: A fully automatic approach for multimodal PET and MR image segmentation in Gamma Knife treatment planning. *Comput. Methods Programs Biomed.* **144**, 77–96 (2017)
16. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Proceedings of AAAI Conference on Artificial Intelligence (AAAI) (2017)
17. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. IEEE (2016)
18. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of GANs for improved quality, stability, and variation. In: Proceedings of International Conference on Learning Representations (ICLR). arXiv preprint [arXiv:1710.10196v3](https://arxiv.org/abs/1710.10196v3) (2018)
19. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training GANs. In: Advances in Neural Information Processing Systems (NIPS), pp. 2234–2242 (2016)
20. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems (NIPS), pp. 2672–2680 (2014)
21. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of International Conference on Computer Vision (ICCV), pp. 2242–2251. IEEE (2017)
22. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of Wasserstein GANs. In: Advances in Neural Information Processing Systems, pp. 5769–5779 (2017)
23. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. In: Proceedings of International Conference on Learning Representations (ICLR). arXiv preprint [arXiv:1511.06434](https://arxiv.org/abs/1511.06434) (2016)
24. Kwak, H., Zhang, B.: Generating images part by part with composite generative adversarial networks. arXiv preprint [arXiv:1607.05387](https://arxiv.org/abs/1607.05387) (2016)
25. Xue, Y., Xu, T., Zhang, H., Long, L.R., Huang, X.: SegAN: Adversarial network with multi-scale  $L_1$  loss for medical image segmentation. *Neuroinformatics* **16**(3–4), 383–392 (2018)
26. Mahapatra, D., Bozorgtabar, B., Hewavitharanage, S., Garnavi, R.: Image super resolution using generative adversarial networks and local saliency maps for retinal image analysis. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 382–390 (2017)
27. Nie, D., Trullo, R., Lian, J., Petitjean, C., Ruan, S., Wang, Q., Shen, D.: Medical image synthesis with context-aware generative adversarial networks. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 417–425 (2017)
28. Menze, B.H., Jakab, A., Bauer, S., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **34**(10), 1993–2024 (2015)
29. van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008)
30. Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: Proceedings of International Conference on Information Processing in Medical Imaging (IPMI), pp. 146–157 (2017)
31. Liu, M.Y., Breuel, T., Kautz, J.: Unsupervised image-to-image translation networks. In: Advances in Neural Information Processing Systems (NIPS), pp. 700–708 (2017)

# Chapter 28

## DNA Microarray Classification: Evolutionary Optimization of Neural Network Hyper-parameters



**Pietro Barbiero, Andrea Bertotti, Gabriele Ciravegna, Giansalvo Cirrincione and Elio Piccolo**

**Abstract** The analysis of complex systems, such as cancer resistance to drugs, requires flexible algorithms but also simple models, as they will be used by biologists in order to get insights on the underlying phenomenon. Exploiting the availability of the largest collection of patient-derived xenografts from metastatic colorectal cancer annotated for response to therapies, this manuscript aims to (i) forecast the response to treatments on human tissues using murine information; (ii) providing a trade-off between model accuracy and interpretability, evolving a shallow neural network using a genetic algorithm.

### 28.1 Introduction

The objective of the work consists of the forecast of patient-derived xenografts (PDX, [1–3]) of some metastatic colorectal cancers (CRC) as responsive or not responsive to cetuximab treatments [4–7]. The analyses are a natural update and extension of those proposed in [8]. While previous works attempt to forecast responsiveness in

---

P. Barbiero (✉)

Department of Mathematical Sciences (DISMA), Politecnico di Torino, 10126 Torino, Italy  
e-mail: [pietro.barbiero@studenti.polito.it](mailto:pietro.barbiero@studenti.polito.it)

A. Bertotti

Dipartimento di Oncologia, Candiolo Cancer Institute - FPO, IRCCS,  
Università degli studi di Torino, Torino, Italy  
e-mail: [andrea.bertotti@ircc.it](mailto:andrea.bertotti@ircc.it)

G. Ciravegna · E. Piccolo

Department of Control and Computer Engineering (DAUIN), Politecnico di Torino,  
10126 Torino, Italy  
e-mail: [elio.piccolo@polito.it](mailto:elio.piccolo@polito.it)

G. Cirrincione

Lab. LTI, University of Picardie Jules Verne, Amiens, France  
e-mail: [exin@u-picardie.fr](mailto:exin@u-picardie.fr)

E. Piccolo

University of South Pacific, Suva, Fiji

© Springer Nature Singapore Pte Ltd. 2020

A. Esposito et al. (eds.), *Neural Approaches to Dynamics of Signal Exchanges*,  
Smart Innovation, Systems and Technologies 151,  
[https://doi.org/10.1007/978-981-13-8950-4\\_28](https://doi.org/10.1007/978-981-13-8950-4_28)

mice tissues, the goal of this work is to predict the response in human tissues using murine information only, exploiting a transfer learning approach.

The two data sets used in this work are DNA microarray sequences composed of the expression of 20,023 genes. The murine data set contains 403 CRC murine tissues, while the human one contains 94 samples [9]. Each cancerous tissue is associated with a Boolean variable describing the tumor response to cetuximab, as two classes, responsive (1) and not responsive to treatments (0), as described in previous works [10]. Since the final goal of the analysis consists of patient classification, genes are considered as features and patient data as samples. The dataset is normalized using a column statistical scaling (zscore) on features.

## 28.2 Proposed Approach

The analysis of complex systems, such as cancer resistance, requires powerful algorithms but also simple models, as they will be used by biologists in order to get insights on the underlying phenomenon. For this reason, the proposed approach exploits an evolutionary algorithm in order to optimize the parameters of a shallow neural network, following similar attempts [11]. The main advantage of this framework consists in providing a trade-off between accuracy and model interpretability.

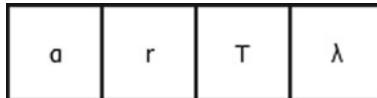
The mathematical model exploited during the evolution is a neural network with a shallow structure known as *SoftMax Adaline* [12–14]. The model can be synthesized in the following:

$$\min_{w,b} J(w, b) = \frac{1}{m} \sum_i^m \mathcal{L}(\hat{y}^{(i)}, y^{(i)}) + \frac{\lambda}{2m} ||w||_F^2 \quad (28.1)$$

$$s.t. \quad \lambda \geq 0 \quad (28.2)$$

where the objective function (or cost function)  $J$  to be minimized is a combination of the average cross-entropy loss  $\mathcal{L}$  over the training data and the L2 regularization of the weights  $w$ . The optimization problem is solvable by means of particular choices for  $w$  and  $b$ . The chosen solver is the Adam optimizer [15], which is a modern and improved version of the backpropagation rule with gradient descent. The main hyperparameters of this simple neural network are: (i) the learning rate  $\alpha$ , (ii) the learning decay rate  $r$ , (iii) the number of epochs  $T$ , and (iv) the regularization parameter  $\lambda$ .

Evolutionary algorithms (EAs) are powerful metaheuristic procedures able to explore efficiently a portion of a complex (NP-hard or NP-complete) problem space and find good approximate solutions. The hyper-parameter optimization of a neural network is a complex problem because there are no polynomial time algorithms able to solve it. One of the most widely spread EA is the genetic algorithm (GA) [16, 17]. GAs are metaheuristics inspired by natural selection processes. Broadly speaking, GAs involve the evolution of a population of candidate solutions toward better ones. A predefined fitness function evaluates the individual goodness.



**Fig. 28.1** Graphical representation of an individual. It is represented as an ordered list of “genetic material”. Each “gene” stands for a neural network hyper-parameter

As before outlined, the objective of this work is the optimization of a neural network model. In particular, the MLP model presented in the previous section can be optimized tuning its hyper-parameters. Since each set of hyper-parameters uniquely identifies a neural network, then each neural network can be represented by its hyper-parameters. For this reason, an ordered list of hyper-parameters is an efficient representation of a MLP model. Having assigned a value to each hyper-parameter from its domain, then the list is called candidate solution or individual (Fig. 28.1).

The generator is an EA method devoted to the initialization of new individuals. One of the most common generators is a random generator. In this case, each hyper-parameter is sampled by a uniform distribution within a predefined range. The ranges chosen are:

- $\alpha \in [10^{-1}, 10^{-7}]$
- $r \in [10^{-1}, 10^{-7}]$
- $T \in [10, 300]$
- $\lambda \in [10^{-1}, 10^{-8}]$ .

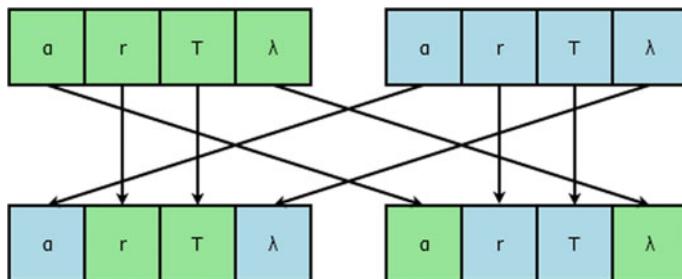
The uniform distribution guarantees that the initial individuals are sufficiently different from each other. This biodiversity will help the EA search since there is more genetic material available for exchanges.

In order to optimize individuals generation by generation, the next population should be better than the previous one. Therefore, after the generation process, each individual is evaluated in order to estimate the goodness of its genetic material. To do this, a neural network is built for each candidate solution, i.e., it is set up by using the corresponding hyper-parameters. The learning process is validated through a 10-fold cross-validation. At the end of the training process, the average validation accuracy is considered as an estimate of the individual fitness.

After the evaluation process, the GA selects a random set of individuals from the population and selects a subset of them through a fitness-based criterion. It ranks the randomly selected individuals in ascending order according to their fitness and it picks out some of the best ones. These small sets of individuals are then used to produce the next generation.

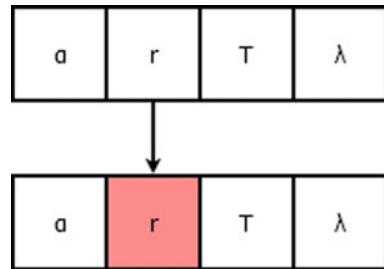
In order to modify and (hopefully) improve the current population, the selected solutions should be slightly modified. At first, they go through a crossover (or recombination) process in which pairs of individuals mix their genetic material to produce two new child solutions (Figs. 28.2 and 28.3).

Secondly, the child solutions randomly mutate one of their components (hyper-parameters).



**Fig. 28.2** During the crossover process, two individuals mix their genetic material in order to produce two child solutions

**Fig. 28.3** Figure shows the mutation process. The individual randomly mutate one of its “genes”



After having generated new candidate solutions, the replacer method selects the best half of old population individuals and the best half of offsprings in order to select the candidate solution of the next generation.

The entire process is repeated for a certain amount of generations exploring the hyper-parameter space and providing better and better configurations.

---

#### Algorithm 1 Genetic Algorithm

---

- 1: **Input:** DNA microarray and cancer growth targets
  - 2: Generate random individuals
  - 3: **for each** generation **do**
  - 4:   **for each** individual **do**
  - 5:     Evaluate the individual through cross-validation
  - 6:   **end for**
  - 7:   Select next generation parents
  - 8:   Breed random parents
  - 9:   Mutate child individuals
  - 10:   Replace old individuals with the new generated ones
  - 11:   **go to** next generation
  - 12: **end for**
  - 13: **Output:** best individual
-

The GA set up includes at least the choice of the population size, the maximum amount of generations and the mutation rate. In this work the following choices are made:

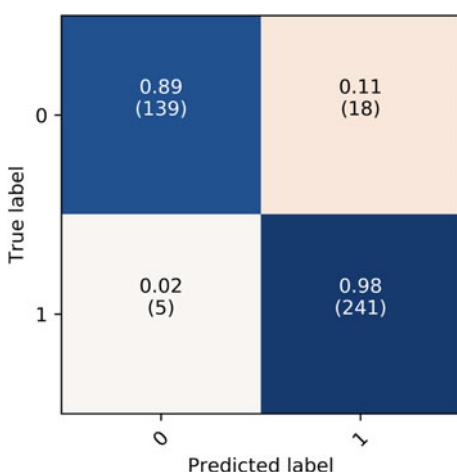
- population size: 20
- number of generations: 100
- mutation rate: 25%.

### 28.3 Experiments and Discussion

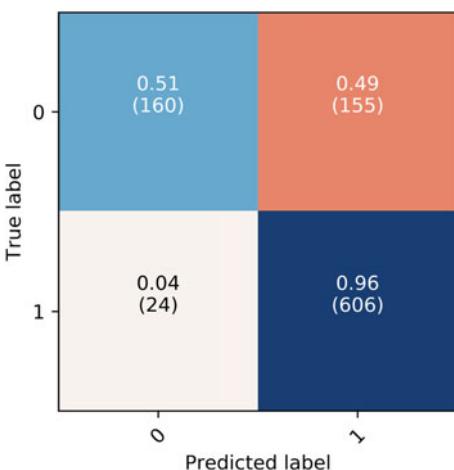
The experiments are carried out using a 10-fold cross-validation framework. At first, the GA optimizes the hyper-parameters of the Adaline model exploiting the murine training set. At the end of the evolution process, the best individual is picked up and the corresponding neural network is evaluated by using a murine test set and the whole human data set as test set. Repeating randomly the same process 10 times shuffling training and test data of the murine data set it is possible to evaluate the noise due to the use of specific murine samples in the training set. Figures 28.4 and 28.5 show the confusion matrices containing the results of all the ten folds for murine and human test sets, respectively. By transferring learning from mice to humans, the accuracy of classifiers reduces from  $\sim 0.94$  to  $\sim 0.81$ .

Compared with deep neural models, a trained Adaline network has a straightforward interpretation as the model parameters have a one-to-one correspondence with features (genes). By adding a weight regularization term in the cost function of the neural network, useless weights are shrunk toward zero. Figure 28.6 shows the distribution of the weights of the best individuals averaged across the ten folds. Features corresponding to zero weights may be considered as useless for the purpose

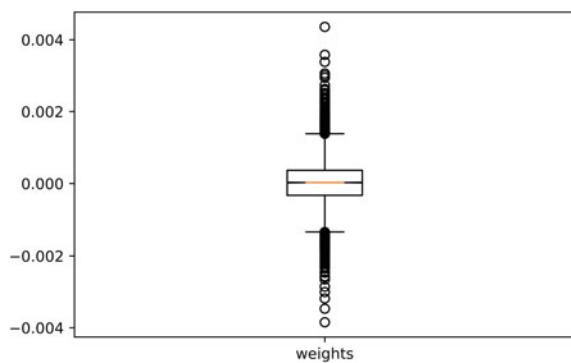
**Fig. 28.4** Confusion matrix containing cumulative results on murine test sets. 0 → not responsive, 1 → responsive samples



**Fig. 28.5** Confusion matrix containing cumulative results on the whole human data set. 0 → not responsive, 1 → responsive samples



**Fig. 28.6** Notched boxplot containing the average value of neural network weights of the best individuals



of classification. As most of the weights are zero, the number of relevant features is small with respect to the whole set. Thus, the evolved Adaline networks may be used as accurate and interpretable models.

## 28.4 Conclusions

While previous works attempt to forecast responsiveness in mice tissues, in this work is addressed the problem of predicting the response in human tissues using murine information only, exploiting a transfer learning approach. The main advantage of the proposed approach consists of providing the best possible trade-off between model accuracy and interpretability, evolving a shallow neural network using a genetic

algorithm. The predictions on unseen murine test sets and on an external human data set seem promising. In future works, biologist may exploit the proposed approach for the analysis of the features selected by the evolved neural networks.

## References

1. Hidalgo, M., et al.: Patient-derived Xenograft models: An emerging platform for translational cancer research. *Cancer Discov.* **4**, 998–1013 (2014)
2. Tentler, J.J., et al.: Patient-derived tumour xenografts as models for oncology drug development. *Nat. Rev. Clin. Oncol.* **9**, 338–350 (2012)
3. Byrne, A.T., et al.: Interrogating open issues in cancer precision medicine with patient derived xenografts. *Nat. Rev. Cancer* (2017). <https://doi.org/10.1038/nrc.2016.140>
4. Bertotti, A., et al.: A molecularly annotated platform of patient-derived xenografts ('xenopatients') identifies HER2 as an effective therapeutic target in cetuximab-resistant colorectal cancer. *Cancer Discov.* **1**, 508–523 (2011)
5. Zanella, E.R., et al.: IGF2 is an actionable target that identifies a distinct subpopulation of colorectal cancer patients with marginal response to anti-EGFR therapies. *Sci. Transl. Med.* **7**, (2015)
6. Bertotti, A., et al.: The genomic landscape of response to EGFR blockade in colorectal cancer. *Nature* **526**, 263–7 (2015)
7. Sartore Bianchi, A. et al., Dual-targeted therapy with trastuzumab and lapatinib in treatment-refractory, KRAS codon 12/13 wild-type, HER2-positive metastatic colorectal cancer (HERA-CLES): a proof-of-concept, multicentre, open-label, phase 2 trial. *Lancet Oncol.* **17**, 738–746 (2016)
8. Barbiero P., Bertotti A., Ciravagna G., Cirrincione G., Pasero E., Piccolo E.: Supervised gene identification in colorectal cancer. In: Quantifying and Processing Biomedical and Behavioral Signals. Springer (2018). ISBN 9783319950945. [https://doi.org/10.1007/978-3-319-95095-2\\_21](https://doi.org/10.1007/978-3-319-95095-2_21)
9. Illumina: Array-based gene expression analysis. Data Sheet Gene Expr. (2011). [http://res.illumina.com/documents/products/datasheets/datasheet\\_gene\\_exp\\_analysis.pdf](http://res.illumina.com/documents/products/datasheets/datasheet_gene_exp_analysis.pdf)
10. Isella, C., et al.: Selective analysis of cancer-cell intrinsic transcriptional traits defines novel clinically relevant subtypes of colorectal cancer. *Nat. Gen.* **8**, (2017). <https://doi.org/10.1038/ncomms15107>
11. Bevilacqua, V., Mastronardi, G., Menolascina, F.: Genetic algorithm and neural network based classification in microarray data analysis with biological validity assessment. In: International Conference on Intelligent Computing, pp. 475–484. Springer (2006)
12. Widrow, B., Lehr, M.A.: Artificial Neural Networks of the perceptron, madaline, and back-propagation family. In: Neurobionics (1993). <https://doi.org/10.1016/B978-0-444-89958-3.50013-9>
13. Haykin, S.: Neural Networks: A Comprehensive Foundation. Prentice Hall (1998). ISBN 0132733501
14. Chollet, F., et al.: Keras (2015). <https://keras.io>
15. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. In: International Conference for Learning Representations (2017). [arXiv:1412.6980v9](https://arxiv.org/abs/1412.6980v9)
16. Michalewicz, Z., Hartley, S.J.: Genetic algorithms + data structures = evolution programs. *Math. Intell.* **18**(3), 71 (1996)
17. Garrett, A.: Inspyred: bio-inspired algorithms in python (2014). <https://pypi.python.org/pypi/inspyred> (visited on 11/28/2016)

## Chapter 29

# Evaluation of a Support Vector Machine Based Method for Crohn's Disease Classification



S. Franchini, M. C. Terranova, G. Lo Re, S. Salerno, M. Midiri  
and Salvatore Vitabile

**Abstract** Crohn's disease (CD) is a chronic, disabling inflammatory bowel disease that affects millions of people worldwide. CD diagnosis is a challenging issue that involves a combination of radiological, endoscopic, histological, and laboratory investigations. Medical imaging plays an important role in the clinical evaluation of CD. Enterography magnetic resonance imaging (E-MRI) has been proven to be a useful diagnostic tool for disease activity assessment. However, the manual classification process by expert radiologists is time-consuming and expensive. This paper proposes the evaluation of an automatic Support Vector Machine (SVM) based supervised learning method for CD classification. A real E-MRI dataset composed of 800 patients from the University of Palermo Policlinico Hospital (400 patients with histologically proved CD and 400 healthy patients) has been used to evaluate the proposed classification technique. For each patient, a team of radiology experts has extracted a vector composed of 20 features, usually associated with CD, from the related E-MRI examination, while the histological specimen results have been used as the ground-truth for CD diagnosis. The dataset composed of 800 vectors has been used to train and validate the SVM classifier. Automatic techniques for feature space reduction have been applied and validated by the radiologists to optimize the proposed classification method, while  $K$ -fold cross-validation has been used to improve the SVM classifier reliability. The measured indexes (sensitivity: 97.07%, specificity: 96.04%, negative predictive value: 97.24%, precision: 95.80%, accuracy: 96.54%, error: 3.46%) are better than the operator-based reference values reported in the literature. Experimental results also show that the proposed method outperforms the main standard classification techniques.

---

S. Franchini (✉) · M. C. Terranova · G. Lo Re · S. Salerno · M. Midiri · S. Vitabile  
Dipartimento di Biomedicina, Neuroscienze e Diagnostica Avanzata, University of Palermo, Via  
del Vespro, 129, 90127 Palermo, Italy  
e-mail: [silvia.franchini@unipa.it](mailto:silvia.franchini@unipa.it)

## 29.1 Introduction

Crohn's disease (CD) is a chronic, disabling disease that causes inflammation of the gastrointestinal tract. Epidemiological data analysis suggests that the incidence and prevalence rates of CD have been rapidly increasing during the past decades. CD is most often diagnosed in people between 15 and 25 years old, although it can appear at any age. CD has a complex etiopathogenesis; its development results from the effect of environmental factors in genetically predisposed subjects. This condition consists in the granulomatous inflammation of the intestinal walls from the mucosal to the serosa layer, and it is also characterized by extra-luminal and extra-intestinal manifestations. The disease presents heterogeneous behaviors in terms of location and extent of the interested bowel tracts and can involve different healing and relapses events during its course [1–3]. Because of these heterogeneous manifestations of CD, its diagnosis is a challenging issue that requires a combination of radiological, endoscopic, histological, and laboratory investigations [4]. Medical imaging allows for a first-step non-invasive clinical evaluation of CD [5–8]. Recent studies have explored the use of enterography magnetic resonance imaging (E-MRI) as a valid and accurate diagnostic technique for CD activity and extension assessment [1, 2, 5] showing sensitivity and specificity indexes of 93% and 90%, respectively [9]. E-MRI-based diagnosis relies on the evaluation of typical E-MRI features that have been demonstrated to be associated with CD [2, 10, 11]. However, the manual classification process by expert radiologists is time-consuming and expensive. Conversely, automated learning techniques that classify patients into positive or negative classes with respect to CD starting from E-MRI images can allow for early diagnosis of CD and reduce the high-economic costs of such a widespread disease.

### 29.1.1 Related Works

Automated learning methods include supervised and unsupervised machine learning techniques. Supervised learning consists in building a predictive model that can map a set of input variables to a response, while unsupervised learning looks for natural patterns within the dataset without reference to a response or true result. Unsupervised learning methods, including  $k$ -means algorithm, as well as self-organizing maps (SOM), are used for clustering tasks. Classification tasks are usually performed by using supervised learning algorithms or feed-forward neural networks. Supervised classification methods include  $K$ -Nearest Neighbor (KNN), Decision Trees, Naïve Bayes, Discriminant Analysis, and Support Vector Machines (SVM). Recently, several approaches to classify MRI images into positive or negative with respect to some kind of disease have been proposed. A supervised learning method, namely  $K$ -Nearest Neighbor, is used in [12] for MR brain tissue classification, while in [13] another, supervised learning technique, based on a Support Vector Machine, is used to classify MR brain images. Other approaches use unsupervised methods,

such as fuzzy C-means [14] and self-organizing maps (SOM) [13] to classify MR brain images. When compared to other supervised classification methods, Support Vector Machines present advantages such as elegant mathematical treatment, direct geometric interpretation, and high accuracy. An SVM classifier with leave-one-out cross-validation has been presented in [15] for predicting medication adherence in heart failure (HF) patients. In [16] a Kernel Support Vector Machine (KSVM) with Radial Basis Function (RBF) kernel, has been used to classify MR brain images as either normal or abnormal. This hybrid method uses digital wavelet transform to extract features and principal component analysis (PCA) [17] to reduce the feature space dimension and, consequently, the computational cost. The KSVM with five-fold cross-validation is then applied for classification. To the best of our knowledge, SVM-based approaches for CD classification starting from E-MRI images have not been proposed in literature.

### 29.1.2 Our Contribution

This paper proposes the evaluation of a Support Vector Machine based method for Crohn's disease classification using Enterography MRI images. An E-MRI dataset composed of 800 patients from the University of Palermo Policlinico Hospital (400 patients with histologically proved CD and 400 healthy patients) has been used to evaluate the proposed classification technique.

Preliminary results of this approach have been presented in [18] for a dataset of 300 patients. In the study presented in [18], 22 features, usually associated with CD, have been extracted by a team of radiologists starting from the E-MRI examination of each patient. However, two of these features, namely activity and pattern, are composite features that require radiology expertise to be derived. Regarding the pattern, CD can show different patterns and, on the basis of MR features, the radiologists define the CD subtype: 1. Active Inflammatory Subtype: high contrast enhancement, both full-thickness and stratified. Mucosal layer appears hyper-intense in T2W fat saturated images and shows cobblestone appearance due to ulcers and pseudo-polyps. Multiple lymph nodes, comb sign, and mesenteric edema may be present; 2. Fibrostenotic Subtype: Chronic fibrotic walls are typically hypo-intense on both T1W and T2W images, with inhomogeneous enhancement. Mesenteric edema, lymph nodes, and comb sign are usually not present; 3. Fistulizing Subtype is defined when sinus or fistulas occur, usually together with active inflammation features. Fibrostenotic subtype can be overlapped as well. When more than one pattern is simultaneously present, the radiologists assign the more severe pattern [10]. Regarding the activity, it is defined on the basis of the total summa of the following features: wall thickening greater than 4 mm, intramural and mesenteric edema, mucosal hyperemia, wall enhancement (and enhancement pattern), transmural ulceration and fistula formation, vascular engorgement, and inflammatory mesenteric lymph nodes [11]. Therefore, activity and pattern values strongly depend on the subjective evaluation of the radiologist and are subject to great variability.

For this reason, in this work we have excluded activity and pattern from the feature set and considered only the remaining 20 parameters so as to avoid operator dependence. Furthermore, in this paper, the CD classification method has been refined by feature space reduction techniques, tested on a larger dataset (800 patients instead of 300) and evaluated by a performance comparison with the conventional classification methods. A team of radiology experts has extracted from each E-MRI image the vector composed of the typical E-MRI features associated with CD-affected patients. Based on the histological specimen results, which are the ground-truth for CD diagnosis, each patient has been classified as either positive or negative with respect to CD. The 800 observations have been then used to train and validate different classifiers. Experimental results have demonstrated that the SVM-based method shows a better performance with respect to the other standard classification algorithms. Different kernels have been compared, while  $K$ -fold cross-validation has been used to improve the classifier reliability. Feature space reduction techniques, such as principal component analysis (PCA), have been applied to reduce feature space dimensionality. The following indexes have been measured using 20 features and a 15-fold cross-validation scheme: sensitivity: 97.07%, specificity: 96.04%, negative predictive value: 97.24%, precision: 95.80%, accuracy: 96.54%, and error: 3.46%. These results are better than the manual reference methods reported in literature [9].

The rest of the paper is organized as follows: Sect. 29.2 describes the dataset used to train and validate the SVM classifier, while the proposed classification method is presented in Sect. 29.3. Section 29.4 outlines the experimental results and Sect. 29.5 concludes the paper.

## 29.2 Materials

The proposed classification method has been tested and evaluated using a real dataset containing medical data related to 800 patients from the University of Palermo Policlinico Hospital. The dataset is composed of 800 magnetic resonance (MR) enterography examinations of 800 patients (427 females, 373 males, mean age 30.1 years): 400 patients with histologically proven Crohn's disease and 400 healthy patients. The following sequences have been used: DWI—axial plane, HASTE thick slab—coronal plane, Single Shot Fast Spin Echo (T2 SPAIR)—axial and coronal planes, Contrast 3D Spoiled Gradient Echo (T1 e-Thrive)—axial and coronal planes, Steady State Free Precession (BFFE)—axial and coronal planes. Starting from this dataset, a team of radiology experts has extracted for each patient a set of features that have been proven to be associated with Crohn's disease [2, 10, 11].

## 29.3 Methods

This paper proposes the evaluation of an automated tool for the classification of CD-affected patients that uses a supervised machine learning method based on a Support Vector Machine.

### 29.3.1 Classification Methods Comparison

First, the study presented in [18] related to a dataset of 300 patients has been extended to a larger dataset composed of 800 patients. Different classification methods, namely Support Vector Machines (SVM),  $K$ -Nearest Neighbor (KNN), Naïve Bayes (NB), and Feed-Forward Neural Network (FFNN) have been compared and evaluated. These classifiers have been trained and validated using the dataset presented in Sect. 29.2, composed of 800 observations each containing the 22 features used in [18]. The comparison results, shown in Table 29.1, demonstrate that the SVM-based classification technique achieves a better performance with respect to the other three methods. Based on these results, the SVM-based method has been chosen for the classification of CD-affected patients. The following subsections describe the different phases of the classification process as well as the methods used in each phase.

**Table 29.1** Comparison of different classification methods: Support Vector Machine (SVM),  $K$ -Nearest Neighbor (KNN), Naïve Bayes (NB), Feed-Forward Neural Network (FFNN)

	Classification method			
	SVM (%)	KNN ( $K = 10$ ) (%)	NB (%)	FFNN (%)
True Positive (TP)	46.46	46.21	46.34	47.43
True Negative (TN)	50.70	49.80	49.29	48.71
False Positive (FP)	1.02	1.92	2.43	0.64
False Negative (FN)	1.79	2.05	1.92	3.20
Sensitivity (Eq. 29.1)	96.27	95.74	96.01	93.67
Specificity (Eq. 29.2)	98.01	96.27	95.28	98.70
Negative predictive value (Eq. 29.3)	96.57	96.03	96.24	93.82
Precision (Eq. 29.4)	97.83	96.00	95.00	98.66
Accuracy (Eq. 29.5)	97.18	96.03	95.64	96.16
Error (Eq. 29.6)	2.82	3.97	4.36	3.84

### 29.3.2 Feature Extraction

For each E-MRI image, a set of 20 parameters, based on the typical E-MRI features associated with CD-affected patients [2, 10, 11], have been derived by expert radiologists. Starting from the 22 parameters considered in [18], we have excluded the composite features activity and pattern, which require great radiology expertise to be derived, and have used only the remaining 20 features. This choice allowed us to avoid operator dependence and to exclude two features that are subject to great variability among different teams of radiologists since they strongly depend on the radiologist subjective evaluation. The 20 considered features are listed in Table 29.2.

### 29.3.3 Feature Reduction Techniques

The extracted features will be used as the predictive variables to train and test the SVM model. Reducing the number of predictors can have significant benefits on computational time and memory consumption. Furthermore, a reduced number of predictors results in a simpler model that is easier to interpret and can be generalized. Automated methods for feature space dimensionality reduction can find noisy or highly correlated predictive variables. These methods include feature transformation methods, such as principal component analysis (PCA) [17], which transform the coordinate space of the observed variables, and feature selection methods [19], which choose a subset of the observed variables to be included in the model. PCA transforms an n-dimensional feature space into a new n-dimensional space of orthogonal components. The principal components are ordered by the variation explained in the data. PCA can therefore be used for dimensionality reduction by discarding the components beyond a chosen threshold of explained variance. We have applied both the PCA technique and a sequential forward feature selection algorithm to reduce the number of predictors. The sequential forward feature selection algorithm selects the subset of predictors by incrementally adding predictors to the model as long as the prediction error is decreasing. Different SVM models with different numbers of predictors have been trained, validated, and compared. Results are reported in Sect. 29.4.

### 29.3.4 Support Vector Machines

Support Vector Machines (SVM) perform data classification by finding the best hyperplane that separates all data points [20, 21]. This optimization problem consists in finding the boundary that is as far as possible from any of the observations, namely maximizing the margin, i.e., the distance between the boundary and the nearest

**Table 29.2** Features extracted from Enterography MR images

Sequence types	Features	Values
DWI	Water diffusion restriction (DWI)	0: Free and physiological diffusion, no hyper-intensity on DWI; 1: Mild hyper-intensity; 2: Severe hyper-intensity
HASTE thick slab	Bowel cleaning protocol	0: No sufficient preparation; 1: Adequate bowel loops cleaning and distention
	Bowel distention protocol	0: No sufficient bowel distention, only stomach or duodeno-jejunal loop; 1: PEG has reached the ileocecal junction
T2 SPAIR	Lumen	0: No changes in lumen caliber; 1: Stenosis/sub-stenosis
	Pseudo-polyps	0: No; 1: yes
	T2W imaging	0: No mural edema; 1: Mild mural edema; 2: Severe hyper-intensity due to noticeable edema
Post-contrast T1 e-Thrive	Breathing/peristalsis artifacts	0: No; 1: yes
	Lymph nodes	0: No; 1: yes
	Post-contrast T1 imaging	0: No wall enhancement; 1: Layered enhancement; 2: Transmural enhancement
T2 SPAIR/post-contrast T1 e-Thrive/BFFE	Complications	0: No; 1: yes
	Fat wrapping	0: Normal mesenteric adipose tissue; 1: Mesenteric hyperplasia
	Fistulas	0: No; 1: yes
	Free fluid	0: No; 1: yes
	Intestinal obstruction	0: No; 1: yes
	Length	Length (in cm) of the affected gastrointestinal tract/tracts
	Mucosal layer	0: No mucosal involvement; 1: edema and post-contrast enhancement; 2: Inflammatory changes and pseudo-polyps
	Single lesion/skip lesions	1: Single tract involved; 2: Multiple tracts involved
	Sinus	0: No; 1: yes

(continued)

**Table 29.2** (continued)

Sequence types	Features	Values
	Surgery	0: No; 1: yes
	Terminal ileum thickening	0: None or less than 3 mm thickening; 1: Thickening greater than 3 mm

observations (support vectors). Since real noisy data can be not linearly separable, the optimization problem is modified to maximize the margin, but with a penalty term for misclassified observations. This is reduced to a quadratic optimization problem that is solved by quadratic programming. Many classification problems are not linearly separable in the space of the input data, but they might be separable in a higher-dimensionality feature space. Support Vector Machines can be still used for non-linear classification problems by applying appropriate kernel functions that map the input data into a higher-dimensional space where the classes are linearly separable. Three common kernels, namely linear, Gaussian, and polynomial, have been used in this work.

### 29.3.5 K-Fold Cross-Validation

Cross-validation can provide a more robust and reliable estimate of the classification accuracy [22]. To evaluate a classification model performance, input data are divided into training and test data. The classification model is first fitted to the training data and then validated using test data. The learning algorithm accuracy is therefore calculated for that specific test data. The classifier could not generalize well to other data. To solve this problem,  $K$ -fold cross-validation randomly divides input data into  $K$  sets or folds. The training and testing process is repeated  $K$  times, each time reserving a different fold for the testing and using the rest of the data for the training. The average error from all the folds is the overall  $K$ -fold error. When  $K$  is equal to the number of observations, then a single observation is used each time for validation. This is known as leave-one-out cross-validation. As described in Sect. 29.4, a cross-validation strategy has been integrated into the proposed SVM classifier in order to obtain a more robust estimation of its generalization capabilities.

## 29.4 Results and Discussions

The classification method described in Sect. 29.3 has been evaluated using the medical dataset presented in Sect. 29.2.

### 29.4.1 Performance Evaluation

The following standard metrics have been used to measure the classifier performance.

*Sensitivity* measures the percentage of actual positives that are correctly classified:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (29.1)$$

*Specificity* measures the percentage of actual negatives that are correctly classified:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (29.2)$$

*Negative Predictive Value* measures the probability that subjects classified as negatives truly do not have the disease:

$$\text{Negative Predictive Value} = \frac{\text{TN}}{\text{TN} + \text{FN}} \quad (29.3)$$

*Precision* measures the probability that subjects classified as positives truly have the disease:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (29.4)$$

*Accuracy* measures the percentage of correctly classified cases among the total number of cases examined:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (29.5)$$

*Error* measures the percentage of misclassified cases among the total number of cases examined:

$$\text{Error} = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (29.6)$$

Table 29.3 lists classification results obtained using the linear SVM kernel for four different cross-validation schemes, while the comparison of three different SVM kernels with 15-fold cross-validation is reported in Table 29.4. The best results are obtained by the linear kernel SVM with 15-fold cross-validation, which achieves an accuracy of 96.54% with an error of 3.46%.

**Table 29.3** Linear SVM classifier with  $K$ -fold cross-validation: comparison for different values of  $K$  (the classification method has been applied to the dataset composed of 800 vectors each containing the 20 features listed in Table 29.2)

	Linear SVM			
	K-fold cross-validation (%)			Leave-one-out cross-validation (%)
	$K = 5$	$K = 10$	$K = 15$	
True Positive (TP)	46.54	46.67	46.79	46.54
True Negative (TN)	49.74	49.61	49.74	49.74
False Positive (FP)	2.05	2.17	2.05	2.05
False Negative (FN)	1.66	1.53	1.41	1.66
Sensitivity (Eq. 29.1)	96.54	96.81	97.07	96.54
Specificity (Eq. 29.2)	96.04	95.79	96.04	96.04
Negative predictive value (Eq. 29.3)	96.76	97.00	97.24	96.76
Precision (Eq. 29.4)	95.78	95.54	95.80	95.78
Accuracy (Eq. 29.5)	96.29	96.29	96.54	96.29
Error (Eq. 29.6)	3.71	3.71	3.46	3.71

**Table 29.4** SVM classifier with 15-fold cross-validation: comparison of 3 different kernels (the classification method has been applied to the dataset composed of 800 vectors each containing the 20 features listed in Table 29.2)

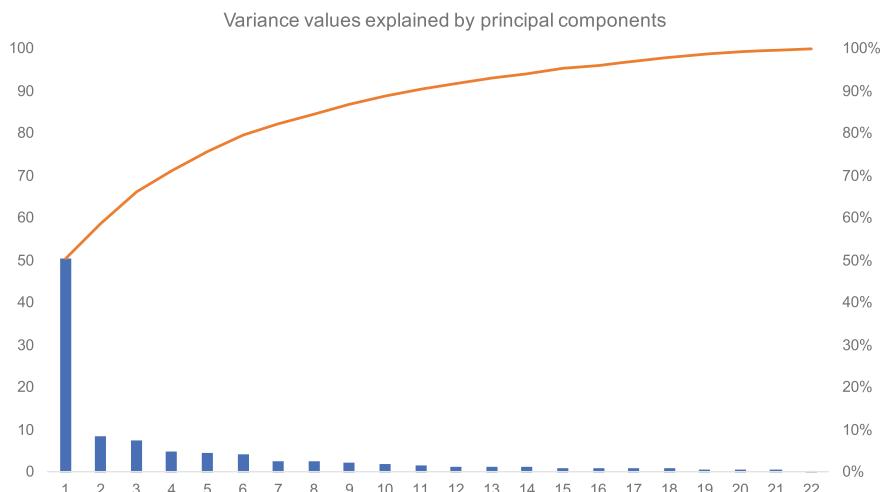
	15-Fold Cross-Validation		
	SVM kernel		
	Linear (%)	Gaussian (%)	Polynomial (%)
True Positive (TP)	46.79	46.79	46.15
True Negative (TN)	49.74	49.36	49.61
False Positive (FP)	2.05	2.43	2.17
False Negative (FN)	1.41	1.41	2.05
Sensitivity (Eq. 29.1)	97.07	97.07	95.74
Specificity (Eq. 29.2)	96.04	95.30	95.79
Negative predictive value (Eq. 29.3)	97.24	97.22	96.03
Precision (eq. 29.4)	95.80	95.05	95.49
Accuracy (Eq. 29.5)	96.54	96.16	95.77
Error (Eq. 29.6)	3.46	3.84	4.23

### 29.4.2 Feature Space Reduction Techniques

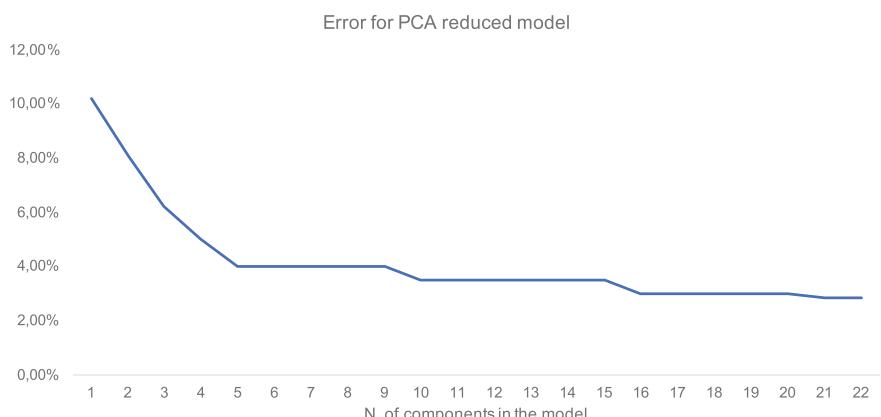
Different feature space dimensionality reduction methods have been applied.

#### Principal Component Analysis (PCA)

The PCA technique described in Sect. 29.3 has been applied to the original dataset composed of the 22 features of [18]. Figure 29.1 reports the variance values explained by principal components, while Fig. 29.2 depicts the classification error for different PCA reduced SVM models that use a different number of principal components. The first 16 principal components explain approximately 97% of the variance and



**Fig. 29.1** Principal component analysis



**Fig. 29.2** Error for different PCA reduced SVM models

**Table 29.5** Sequential feature selection results

10 selected features	SVM reduced model error
Activity, bowel distention protocol, fat wrapping, length, lumen, lymph nodes, pattern, post-contrast t1 imaging, single lesion/skip lesions, T2W imaging	3.33%

the corresponding SVM model, obtained by discarding the components beyond this threshold, shows an error of 2.97%, comparable to the complete model error (2.82%). It can be also observed that the PCA reduced model based on the first 5 principal components shows a classification error of about 4%.

### Sequential Feature Selection (SFS)

A sequential feature selection algorithm, as described in Sect. 29.3, has been also used to select the features that are the most important for creating an accurate model. Table 29.5 shows the 10 selected features and the related classification error of the reduced SVM model.

#### 29.4.3 Radiologist Driven Reduction Techniques

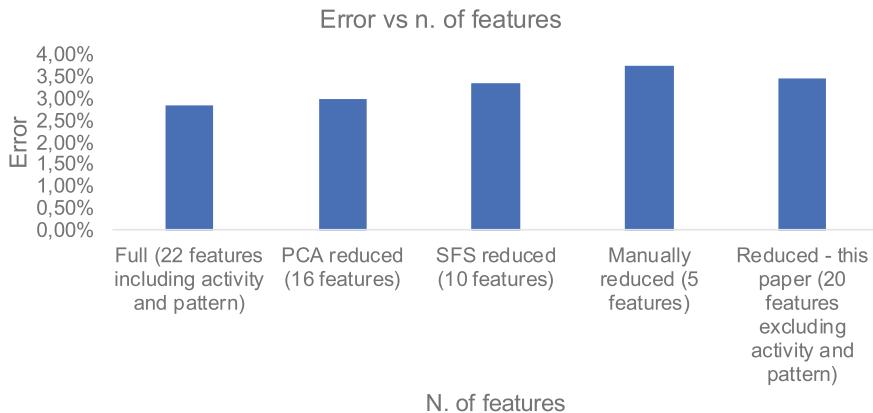
Based on the radiologist expertise, the 5 features that have the higher importance have been manually selected and the related error of the reduced SVM model has been measured. Results are reported in Table 29.6.

The latter result confirms that activity and pattern are the two most important and significant predictors. It can be observed that the 5 selected predictors shown in Table 29.6, which include activity and pattern, allow to achieve a classification error (3.72%) comparable to that achieved using the 20 predictors other than activity and pattern listed in Table 29.2 (3.46%). We can therefore conclude that the 20 predictors excluding activity and pattern still give an acceptable classification accuracy allowing us at the same time to avoid the operator dependence and variability given by these two composite features.

Figure 29.3 reports the errors measured for the five different SVM classifier models: the full model that uses the 22 features including activity and pattern considered in [18], the PCA reduced model that uses the first 16 principal components, the sequential feature selection (SFS) reduced model that uses the 10 features reported in Table 29.5, the manually reduced model that uses the 5 features manually selected by the radiologists and reported in Table 29.6, and finally, the model based on the 20 features excluding activity and pattern reported in Table 29.2 and used in this work.

**Table 29.6** Most important feature selection based on the radiologist expertise

5 selected features	SVM reduced model error
Activity, DWI, lymph nodes, pattern, T2W imaging	3.72%



**Fig. 29.3** Comparison of different SVM models: full model (based on the 22 features including activity and pattern), PCA reduced model (based on the first 16 principal components), sequential feature selection (SFS) reduced model (based on the 10 features of Table 29.5), manually reduced model (based on the 5 features of Table 29.6), and reduced model (based on the 20 features excluding activity and pattern used in this work)

## 29.5 Conclusions

The evaluation of an SVM-based method for classifying Crohn's disease affected patients has been presented. A real dataset composed of E-MRI images of 800 patients from the University of Palermo Policlinico Hospital has been used to evaluate the proposed method. A team of radiology experts has extracted for each patient two vectors: The first composed of 22 parameters and the second composed of 20 parameters. Principal component analysis and feature selection techniques have been applied to verify the possibility to reduce feature space dimensionality, while a  $K$ -fold cross-validation strategy has been integrated into the classifier to better measure results accuracy. Classification results have been compared with the histological specimen results, which are the adopted clinical ground-truth for CD diagnosis. It has been proven that the proposed SVM-based classification method outperforms the main standard classification techniques. Furthermore, the performance metrics measured using 20 parameters and a 15-fold cross-validation (sensitivity: 97.07%, specificity: 96.04%, negative predictive value: 97.24%, precision: 95.80%, accuracy: 96.54%, error: 3.46%) are better than the operator-based reference values reported in literature [9], namely sensitivity: 93% and specificity: 90%.

Future work will focus on the design of a multi-class SVM classifier able to not only detect the presence/absence of the Crohn's disease, but also grade the disease activity and classify patients into different classes according to the disease activity level (mild, moderate, or severe).

## References

1. Bhatnagar, G., Stempel, C., Halligan, S., Taylor, S.A.: Utility of MR enterography and ultrasound for the investigation of small bowel CD. *J. Magn. Reson. Imaging* **45**, 1573–1588 (2016)
2. Lo Re, G., Midiri, M.: Crohn's disease: radiological features and clinical-surgical correlations. Springer, Heidelberg (2016)
3. Maglinte, D.D., Gourtsoyiannis, N., Rex, D., Howard, T.J., Kelvin, F.M.: Classification of small bowel Crohn's subtypes based on multimodality imaging. *Radiol. Clin. North Am.* **41**(2), 285–303 (2003)
4. Gomollón, F., Dignass, A., Annese, V., Tilg, H., Van Assche, G., Lindsay, J.O., Peyrin-Biroulet, L., Cullen, G.J., Daperno, M., Kucharzik, T., et al.: 3rd European evidence-based consensus on the diagnosis and management of Crohn's disease 2016: part 1: diagnosis and medical management. *J. Crohns Colitis* **11**, 3–25 (2016)
5. Sinha, R., Verma, R., Verma, S., Rajesh, A.: Mr enterography of Crohn disease: part 1, rationale, technique, and pitfalls. *Am. J. Roentgenol.* **197**(1), 76–79 (2011)
6. Peloquin, J.M., Pardi, D.S., Sandborn, W.J., Fletcher, J.G., McCollough, C.H., Schueler, B.A., Kofler, J.A., Enders, F.T., Achenbach, S.J., Loftus, E.V.: Diagnostic ionizing radiation exposure in a population-based cohort of patients with inflammatory bowel disease. *Am. J. Gastroenterol.* **103**(8), 2015–2022 (2008)
7. Lo Re, G., Cappello, M., Tudisca, C., Galia, M., Randazzo, C., Craxi, A., Camma, C., Giavagnoni, A., Midiri, M.: CT enterography as a powerful tool for the evaluation of inflammatory activity in Crohn's disease: relationship of CT findings with CDAI and acute-phase reactants. *Radiol. Med. (Torino)* **119**(9), 658–666 (2014)
8. Steward, M.J., Punwani, S., Proctor, I., Adjei-Gyamfi, Y., Chatterjee, F., Bloom, S., Novelli, M., Halligan, S., Rodriguez-Justo, M., Taylor, S.A.: Non-perforating small bowel CD assessed by MRI enterography: derivation and histopathological validation of an MR-based activity index. *Eur. J. Radiol.* **81**(9), 2080–2088 (2012)
9. Panes, J., Bouzas, R., Chaparro, M., García-Sánchez, V., Gisbert, J., Martínez de Guereñu, B., Mendoza, J.L., Paredes, J.M., Quiroga, S., Ripollés, T., et al.: Systematic review: the use of ultrasonography, computed tomography and magnetic resonance imaging for the diagnosis, assessment of activity and abdominal complications of Crohn's disease. *Aliment. Pharmacol. Ther.* **34**(2), 125–145 (2011)
10. Sinha, R., Verma, R., Verma, S., Rajesh, A.: Mr enterography of Crohn disease: part 2, imaging and pathologic findings. *Am. J. Roentgenol.* **197**(1), 80–85 (2011)
11. Tolan, D.J., Greenhalgh, R., Zealley, I.A., Halligan, S., Taylor, S.A.: Mr enterographic manifestations of small bowel Crohn disease 1. *Radiographics* **30**(2), 367–384 (2010)
12. Cocosco, C.A., Zijdenbos, A.P., Evans, A.C.: A fully automatic and robust brain MRI tissue classification method. *Med. Image Anal.* **7**(4), 513–527 (2003)
13. Chaplot, S., Patnaik, L., Jagannathan, N.: Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomed. Signal Process. Control* **1**(1), 86–92 (2006)
14. Agnello, L., Comelli, A., Ardizzone, E., Vitabile, S.: Unsupervised tissue classification of brain MR images for voxel-based morphometry analysis. *Int. J. Imaging Syst. Technol.* **26**(2), 136–150 (2016)
15. Son, Y.J., Kim, H.G., Kim, E.H., Choi, S., Lee, S.K.: Application of support vector machine for prediction of medication adherence in heart failure patients. *Healthc. Inform. Res.* **16**(4), 253–259 (2010)
16. Zhang, Y., Wang, S., Ji, G., Dong, Z.: An MR brain images classifier system via particle swarm optimization and Kernel support vector machine. *Sci. World J.* **2013**, 9 (2013)
17. Jolliffe, I.T.: Principal Component Analysis, 2nd edn. Springer, New York (2002)
18. Comelli, A., Terranova, M. C., Scopelliti, L., Salerno, S., Midiri, F., Lo Re, G., Petrucci, G., Vitabile, S.: A kernel support vector machine based technique for Crohn's disease classification in human patients. In: Barolli, L., Terzo, O. (eds.) Complex, Intelligent, and Software Intensive

- Systems. CISIS 2017. Advances in Intelligent Systems and Computing, vol 611. Springer, Cham (2018)
19. Jain, A., Zongker, D.: Feature selection: evaluation, application, and small sample performance. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(2), 153–158 (1997)
  20. Christianini, N., Shawe-Taylor, J.C.: An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, Cambridge, UK (2000)
  21. Scholkopf, B., Smola, A.: Learning with kernels: support vector machines, regularization, optimization and beyond, adaptive computation and machine learning. The MIT Press, Cambridge, MA (2002)
  22. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer, New York (2001)

**Part IV**

**Dynamics of Signal Exchanges**

# Chapter 30

## Seniors' Appreciation of Humanoid Robots



**Anna Esposito, Marialucia Cuciniello, Terry Amorese,  
Antonietta M. Esposito, Alda Troncone, Mauro N. Maldonato, Carl Vogel,  
Nikolaos Bourbakis and Gennaro Cordasco**

**Abstract** This paper is positioned inside a research project investigating elders' preferences and acceptance toward robots, in order to collect insights for the design and implementation of socially assistive robots. To this aim, short video clips of five manufactured robots (Roomba, Nao, Pepper, Ishiguro, and Erica) were shown to 100 seniors (50 Female) aged 65+ years (average age: 71.34 years, DS:  $\pm 5.60$ ). After watching each robot video clip, seniors were administered a short questionnaire assessing their willingness to interact with robots, feelings robots aroused, and duties they would entrust to robots. The questionnaire's scores were assessed through repeated measures ANOVA in order to ascertain statistically significant differences among seniors' preferences. A clear uncanny valley effect was identified. The robot Pepper received significantly higher scores than Roomba, Nao, Ishiguro, and Erica on communication skills, ability to remind friendly and pleasant memories, comprehension, and ability to provide emotional support. In addition, Pepper was considered the most suitable, among the five proposed robots, in performing welfare duties for elders, children and disabled, protection and security, and front desk occupations.

---

A. Esposito (✉) · M. Cuciniello · T. Amorese · A. Troncone · G. Cordasco  
Dipartimento di Psicologia, Università della Campania "Luigi Vanvitelli," and IIASS,  
Caserta, Italy  
e-mail: [iiass.annaesp@tin.it](mailto:iiass.annaesp@tin.it)

A. M. Esposito  
Istituto Nazionale di Geofisica e Vulcanologia,  
Sez. di Napoli Osservatorio Vesuviano, Napoli, Italy

M. N. Maldonato  
Dipartimento di Neuroscienze, Università di Napoli "Federico II", Napoli, Italy

C. Vogel  
School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland

N. Bourbakis  
Department of Comp. Science & Eng., WSU, Pullman, OH, USA

### 30.1 Introduction

Given the foreseen aging of the European populations, it has become incumbent on society to improve the effectiveness of healthcare systems in providing social assistance and continuous monitoring of elderly physical and cognitive quality of life in order to set up prevention measures and timely treatments, provide assistance and rehabilitation tools, and concurrently lighten the costs for social care.

To this aim, robotic and virtual assistive technologies have been considered. In particular, several robotic assistants have been proposed for supporting elder's caregivers and relatives in their daily assistance routines. However, pursuing this, researches have neglected to investigate to which degree elders accept to undertake an interaction with such robots.

For elders, accepting to be assisted by robots requires cognitive loads related to the difficulties to understand their functioning and adaptive efforts to allocate these resources in their personal environments. In addition, it has been observed that robots' appearance plays a fundamental role in terms of their acceptance [11]. To be accepted, robots must be reassuring, friendly, pleasant, able to interpret elders' emotional and social behavior, and must establish with them trustworthy relationships, supporting their health, and simplifying their access to telemedicine and tele-care support services [5, 4]. Acceptance to use a robot on a daily basis is a complex concept. It requires: (a) strong users' motivations; (b) accessibility and easiness of use; (c) trustworthiness; (d) comfortable and reassuring appearance of robots and (e) robot's ability to interpret individuals' social rules and cognitive competencies [2, 3, 6, 8, 10, 12]. Users can experience the "uncanny valley" effect [9], a situation in which a seemingly "intelligent" humanlike artifact trigger feelings of oddness and repugnance—the exact opposite of acceptance—generating an interactional failure. The "uncanny valley" effect manifests itself when a person is faced with human-like artifacts characterized by a certain percentage of resemblance to human beings. Familiarity with the stimulus and probably positive feelings connected to it increase up to a certain point and then begin to decrease dramatically as the percentage of human likeness increases, associated with long-lasting feelings of discomfort and strangeness. Wang et al. [13] observed that the uncanny feeling is a multifaceted construct that cannot be reduced neither to a negative emotional response aroused by the excessive resemblance of the artifact to human semblances [14], nor to the emergence of an innate survival instinct in response to a threat, being the artifact deceptive and then, perceived as dangerous or harmful [7]. Rather, the uncanny valley effect is due to a "*dehumanization process*" triggered by the detection of mechanistic features in artifacts assuming humanlike resemblances. "*The more human observers attribute humanlike characteristics to (i.e., anthropomorphize) a human replica, the more likely detecting its mechanistic features triggers the dehumanization process that would lead to the uncanny feeling*" [13], first column). The proposed research suggests that, as long as, important human features such as voice, body movements, and facial expressions cannot be appropriately rendered in robots, a *dehumanization process* is observed leading to the uncanny valley effect [1]. In order to assess

these effects, the present research investigates elders' willingness to interact with five manufactured robots (Roomba, Nao, Pepper, Ishiguro, and Erica), evaluating their perception of robot's human likeness, feelings aroused, and mansions entrusted to robots. It is essential to highlight that while the concept of appearance is very complex and involves many physical and social factors, the proposed research restricts its investigation to the level of human likeness displayed by robots.

## 30.2 Materials and Methods

### 30.2.1 Stimuli

Five mute video clips depicting five well-known robots were exploited. The videos were downloaded by "YouTube" search engine and edited in order to create short clips of approximately 40-s duration. The selected robots were Roomba, Pepper, Nao, Erica, and Ishiguro (Fig. 30.1). *Roomba* is vacuum cleaner robot ([www.irobot.it/roomba/](http://www.irobot.it/roomba/)), designed to help people in daily household cleaning; *Nao* ([www.softbankrobotics.com/emea/en/nao](http://www.softbankrobotics.com/emea/en/nao)) and *Pepper* are humanoid robots proposed by SoftBank Robotics ([www.softbankrobotics.com/emea/en/robots/pepper](http://www.softbankrobotics.com/emea/en/robots/pepper)) with the aim of assisting people in their daily life. *Ishiguro* is the exact copy of Ishiguro Hiroshi (<http://www.geminoid.jp/en/index.html>), professor of artificial intelligence at University of Osaka, Japan. Hiroshi Ishiguro is also the creator of *Erica*, a robot capable of holding a conversation with humans while moving her facial features, neck, shoulders, and waist, with very closely resembling human movements. As previously stated, and as it appears evident from Fig. 30.1, the five proposed robots show different degrees of similarity to human beings. Roomba is a non-humanoid robot, the one among the five that is less reminiscent of human features. Nao and Pepper are humanoid robots, characterized by features vaguely resembling humans. Ishiguro and Erica are android robots showing extreme levels of human likeness in terms of shape, body movements, and facial expressions.

### 30.2.2 Participants

The experiments involved 100 participants (50 females) all aged 65+ years (mean age = 71.34, SD =  $\pm 5.60$ ) recruited in Campania, a region in the south of Italy. Participants reported vision corrected with glasses, several chronic diseases (such as prostatitis, benign tumor, esophageal reflux, diabetes, and arterial hypertension) but no psychological disorders. Participants joined the experiment on a voluntary basis and signed an informed consent formulated in accord with the privacy and data protection procedures established by the current Italian and European laws. The



**Fig. 30.1** Five manufactured robots exploited for the proposed investigation

experiment was approved by the ethical committee of the Università della Campania “Luigi Vanvitelli,” Department of Psychology, Code Number 25/2017.

### 30.2.3 Tools and Procedures

Participants were asked to watch each robot's video clip and immediately after complete a questionnaire. The questionnaire was divided into three clusters. The first cluster comprised two single items aimed at evaluating, respectively, seniors' willingness to interact with each robot, and robot's human likeness ("appearance"). The second cluster, named "impressions," was constituted by six items, aimed to assess feelings aroused by robots in terms of (1) communication skills, (2) ability to arouse memories of someone or something, (3) efficiency, (4) language comprehension, (5) reliability, and (6) ability to provide emotional support. These items were assessed both as a whole under the label "impressions" and singularly. The last cluster investigated occupations seniors would entrust to robots among welfare, housework, security/protection, and front office jobs. Each questionnaire item was rated on a 7-point Likert scale ranging from 1 = not at all to 7 = very much. Participants were initially asked to read and sign an informed consent and fill in a datasheet providing information about their age, gender, educational level, and health problems. Subsequently, they were asked to watch each robot video clip and immediately after fill in the questionnaire. Participants were grouped into two groups. A first group of 50 subjects (25 males and 25 females) saw the video clips of Roomba, Pepper, Nao, and Ishiguro, while the second group (50 subjects, 25 males and 25 females) saw the video clips of Roomba, Pepper, Nao, and Erica. The experiment was conducted in participants' private dwellings. In order to avoid visual interference, a laptop with a bright, non-reflective screen was used. Participants were positioned on a chair next to a table in a quiet room. The experiment lasted approximatively 15 min.

## 30.3 Analysis and Results

Several repeated measures ANOVA, using the IBM SPSS (Statistical Package for Social Science) software, were conducted on the questionnaire's scores in order to assess participants' preferences toward robots in terms of willingness of interact ("interaction"), degree of human likeness ("appearance"), "impressions" (both including as a whole the six items mentioned in Sect. 30.2.3, and considering each item singularly) and "Occupations" entrusted to robots. Repeated measures ANOVA was carried out considering participants (and only for Ishiguro and Erica robots' gender) as between factors. The scores obtained by each robot at items evaluating willingness to interact "interaction," human likeness "appearance," and "impressions," as well as those obtained for each occupation entrusted to robots, were considered as within factors.

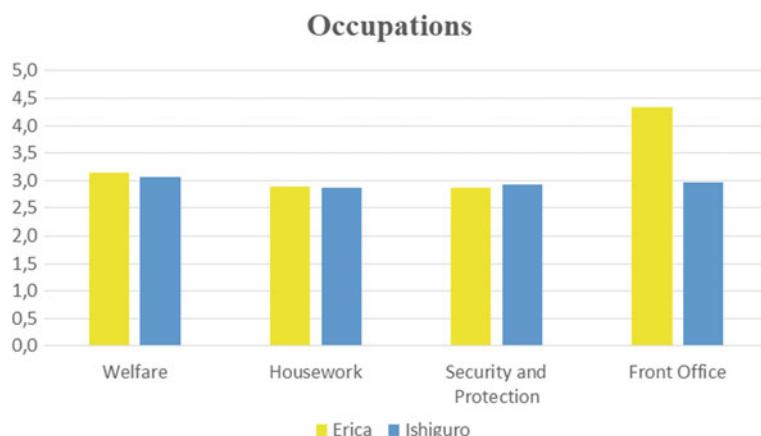
The following sections compare the scores obtained, first for the android robots Ishiguro and Erica and then for Roomba, Nao and Pepper. Since Pepper emerged as being favorite with respect to Roomba and Nao, the scores obtained by Pepper were compared with those obtained by Ishiguro and Erica, respectively.

### *Ishiguro versus Erica*

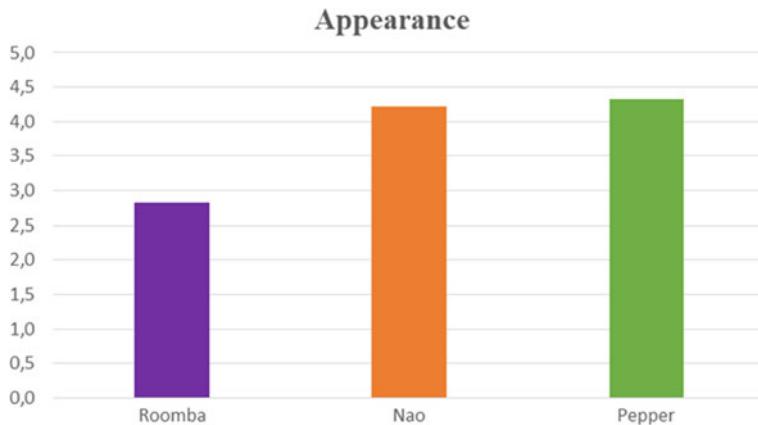
No significant differences were found both for participants and robot gender (Ishiguro and Erica), nor for the within factors interaction, appearance, and impressions (neither as a whole or as for each single item constituting the cluster). These results suggested that none of the two robots resulted favorite by elders. Significant differences were found among occupations seniors would entrust to robots ( $F(3.294) = 7.261 p \ll 0.01$ ), with welfare (mean = 3.310) and front office (mean = 3.658) considered largely more suitable for them than protection/security (mean = 3.000) and housework (mean = 3.050). A significant interaction emerged between robot gender and occupations ( $F(3.294) = 5.869 p < 0.01$ ). Bonferroni post hoc tests showed that Erica (mean = 4.340) was considered significantly more suitable than Ishiguro (mean = 2.960) for front office occupations (see Fig. 30.2). No significant differences between the two android robots emerged for welfare, housework, and security/protection.

### *Roomba versus Nao versus Pepper*

Unlike Ishiguro and Erica, the whole sample of participants (100 subjects) watched the video clips of Roomba, Nao, and Pepper. Repeated measures ANOVA was conducted considering participants' gender as a between factor and "interaction," "appearance," "impressions," and "Occupations" as within factors. No participants' gender effect was found. Significant differences emerged for the within factors. Detailed repeated measures ANOVA were performed on the single variables willingness to interact, appearance, and impressions. No significant differences among robots (Roomba mean = 4.080, Nao mean = 4.450, Pepper mean = 4.430) were observed for the elders' willingness to interact with them ( $F(2.196) = 2.230, p = 0.110$ ). Pepper's appearance (Pepper mean = 4.320, see Fig. 30.3) was rated significantly better ( $F(2.196) = 50.946, p \ll 0.01$ ) than Roomba (Roomba mean =

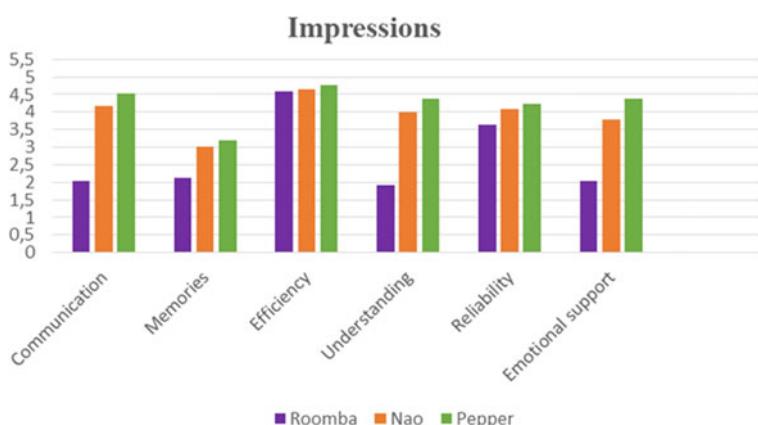


**Fig. 30.2** Scores obtained for occupations entrusted to Erica and Ishiguro. Erica was considered more able to perform front office tasks



**Fig. 30.3** Preferences expressed by elders for the variable “appearance”

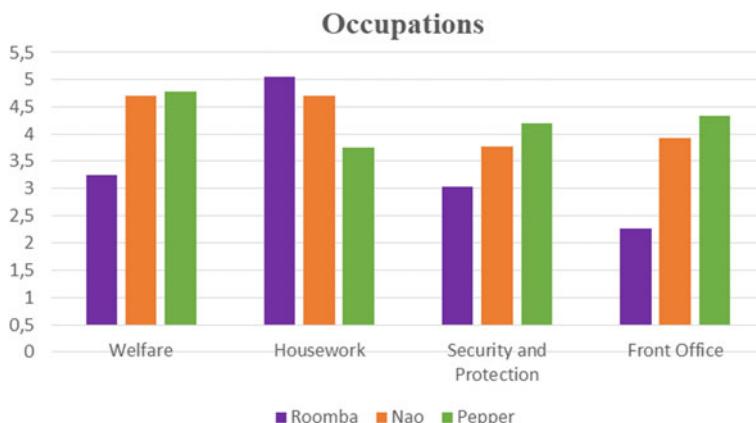
$2.820, p \ll 0.01$ ) and Nao (Nao mean = 4.220,  $p \ll 0.01$ ). Pepper was also able to arouse significantly overall more positive (Pepper mean = 25.480) “impressions” ( $F(2.196) = 82.935, p \ll 0.01$ ) than Roomba (Roomba mean = 16.380,  $p \ll 0.01$ ) and Nao (Nao mean = 23.700,  $p \ll 0.01$ ). Nao was considered slightly significantly more humanlike than Roomba ( $p = 0.017$ ). Detailed analyses concerning items constituting the cluster “impressions” revealed that Pepper (Pepper mean = 4.530, see Fig. 30.4) was considered significantly more communicative ( $F(2.196) = 97.338, p \ll 0.01$ ) than Roomba (Roomba mean = 2.050,  $p \ll 0.01$ ) and Nao (Nao mean = 4.160,  $p = 0.038$ ), more able to arouse memories of someone or something ( $F(2.196) = 21.146, p \ll 0.01$ ) than Roomba ( $p \ll 0.01$ ) (Roomba mean = 2.130, Nao mean = 3.010, Pepper mean score = 3.200), more able to understand participants’ language



**Fig. 30.4** Impressions aroused by Roomba, Nao, and Pepper in terms of communication, memories, efficiency, ability to understand language, reliability, and providing emotional support

( $F(2.196) = 109.712, p \ll 0.01$ ) than Roomba ( $p \ll 0.01$ ) (Roomba mean = 1.910, Nao mean = 4.000, Pepper mean = 4.380), slightly more reliable ( $F(2.196) = 5.106, p = 0.007$ ) than Roomba ( $p = 0.028$ ) (Roomba mean = 3.850, Nao mean = 4.080, Pepper mean = 4.220), and more able to provide emotional support ( $F(2.196) = 75.982, p \ll 0.01$ ) than Roomba ( $p \ll 0.01$ ) and Nao ( $p \ll 0.01$ ) (Roomba mean = 2.050, Nao mean = 3.790, Pepper mean = 4.380). In addition, Nao was considered significantly more communicative, more able to arouse memories of something, more able to understand participant language, and more able to provide emotional support than Roomba ( $p \ll 0.01$ ). These results proved that the two android robots were preferred to the non-android robot Roomba. In addition, Pepper was preferred (it scored significantly higher) to Nao for emotional support and communication abilities. The analyses assessing preferences for robots' potential occupations revealed significant differences among robots for welfare ( $F(2.196) = 47.248, p \ll 0.01$ ), protection/security ( $F(2.196) = 18.122, p \ll 0.01$ ), front office ( $F(2.196) = 57.269, p \ll 0.01$ ), and housework ( $F(2.196) = 41.932, p \ll 0.01$ ). Bonferroni post hoc tests showed that Pepper and Nao were considered significantly ( $p \ll 0.01$ ) more suitable than Roomba ( $p \ll 0.01$ ) for welfare (Roomba mean = 3.250, Nao mean = 4.700, Pepper mean = 4.780). Pepper was considered significantly more suitable than Nao ( $p = 0.028$ ) and Roomba ( $p \ll 0.01$ ) and Nao significantly more suitable than Roomba ( $p < 0.01$ ) for protection/security (Roomba mean = 3.020, Nao mean = 3.760, Pepper mean = 4.190). Pepper was considered significantly more suitable than Nao ( $p = 0.017$ ) and Roomba ( $p \ll 0.01$ ), and Nao significantly more suitable than Roomba ( $p \ll 0.01$ ) for front office (Roomba mean = 2.270, Nao mean = 3.930, Pepper mean = 4.340, see Fig. 30.5). Roomba was considered significantly more suitable than Pepper ( $p \ll 0.01$ ) and Nao ( $p \ll 0.01$ ) for housework (Roomba mean = 5.050, Nao mean = 3.420, Pepper mean = 3.740). No participants' gender effects were found.

### *Androids versus Humanoids Robots*

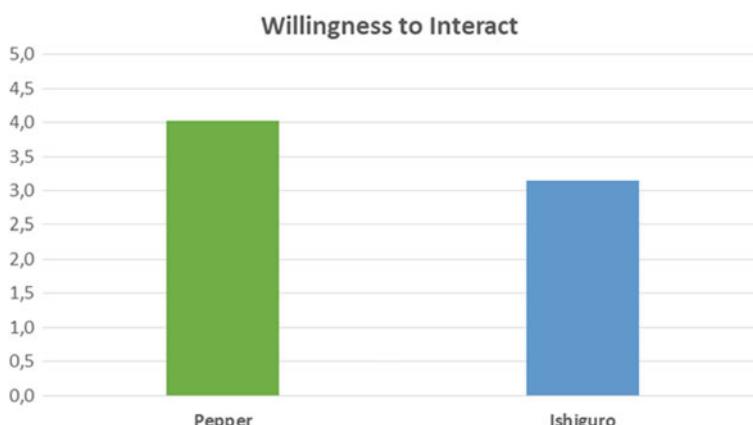


**Fig. 30.5** Occupations entrusted to robots by elders

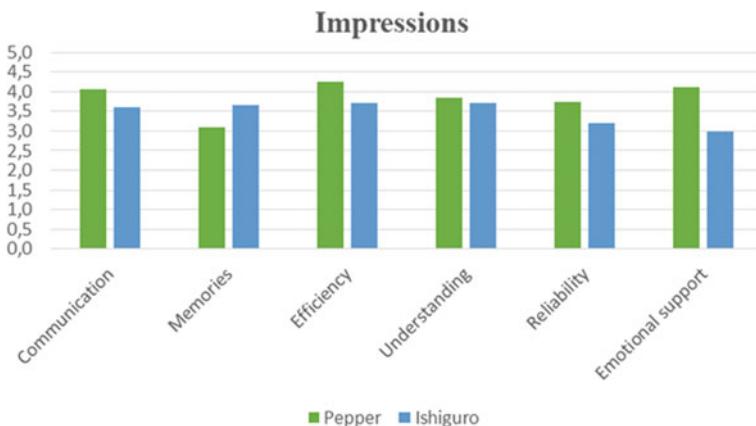
As previously mentioned, 50 participants (Group 1) participated to the assessment of Roomba, Nao, Pepper, and Ishiguro and the remaining 50 (defined Group 2) to the assessment of Roomba, Nao, Pepper and Erica. Since, as described in the above paragraph, a clear preference toward Pepper among seniors appeared, it was decided to compare Pepper's scores with those obtained by Ishiguro and Erica, respectively. These comparisons were carried out through repeated measures ANOVA, considering participants' gender as a between factor and "interaction," "appearance," "impressions," and "Occupations" as within factors.

### ***Pepper versus Ishiguro***

When comparing Pepper and Ishiguro, no participants' gender effect was found, for all the within factors under considerations. No significant differences were found between Pepper and Ishiguro for the variable appearance (Pepper mean = 4.300, Ishiguro mean = 4.260). Seniors scored the two robots equally well on their appearance. Significant differences emerged for "interaction" ( $F(1.48) = 9.984, p < 0.01$ ). Bonferroni post hoc tests revealed a significant preference among seniors to interact with Pepper (mean = 4.020) rather than Ishiguro (mean = 3.140,  $p < 0.01$ ). Figure 30.6 illustrates these results. Significant differences were observed for the cluster "impressions" ( $F(1.48) = 5.548, p = 0.023$ ). Separate ANOVA on each item of the cluster "impressions" showed significant differences between Pepper and Ishiguro for efficiency (Pepper mean = 4.240, Ishiguro mean = 3.700,  $p = 0.019$ ), reliability (Pepper mean = 3.740, Ishiguro mean = 3.200,  $p = 0.035$ ), and emotional support (Pepper mean = 4.120, Ishiguro mean = 2.980,  $p \ll 0.01$ ), in favor of Pepper, and for memories (Pepper mean = 3.080, Ishiguro mean = 3.660,  $p = 0.023$ ) in favor of Ishiguro. No differences were observed for communication (Pepper mean = 4.060, Ishiguro mean = 3.600) and language understanding (Pepper mean = 3.840, Ishiguro mean = 3.720), even though Pepper scored better than Ishiguro. Figure 30.7 illustrates these results. Concerning potential occupations seniors entrusted to robots,



**Fig. 30.6** Senior's willingness to interact with Pepper and Ishiguro

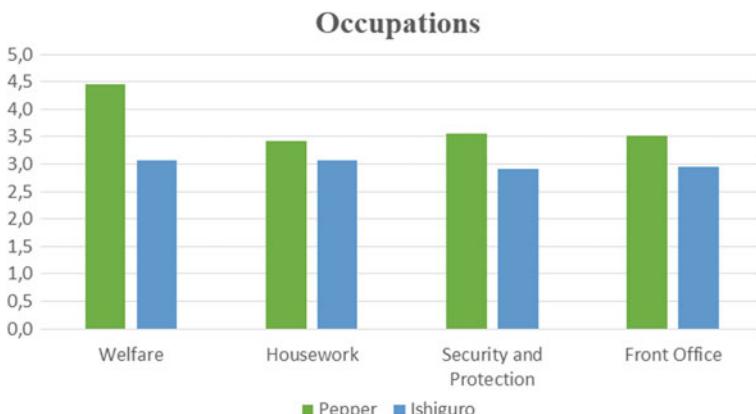


**Fig. 30.7** Pepper and Ishiguro's differences in terms of impressions aroused

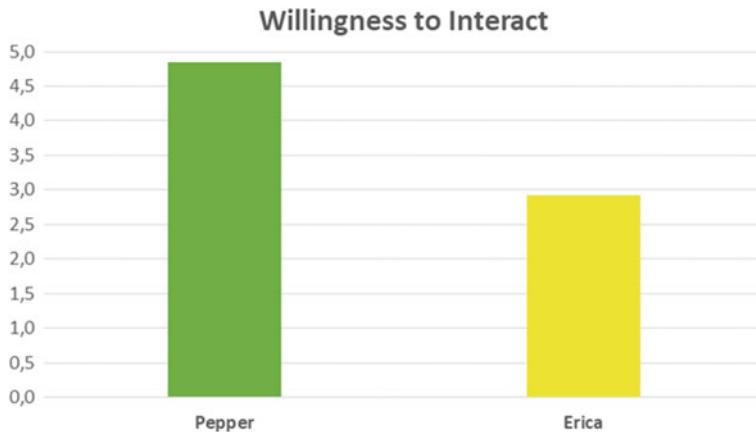
a significant difference ( $F(7.336) = 8.536, p < 0.01$ ) emerged between Pepper and Ishiguro. Bonferroni post hoc tests revealed that seniors considered Pepper significantly more suitable than Ishiguro for welfare (Pepper mean = 4.460, Ishiguro mean = 3.060,  $p < 0.01$ ), security/protection (Pepper mean = 3.560, Ishiguro mean = 2.920,  $p = 0.032$ ), housework (Pepper mean = 3.420, Ishiguro mean = 2.860,  $p = 0.049$ ), and front office (Pepper mean = 3.320, Ishiguro mean = 2.960,  $p = 0.021$ ). These data are illustrated in Fig. 30.8.

### ***Pepper versus Erica***

When comparing Pepper and Erica, no participants' gender effect was found, for all the within factors under considerations. No significant differences were found between Pepper and Erica for the variable appearance (Pepper mean = 4.340, Erica



**Fig. 30.8** Pepper and Ishiguro's scores in terms of occupations entrusted by seniors



**Fig. 30.9** Pepper and Erica's scores representative of seniors' willingness to interact with them

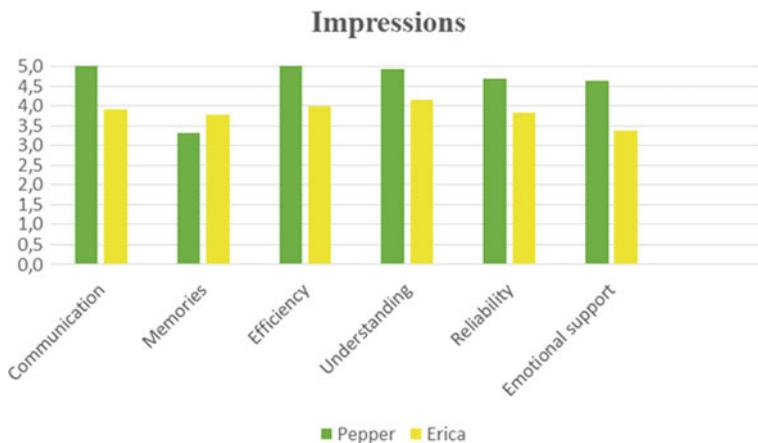
mean = 4.680). Seniors scored the two robots equally well on their appearance. Significant differences emerged for "interaction" ( $F(1.48) = 43.626, p \ll 0.01$ ). Bonferroni post hoc tests revealed a significant senior's preference to interact with Pepper (mean = 4.840) rather than Erica (mean = 2.920,  $p \ll 0.01$ ). Figure 30.9 illustrates these results.

Significant differences were observed for the cluster "impressions" ( $F(1.48) = 24.088, p \ll 0.01$ ). Separate ANOVA on each item of the cluster "impressions" showed significant differences between Pepper and Erica for communication (Pepper mean = 5.000, Erica mean = 3.920,  $p \ll 0.01$ ), efficiency (Pepper mean = 5.300, Erica mean = 4.000,  $p \ll 0.01$ ), reliability (Pepper mean = 4.700, Erica mean = 3.840,  $p \ll 0.01$ ), language understanding (Pepper mean = 4.920, Erica mean = 4.160,  $p < 0.01$ ), and emotional support (Pepper mean = 4.640, Erica mean = 3.360,  $p \ll 0.01$ ), in favor of Pepper, and for memories (Pepper mean = 3.320, Erica mean = 3.760,  $p = 0.043$ ) in favor of Erica. These results are illustrated in Fig. 30.10.

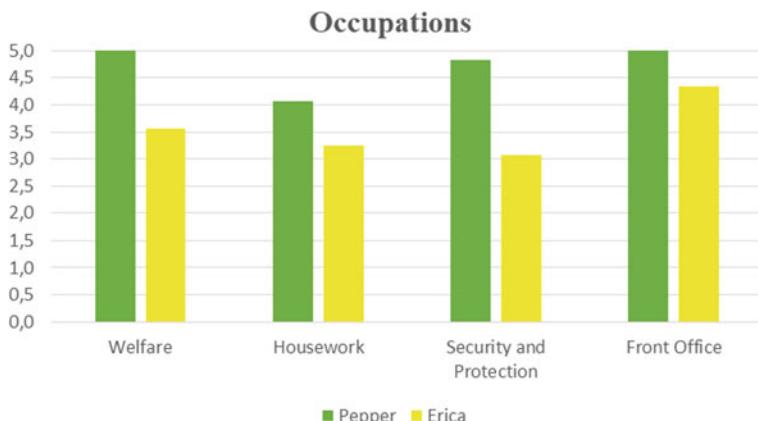
Concerning potential occupations seniors entrusted to robots, a significant difference ( $F(7.336) = 21.029, p \ll 0.01$ ) emerged between Pepper and Erica. Bonferroni post hoc tests revealed that seniors considered Pepper significantly more suitable than Erica for welfare (Pepper mean = 5.100, Erica mean = 3.560,  $p \ll 0.01$ ), security/protection (Pepper mean = 4.820, Erica mean = 3.080,  $p \ll 0.01$ ), housework (Pepper mean = 4.060, Erica mean = 3.240,  $p < 0.01$ ), and front office (Pepper mean = 5.100, Erica mean = 3.560,  $p < 0.01$ ). These data are illustrated in Fig. 30.11.

## 30.4 Discussion and Conclusion

The present investigation was carried out with the aim to investigate elderly people's preferences among five robots, characterized by different levels of human likeness.



**Fig. 30.10** Pepper and Erica's differences in terms of impressions aroused



**Fig. 30.11** Pepper and Erica's differences in terms of occupations entrusted by elders

As previously mentioned, the selected robots show different degrees of similarity to human beings. Roomba is certainly the one whose features are least reminiscent of a person, being a tool useful for house cleaning. Nao and Pepper are humanoid robots with features remotely resembling human beings, and finally, Ishiguro and Erica are characterized by a high level of human likeness to the extent that, at first glance, they can be mistaken for real people. The proposed investigation tested first who, between Ishiguro and Erica, was favored by elders, then compared them with the elders' most favorite robot among Roomba, Nao, and Pepper. For each robot, the following attributes were tested: participants' willingness to interact with them, robots' appearance, and general impressions aroused by them. Being appearance a complex concept is noteworthy to underline here that the present investigation takes

into account only one aspect of it, namely robots' level of human likeness. In addition, seniors were required to select among welfare, housework, protection/security, and front office occupations which one is more suitable for the proposed robots. Results showed that seniors did not express any major preference between Erica and Ishiguro, either in terms of willingness to interact, or appearance, or impressions and robots' gender. This can be probably attributed to the fact that the two robots are very similar in terms of human likeness, both in body shape and movement skills. Seniors, however, considered Erica more suitable than Ishiguro for front office occupations, suggesting a senior gender preference toward Erica for this task.

The comparison among Roomba, Nao, and Pepper was carried out in order to understand whether humanoid features are more attractive than plain robot features and to what extent small differences among robots play a role on elders' robot's acceptance, considering that Nao and Pepper have similar humanoid features while Roomba acted as control to test acceptance in terms of practical purposes. It emerged that Pepper's appearance was greatly appreciated by seniors, to the extent that their willingness to interact, their feelings about it, and its appearance scored significantly higher compared to Roomba and Nao. More specifically, Pepper's communication skills and its ability to provide emotional support scored significantly better than Roomba and Nao, while its ability to arouse memories of someone or something, to understand people and be reliable, scored significantly better than Roomba and equally well than Nao. Interestingly, when seniors were asked to indicate how much suitable they considered Pepper, Nao, and Roomba to perform welfare protection/security tasks and front office occupations, Pepper scored significantly higher than Nao and Roomba and Nao significantly higher than Roomba, suggesting a clear senior preference toward Pepper. Roomba was considered the most suitable among the three robots in performing housework task, which was a clearly expected result since Roomba is essentially a vacuum cleaner.

In order to understand which level of human likeness seniors would accept in robots and assess also gender differences, comparisons were made between Ishiguro and Pepper and then between Erica and Pepper, within two different groups of 50 seniors each.

The results suggest that seniors did not have a gender preference since neither Ishiguro nor Erica was preferred to Pepper, although Erica was preferred over Ishiguro for front office works. Seniors expressed a greater willingness to interact with Pepper, considering it more efficient, communicative, reliable, able to understand language, able to provide emotional support and performing welfare, security/protection, housework, and front office occupation than Ishiguro. Similarly, seniors' preferences were all significantly in favor of Pepper, when Erica and Pepper were compared. Pepper was considered more communicative, efficient, reliable, comprehensive, and emotionally supportive than Erica, as well as more able to perform welfare, protection/security, housework, and front office occupations.

In summary, the humanoid robot Pepper was the preferred by seniors. An interesting detail concerns Pepper's scores in terms of *appearance*. When Pepper was compared with Ishiguro and Erica, who were characterized by a higher level of human likeness than Pepper, no significant differences emerged among them, suggesting

that seniors were not enthusiast of the high degree of human likeness of Ishiguro and Erica. This reaction of seniors is observed probably because the excessive level of resemblance of robots to human beings negatively affects their evaluation and leads people to refuse their daily use, because of the “uncanny valley effect,” i.e., a feeling of discomfort and eeriness caused by their excessive resemblance to a human being. Our data go in this direction. However, the current investigation accounts only of few features driving user’s acceptance of robots. More investigations are needed in order to design socially believable human–robot interactions and increase users’ acceptance of such assistive technologies.

### Acknowledgements



The research leading to these results has received funding from the European Union Horizon 2020 research and innovation programme under grant agreement N. 769872 (EMPATHIC) and N. 823907 (MENHIR) and from the project SIROBOTICS that received funding from Ministero dell’Istruzione, dell’Università, e della Ricerca (MIUR), PNR 2015-2020, Decreto Direttoriale 1735 del 13 luglio 2017.

## References

1. Brenton, H., Gillies, M., Ballin, D., Chatting D.: The Uncanny Valley: does it exist? In: Proceeding of the Workshop on Human-Animated Characters Interaction at The 19th British HCI Group Annual Conference HCI 2005: The Bigger Picture, Edinburgh (2005)
2. Broadbent, E., Stafford, R., MacDonald, B.: Acceptance of healthcare robots for the older population: review and future directions. **1**, 319–330. Springer
3. Esposito, A., Esposito, A.M.: On the recognition of emotional vocal expressions: motivations for an holistic approach. *Cogn. Process.* **13**(2), 541–550 (2012)
4. Esposito, A., Esposito, A.M., Vogel, C.: Needs and challenges in human computer interaction for processing social emotional information. *Pattern Recogn. Lett.* **66**, 41–51 (2015)
5. Esposito, A., Jain, L.C.: Modeling social signals and contexts in robotic socially believable behaving systems. In: Esposito, A., Jain L.C. (eds.) *Toward Robotic Socially Believable Behaving Systems Volume II—“Modeling Social Signals”*. ISRL series 106, pp. 5–13. Springer, Switzerland
6. Komatsu, T., Kurosawa, R., Yamada, S.: How does the difference between users’ expectations and perceptions about a robotic agent affect their behavior? *Int. J. Soc. Robot.* **4**, 109–116 (2012)
7. MacDorman, K.F., Chattopadhyay D.: Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. 190–205 (2015). Elsevier
8. Maldonato, N.M., Dell’Orco, S.: Making decision under uncertainty, emotions, risk and biases. In: Bassis S., Esposito A., Morabito F.C. (eds.) *Advances in Neural Networks: Computational and Theoretical Issues*. SIST series 37, pp. 293–302. Springer, Switzerland (2015)
9. Mori, M.: The uncanny valley. *Energy* **7**(4), 33–35 (1970)
10. Peek, S.T., Wouters, E.J., van Hoof, J., Luijckx, K.G., Boeije, H.R., Vrijhoef, H.J.: Factors influencing acceptance of technology for aging in place: a systematic review. *Int. J. Med. Inform.* **83**(4), 235–248 (2014)

11. Phillips, E., Ullman, D., de Graaf, M.M.A., Malle, B.F.: What does a robot look like? A multi-site examination of user expectations about robot appearance. In: Proceedings of the Human Factors and Ergonomics Society 2017 Annual Meeting, pp. 1215–1219
12. Troncone, A., Palumbo, D., Esposito, A.: Mood effects on the decoding of emotional voices. In: Bassis, S., et al. (eds.), Recent Advances of Neural Network Models and Applications. SIST 26, pp. 325–332. International Publishing, Switzerland
13. Wang, S., Lilienfeld, S.O., Rochat, P.: The uncanny valley: existence and explanations. *Am Psychol. Assoc.* **4**, 393–407 (2015)
14. Złotowski, J.A., Sumioka, H., Nishio, S., Glas, D.F., Bartneck, C., Ishiguro, H.: Persistence of the uncanny valley: the influence of repeated interactions and a robot's attitude on its perception. *Front. Psychol.* **6**, 883 (2015)

## Chapter 31

# The Influence of Personality Traits on the Measure of Restorativeness in an Urban Park: A Multisensory Immersive Virtual Reality Study



Vincenzo Paolo Senese, Aniello Pascale, Luigi Maffei, Federico Cioffi,  
Ida Sergi, Augusto Gnisci and Massimiliano Masullo

**Abstract** This study investigates the influence of personality traits and water masking installations on the perceived restorativeness of an urban park by means of Multisensory Immersive Virtual Reality (M-IVR) methodology. To this aim, 95 adults (67 females, 28 males) were administered the NEO-FFI to measure personality and were presented two kinds of M-IVR scenarios, representing an urban park without any installation (S0) and the same park with a water installation (S1); in both scenarios, the perceived restorativeness was measured by means of the Perceived Restorativeness Scale (PRS-11). Results of ANOVAs showed that the perceived restorativeness (fascination and being-away components) was increased by the water installations, but that the effect was attenuated by personality. Correlation analysis showed that the extroversion dimension was weakly and negatively related to the fascination and being-away change score. These results suggest that M-IVR is a valid paradigm to investigate in a controlled but ecological way the effect of installations on the

---

V. P. Senese (✉) · F. Cioffi · I. Sergi · A. Gnisci

Department of Psychology, University of Campania “Luigi Vanvitelli”, Caserta, Italy  
e-mail: [vincenzopaolet.senese@unicampania.it](mailto:vincenzopaolet.senese@unicampania.it)

F. Cioffi

e-mail: [federico.cioffi.88@gmail.com](mailto:federico.cioffi.88@gmail.com)

I. Sergi

e-mail: [id.a.sergi@unicampania.it](mailto:id.a.sergi@unicampania.it)

A. Gnisci

e-mail: [augusto.gnisci@unicampania.it](mailto:augusto.gnisci@unicampania.it)

A. Pascale · L. Maffei · M. Masullo

Department of Architecture and Industrial Design, University of Campania “Luigi Vanvitelli”,  
Caserta, Italy

e-mail: [aniello.pascale@unicampania.it](mailto:aniello.pascale@unicampania.it)

L. Maffei

e-mail: [luigi.maffei@unicampania.it](mailto:luigi.maffei@unicampania.it)

M. Masullo

e-mail: [massimiliano.masullo@unicampania.it](mailto:massimiliano.masullo@unicampania.it)

perceived restorativeness of environment and showed that beneficial effect of water installations on the evaluation of urban green parks is also related to personality characteristics.

### 31.1 Introduction

An urban park is a public open space within a city that incorporates green spaces and recreations for residents and visitors. Historically, urban parks have been designed for several functions; among them, they have been considered because they can restore mental fatigue [1, 2]. This property of natural environments has been defined as “restorativeness” of places [3], that is the potential of specific environments of re-establishing cognitive capacities related to human information processing and executive functioning, particularly attention and concentration [4]. According to the Attention Restoration Theory (ART) [2], spending time in nature or looking at scenes of nature has a positive effect because nature facilitates the recovery of directed attention capacities. Moreover, it has been showed that nature provides benefits beyond the visual components and that the auditory components are also critical [5]. According to the ART model [6], four characteristics are identified as environmental restorativeness: (1) fascination: the capacity of the environment to effortlessly draw individuals’ attention; (2) being away: the capacity of a place to make the individual feel away from daily concerns; (3) extent: refers to the coherence of the environment, encouraging exploration by the individual; and (4) compatibility: capacity of a place to meet individual’s inclinations and interests. Several authors found that the restorativeness of a place is correlated with several psychophysiological benefits, such as preference judgements, positive emotions, quality of life, stress reduction and sustainable behaviour [1].

Although urban parks are designed to have a restorativeness function, they can fall short on this because studies showed that the subjective perception of the park environment is multifactorial and depends either from the landscape and soundscape of the environment [7, 8], or on individual differences [9, 10].

As regards the first aspect, it has been noticed that the soundscape of urban parks can suffer because of traffic noise and that alternative strategies (e.g. informative masking) could be adopted [11–16]. In this view, a recent experimental study has shown that the inclusion of multisensory (audiovisual) installation of natural features in a city park has a positive influence on the perceived restorativeness of the environment [13]. But more studies are needed to better understand the specific components that are responsible for the observed effect.

As regards the individual differences, several studies have shown that personality factors influence natural environment perception and individual behaviours [10], and that the psychological benefits associated with exposure to nature are a function of personality [9]. To our knowledge, no study has investigated yet if the inclusion of informative masking installations in the urban parks has the same positive effect on individuals independently of personality factors.

Therefore, the aim of the present paper was to investigate the association between personality factors and changes in the perception of the urban park restorativeness after the inclusion of a multisensory natural informative masking installation. To this aim, a Multisensory Immersive Virtual Reality (M-IVR) [17] methodology was used to investigate the benefits of introducing multisensory water installations on perceived restorativeness of an existing urban park and to verify if the effect is observed over and above individual differences.

We considered M-IVR system because it allows the presentation and the exploration of complex, multisensory and dynamic virtual environments that provide a continuous stream of stimuli congruent with the actual movements that deliver a vivid illusion of reality to the participant [18]. This relates to the concept of “presence”, that is, the subjective feeling of “being in the virtual environment” [18], thus making it a key concept in the research related to the effectiveness of virtual reality. The widespread availability and pros deriving from virtual reality technology made it a very promising tool to investigate human perceptions and responses to surrogate real environments [19]. The degrees of freedom provided by virtual environment in the ability of creating and manipulating simulated environments for research purposes offer the opportunity for researchers to study a wide range of individuals’ perceptions, preferences and behaviours [20].

Several researchers that are using virtual reality to investigate individual reactions to virtual environment highlighted that in general, individuals show similar responses in the real or virtual environment [21, 22]. Participants exposed to a natural environment simulated virtually showed higher levels of positive effect and a greater perception of restorativeness if compared to participants exposed to a simulated urban environment [23]. Moreover, positive psychological and psychophysiological effects have also been observed for participants exposed to a virtual environment depicting a forest [24] or to a simulated nature walk [21].

In this study, in line with previous researches [12, 13], two kinds of M-IVR scenarios were created and administered to participants: a baseline scenario, representing an urban park without any installation (S0); and the same park with a water installation (S1). For both scenarios, the perceived restorativeness was measured. Moreover, given the relevance of personality on the environment perception, the personality traits were also measured. We expected that scenario with the water installation was perceived as more restorative than the baseline scenario and that the perceived restorativeness was associated with the personality traits.

## 31.2 Method

### 31.2.1 Sample

A total of 95 adults (67 females, 28 males) participated in a within-subject experimental design. Their ages ranged from 20 to 29 years ( $M = 22.2$ ,  $SD = 1.9$ ), and

their educational level varied from middle school to college levels. All participants were tested individually.

### 31.2.2 Procedure

The experimental session was divided into two phases. In the first phase, a basic sociodemographic questionnaire and a self-report scale of personality were administered. In the second phase, participants wear the Oculus head-mounted display and Sennheiser HD201 headphones, and two multisensory virtual environments were presented: a first baseline scenario representing an urban park without any installation (S0) and a second scenario representing the same urban park with the addition of a water installation (S1). For each scenario, participants were virtually seated on a bench inside the park (a chair in the real environment), and they were invited to explore the surrounding environment with free movements of the head for about 3 min. After that, the testing session started. Participants were administered auditorily by headphones the items of the restorativeness scale and responded verbally. Tests were carried out in conformity with the local Ethics Committee requirements and the Declaration of Helsinki, and all participants signed a written informed consent before starting the experimental session. The session lasted about 25 min.

### 31.2.3 Measures

**NEO Five-Factor Inventory (NEO-FFI).** The 60-item Italian short form of the NEO-PI-R [25] was administered to measure personality traits. The scale was designed to measure the five-factor model of personality [26]. In the NEO-FFI, twelve-item scales are used to measure each of the five facets: Neuroticism, Extraversion, Openness to experience, Agreeableness and Conscientiousness. Each item presents a description, and participants are asked to rate their level of agreement on a 5-point Likert scale ranging from “strongly disagree” (=1) to “strongly agree” (=5). Raw scores were considered, with higher scores indicating higher Neuroticism, Extraversion, Openness to experience, Agreeableness, or Conscientiousness. All the NEO-FFI facets showed adequate reliability ( $\alpha_s > 0.70$ ).

**The Perceived Restorativeness Scale (PRS-11).** The short Italian version of the Perceived Restorativeness Scale (PRS-11) [27] was administered to evaluate the perceived restorativeness of the two virtual scenarios. The PRS-11 is a self-report scale developed according to the Attention Restoration Theory (ART) [2] to measure the four restorative factors of the environment: (a) fascination (three items), a type of attention assumed to be effortless and without capacity limitations (e.g. “Places like that are fascinating”); (b) being away (three items), that is the effect of physical and/or psychological being away from demands on directed attention (e.g. “Places like that are a refuge from nuisances”); (c) coherence (three items) perceived in an

environment (e.g. “There is a clear order in the physical arrangement of places like this”); and (d) scope (two items) of the environment (e.g. That place is large enough to allow exploration in many directions”). Each item presents a description, and participants are asked to rate their level of agreement on a 11-point scale: from “not at all” (=0) to “completely” (=10). Higher scores indicate greater restorativeness. All items were recorded by means of a recorder into an anechoic chamber and were playback auditorily and in a random order to the participants. The experimenter recorded the verbal responses of participants for each item. All the PRS-11 subscales showed adequate reliability ( $\alpha_s > 0.70$ ).

### 31.2.4 Virtual Scenarios

**Graphical stimuli.** A virtual simulation of an existing urban park, the Villa Comunale of Naples (Italy), already used in previous experiments [12, 13], was created starting from a model made in 3ds Max from technical draws and field surveys. Subsequently, the basic visual scenario (S0) was built into Unreal Engine by adding materials, lights, the wind effect on trees and the animations of vehicles’ passing on the park boundaries. The 3D model of the site consists of a public park with rectangular shape framed by city roads. Inside the park, the natural environment is prevalent, even though the vehicular traffic is visible from several points of the park and the road traffic noise perceivable. All the elements that are present in the real environment were inserted into the 3D model. A bench was chosen as the point of view (POV) of the subject during the test. A second scenario (a “new scenario”; S1) was created by manipulating the basic scenario. In S1, a traditional water installation (a stream) was modelled. The installation was characterized by a water movement of a stream.

**Audio stimuli.** Recordings of all the selected urban park sound elements [12, 13] were carried by means of a Zoom H6 recorder equipped with a SoundField SPS200 microphone. The background noise, produced mainly by road traffic, the wind and the people in the park, was recorded. To determine the sound equivalent level of recordings, a sound level meter “Solo 01-dB” was positioned close to the SoundField microphone. The recordings in B-format were then processed by means of the plug-in “Surrounding Zone” for playback in a 5.1 surround system set-up. A 5.1 virtual speaker configuration has been then built into the virtual world to reproduce the sound field around the participant position. The directivity of each virtual speaker has been set according to the directivity diagram of a real speaker (Dynaudio X3). The water installation sound was downloaded from a free web library and represents the recording of a water stream sound. The level of the water sound reproduced for the experiment has been set to be  $-3$  dB of the background road traffic noise. The playback system was calibrated by mean of an Mk1 Cortex dummy head and a Symphonie sound card to reproduce, in the virtual position of the subject, the same equivalent level of 58.0 dB(A) measured in situ.

### 31.3 Data Analysis

Normality of univariate distributions of PRS-11 subscale scores was preliminarily checked. To analyse the effect of the installation on the perceived restorativeness of the scenarios, for each dimension of the PRS-11 scale (fascination, being away, coherence, scope), restorativeness scores were analysed by means two-way within-subject ANOVAs that treated scenarios (S0 and S1) as a two-level within-subjects factor. Moreover, to investigate whether the effect was observed over and above personality, the same analysis was repeated by including each facet of personality as covariate. Bonferroni's correction was used to analyse post hoc effects of significant factors, and partial eta-squared ( $\eta_p^2$ ) was used to evaluate the magnitude of significant effects. To investigate the specific association between personality score and the change in the perception of restorativeness as a function of the water installation, a change score was computed, by subtracting the restorativeness scores of the basic scenario to the scores of the new scenario (S1–S0), and then Pearson correlation coefficients were computed between personality factors and change scores.

### 31.4 Results

The ANOVA on the fascination dimension showed that judgements were strongly influenced by the water installation,  $F(1,94) = 99.97, p < 0.001, \eta_p^2 = 0.515$ . Mean comparisons showed that “new scenario” (S1),  $M = 5.56, 95\% \text{ CI} [5.23; 5.88]$ , was evaluated as more fascinating than the baseline scenario (S0),  $M = 4.27, 95\% \text{ CI} [3.93; 4.62]$  (see Table 31.1).

The ANOVA on the being-away dimension showed that judgements were strongly influenced by the water installation,  $F(1,94) = 54.54, p < 0.001, \eta_p^2 = 0.367$ . Mean comparisons showed that scenario S1,  $M = 5.72, 95\% \text{ CI} [5.37; 6.06]$ , was evaluated as stimulating more the feel of being away than the scenario S0,  $M = 4.95, 95\% \text{ CI} [4.57; 5.32]$ .

The ANOVAs on the coherence and scope dimensions did not show any significant effect of installation on restorativeness,  $F(1,94) = 1.16, p = 0.284, \eta_p^2 = 0.012$  and  $F(1,94) = 0.54, p = 0.462, \eta_p^2 = 0.006$ , respectively, for coherence and scope.

With regards the fascination dimension, the ANCOVAs showed that when controlling for the five personality dimensions, the effect of installation on the perception of fascination was strongly reduced; and that when considering Openness to experience and Conscientiousness, in particular, the effect was not significant. Moreover, data showed that the scenario moderated the effect of extraversion on the fascination scores.

As regards the being-away dimension, the ANCOVAs confirmed that when controlling for the five personality dimensions, the effect of installation on the perception of being away was reduced and showed that when considering Neuroticism, Openness to experience and Agreeableness, in particular, the effect was not significant.

**Table 31.1** Results of two-way within-subject ANOVAs and ANCOVAs (partial eta-squared) on restorativeness scores as a function of personality factors and PRS-11 dimensions

Effect <sup>a</sup>	PRS-11			
	Fascination $\eta_p^2$	Being away $\eta_p^2$	Coherence $\eta_p^2$	Scope $\eta_p^2$
Scenario	0.515***	0.367***	0.012	0.006
<i>Neuroticism</i>				
Scenario	0.049*	0.035	0.006	0.003
N	0.006	0.018	0.018	0.003
Scenario × N	0.017	0.005	0.015	0.001
<i>Extraversion</i>				
Scenario	0.187***	0.145***	0.004	0.001
E	0.002	0.010	0.004	0.005
Scenario × E	0.051*	0.051*	0.001	<0.001
<i>Openness to experience</i>				
Scenario	0.037	0.025	<0.001	0.003
O	0.016	0.008	0.015	0.004
Scenario × O	0.001	0.002	0.001	0.002
<i>Agreeableness</i>				
Scenario	0.106***	0.014	0.011	0.021
A	0.008	0.056*	0.014	0.005
Scenario × A	0.016	0.002	0.007	0.017
<i>Conscientiousness</i>				
Scenario	0.034	0.052*	0.004	<0.001
C	<0.001	0.045*	0.006	<0.001
Scenario × C	0.003	0.003	0.002	0.008

<sup>a</sup>N—Neuroticism; E—Extraversion; O—Openness to experience; A—Agreeableness; C—Conscientiousness

\*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

Moreover, data showed that Agreeableness ( $r = -0.232$ ,  $p < 0.05$ ,  $N = 95$  and  $r = -0.224$ ,  $p < 0.05$ ,  $N = 95$ , respectively for S0 and S1) and Conscientiousness ( $r = -0.183$ ,  $p = 0.07$ ,  $N = 95$  and  $r = -0.228$ ,  $p < 0.05$ ,  $N = 95$ , respectively for S0 and S1) scores were weakly and negatively associated with being-away scores, and that the scenario moderated the effect of extraversion on the being-away scores.

Finally, correlation analysis between personality score and the change in the perception of restorativeness as a function of the water installation confirmed that only the extraversion scores were significantly but weakly associated with change scores (see Table 31.2). The higher was the self-reported extraversion; the lower was the change in the fascination and being-away scores between the baseline scenario and the scenario with the water installation.

**Table 31.2** Pearson's correlation coefficients between personality factors and change scores

Personality factor <sup>a</sup>	Change score			
	Fascination	Being away	Coherence	Scope
N	0.130	0.070	0.123	-0.035
E	-0.227*	-0.226*	-0.035	-0.014
O	-0.038	-0.043	0.030	-0.048
A	-0.126	0.042	-0.083	-0.132
C	0.059	-0.054	-0.041	0.013

<sup>a</sup>N—Neuroticism; E—Extraversion; O—Openness to experience; A—Agreeableness; C—Conscientiousness

\* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$

## 31.5 Discussion

The aim of the present study was to investigate the influence of personality traits and water masking installations on the perceived restorativeness of an urban park by means of Multisensory Immersive Virtual Reality (M-IVR) methodology. The literature showed that spending time in nature or looking at scenes of nature has a positive effect because nature facilitates the recovery of directed attention capacities [1, 2]. Moreover, researchers showed that the auditory components of natural environments provide specific benefits [5]. Starting from these considerations, in this study, an M-IVR methodology [17] was used to investigate the effect of (audiovisual) water installations on perceived restorativeness of an urban park. We considered M-IVR system because it allows the presentation and the manipulation of complex scenarios recreating existing environment, and studies have shown that individuals exhibit similar responses in the real or virtual environment [21–24].

Moreover, considering that subjective perception of the park environment is multifactorial and depends either from landscape and soundscape of the environment or on individual differences [7–10], we investigated whether the effect of installations on restorativeness perception was influenced by the personality factors.

In line with the previous literature [12–16], results showed that the water installation strongly increases the restorativeness perception of the urban park. In particular, data showed that only the fascination and the being-away components were significantly affected by masking installation, not coherence and scope. This result confirms that the presence of water installation (stream), with the relative audiovideo components, affects the perception of the pleasantness of an urban park [14–16]. Moreover, this result confirms that the M-IVR methodology can be considered a valuable approach to investigate human perceptions and responses to surrogate real environments [19, 20], and to investigate in advance the potential benefits of environmental manipulations.

With regard to the personality dimensions, results showed that individual differences influence the perception of restorativeness and the impact of the water installation. This result also is in line with the previous literature and confirms that

personality affects the way individual perceives the natural environment [9, 10]. This is the first time that personality is considered as a critical factor in the evaluation of environmental manipulation for masking. The results suggest that individual differences affect the perception of the benefit of specific masking strategies. In other terms, this indicates that if we are interested in environment interventions that have universal benefits, individual difference should be considered to define the optimal solutions, though it is worth to notice that the effect was weak. Indeed, our results showed that the personality dimensions (extraversion in particular) influence the degree of benefit for the informational masking interventions. Further studies are needed to replicate our findings and to better understand how personality and the specific components of installations interact in the regulation of restorativeness perception. In this perspective, we believe that the M-IVR methodology can be considered a gold standard to test in advance the effect of different solutions on environmental perception and to investigate which installations have universal effects and why.

The results of this study should be interpreted with certain limitations in mind. First, we considered only a single urban park, and future studies should verify the generalizability of our findings. We considered only a stream water installation; future studies should compare the effect of different masking installations. Finally, it is possible that cultural factors can influence the perception of installations, future studies should consider the role of cultural differences.

## 31.6 Conclusions

The results of this study showed that M-IVR methodology can be considered a gold standard to test individuals' perceptions, preferences and behaviours in simulated environments. Moreover, data showed that personality influences the perception of restorativeness of the environment and the impact of the water installation. If replicated, these results indicate that individual differences are a critical aspect that influences the perception of the quality of the urban parks, in terms of both soundscape and landscape. Therefore, we suggest that to better design optimal mitigation interventions, researchers should consider to what extent the planned interventions are beneficial over and above individual differences.

## References

1. Berto, R.: The role of nature in coping with psycho-physiological stress: a literature review on restorativeness. *Behav. Sci.* **4**, 394–409 (2014). <https://doi.org/10.3390/bs4040392>
2. Kaplan, S.: The restorative benefits of nature: toward an integrative framework. *J. Environ. Psychol.* **15**, 169–182 (1995)
3. Hartig, T.: Restorative environments. In: Spielberger, C. (ed.) *Encyclopedia of Applied Psychology*, vol. 3, pp. 273–279. Academic Press, San Diego, CA (2004)

4. Hernández, B., Hidalgo, M.C., Berto, R., Perón, E.: The role of familiarity on the restorative value of a place, research on a Spanish sample. *Bullettin People Environ. Stud.* **18**, 22–25 (2001)
5. Franco, L.S., Shanahan, D.F., Fuller, R.A.: A review of the benefits of nature experiences: more than meets the eye. *Int. J. Environ. Res. Public Health* **14**(8), E864 (2017). <https://doi.org/10.3390/ijerph14080864>
6. Kaplan, R., Kaplan, S.: *The Experience of Nature: A Psychological Perspective*. Cambridge University Press, Cambridge (1989)
7. Brambilla, G., Maffei, L.: Responses to noise in urban parks and in rural quiet areas. *Acta Acustica United with Acustica* **92**, 881–886 (2006)
8. Tse, M.S., Chau, C.K., Choy, Y.S., Tsui, W.K., Chan, C.N., Tang, S.K.: Perception of urban park soundscape. *J. Acoust. Soc. Am.* **131**, 2762–2771 (2012)
9. Ambrey, C.L., Cartlidge, N.: Do the psychological benefits of greenspace depend on one's personality? *Pers. Individ. Differ.* **116**, 233–239 (2017)
10. Johnsen, S.Å.K.: Exploring the use of nature for emotion regulation: associations with personality, perceived stress, and restorative outcomes. *Nord. Psychol.* **65**(4), 306–321 (2013)
11. Durlach, N., Mason, C.R., Kidd, G., Arbogast, T.L., Colburn, H.S., Shinn-Cunningham, B.G.: Note on informational masking. *J. Acoust. Soc. Am.* **113**(6), 2984–2987 (2003)
12. Masullo, M., Maffei, L., Pascale, A.: Effects of combination of water sounds and visual elements on the traffic noise mitigation in urban green parks. In: *Proceeding of INTERNOISE 2016*, 21–24 Aug, pp. 3910–3916. Hamburg, Germany (2016)
13. Masullo, M., Maffei, L., Pascale, A., Senese, V.P.: An alternative noise mitigation strategy in urban green park: a laboratory experiment. Conference paper presented at the 46th international congress and exposition on noise control engineering, 27–30 Aug. Hong Kong Convention and Exhibition Centre (2017)
14. Galbrun, L., Ali, T.T.: Acoustical and perceptual assessment of water sounds and their use over road traffic noise. *J. Acoust. Soc. Am.* **133**, 227–237 (2013)
15. Hong, J.: Designing sound and visual components enhancement of urban soundscapes. *J. Acoust. Soc. Am.* **134**, 2026–2036 (2013)
16. Leung, T.M., Chau, C.K., Tang, S.K., Xu, J.M.: Developing a multivariate model for predicting the noise annoyance responses due to combined water sound and road traffic noise exposure. *Appl. Acoust.* **127**, 284–291 (2017)
17. Iachini, T., Maffei, L., Ruotolo, F., Senese, V.P., Ruggiero, G., Masullo, M., Alekseeva, N.: Multisensory assessment of acoustic comfort aboard metros: an immersive virtual reality study. *Appl. Cogn. Psychol.* **26**, 757–767 (2012). <https://doi.org/10.1002/acp.2856>
18. Slater, M., Wilbur, S.A.: Framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments. *Presence Teleoperators Virtual Environ.* **6**(6), 603–616 (1997)
19. Smith, J.W.: Immersive virtual environment technology to supplement environmental perception, preference and behavior research: a review with applications. *Int. J. Environ. Res. Public Health* **12**(9), 11486–11505 (2015)
20. Slater, M.: A note on presence terminology. *Presence Connect* **3**, 1–5 (2003)
21. Calogiuri, G., Littleskare, S., Fagerheim, K.A., Rydgren, T.L., Brambilla, E., Thurston, M.: Experiencing nature through immersive virtual environments: environmental perceptions, physical engagement, and affective responses during a simulated nature walk. *Front. Psychol.* **8**, 2321 (2018). <https://doi.org/10.3389/fpsyg.2017.02321>
22. Iachini, T., Coello, Y., Frassinetti, F., Senese, V.P., Galante, F., Ruggiero, G.: Peripersonal and interpersonal space in virtual and real environments: effects of gender and age. *J. Environ. Psychol.* **45**, 154–164 (2016)
23. Schutte, N.S., Bhullar, N., Stilinović, E.J., Richardson, K.: The impact of virtual environments on restorativeness and affect ecopsychology. *Ecopsychology* **9**(1), 1–7 (2017). <https://doi.org/10.1089/eco.2016.0042>
24. Valtchanov, D., Barton, K.R., Ellard, C.: Restorative effects of virtual nature settings. *Cyberpsychol. Behav. Soc. Networking* **13**, 503–512 (2010)

25. Caprara, G.V., Barbaranelli, C., Hahn, R., Comrey, A.L.: Factor analyses of the NEO-PI-R inventory and the Comrey personality scales in Italy and the United States. *Pers. Individ. Differ.* **30**, 217–228 (2001)
26. McCrae, R.R., John, O.P.: An introduction to the five-factor model and its applications. *J. Pers.* **60**, 175–216 (1992)
27. Pasini, M., Berto, R., Brondino, M., Hall, R., Ortner, C.: How to measure the restorative quality of environments: the PRS-11. *Procedia Soc. Behav. Sci.* **159**, 293–297 (2014). <https://doi.org/10.1016/j.sbspro.2014.12.375>

# Chapter 32

## Linguistic Repetition in Three-Party Conversations



Justine Reverdy, Maria Koutsombogera and Carl Vogel

**Abstract** The conversational mechanism of repetition appears to be strongly connected to the development of common ground among conversation participants. We report on three-party game-based interactions where two players participate in a quiz supervised by a facilitator. We use a semi-automatic method to detect alignment between players by observing linguistic repetitions in the dialogue transcripts and investigate the relation of the alignment to the type of the facilitator's feedback. Results suggest that the repetitions detected with this method have a function in the interaction, as it is reflected in the verbal and non-verbal behaviours of an interaction facilitator: facilitators provided more encouragement than expected where alignment lacked evidence and less than expected where alignment was ample.

### 32.1 Introduction

The success of dialogue participants in jointly achieving a collaborative task largely depends on the way participants co-construct common knowledge. Linguistic repetitions are strongly connected to the process of establishing common ground [1] and are frequent within communicative behaviours, associated with multiple functions [2]. For example, linguistic repetition is often used in conversational repair [3], which may be distinguished from establishing common ground.<sup>1</sup> They are often used to diminish chances of miscommunication [4], but can also signal engagement or in-

---

<sup>1</sup>If A says, “Let’s order cold brew coffee”, B might respond with “Cold brew coffee is nice”, to establish that common ground exists or with “Cold brew coffee?” to signal that repair is necessary.

J. Reverdy (✉) · M. Koutsombogera · C. Vogel  
School of Computer Science and Statistics, Trinity College Dublin, Dublin 2, Ireland  
e-mail: [reverdyj@tcd.ie](mailto:reverdyj@tcd.ie)

M. Koutsombogera  
e-mail: [koutsomm@scss.tcd.ie](mailto:koutsomm@scss.tcd.ie)

C. Vogel  
e-mail: [vogel@scss.tcd.ie](mailto:vogel@scss.tcd.ie)

vovement in an interaction, where involvement is defined as the action to achieve mutual understanding.

Two previous studies conducted using the data of the Edinburgh Map Task Corpus [5] reported that in task-based interaction, repetitions that occur with a higher probability than chance have an impact on task success [6, 7].

This impact, when defined social-psychological factors are known (such as gender, familiarity and eye-contact), reveals that above chance repetitions of the dialogue partner are in a number of conditions related to higher success in a given task. These results are consistent with other findings in human-to-human communication that discuss how alignment is reflected at different linguistic levels, for example in lexical choices and syntactic structures [8–10] or prosodic features [11].<sup>2</sup>

The achievement of mutual understanding is never entirely certain; however, interlocutors can achieve a state in which they lack direct evidence of misunderstanding [14], i.e. achieve a level of understanding that is adequate to accomplish a given task. Constraints imposed by the external world are shared by all, even if perspectives on the world may differ, and therefore, the pragmatics of some tasks may enable task success independently of interlocutors achieving mutual understanding. The map task, from its inception [15], provides a context in which task-based success is not possible without some level of success in communication. However, arriving at shared denotations is only one dimension of successful communication. Another is in achieving consensus. Here we report on repetition effects in dialogues designed to include joint idea formulation and consensus building.

Measuring repetitions in dialogues at different levels of linguistic representation can inform the assessment of involvement and possible mutual understanding between the participants. Mutual understanding is not perfectly indexed by repetition counts, but these counts provide useful information. In this work, we use a data set consisting of dialogues among three participants, involving two conversational roles: the role of the player, attributed to two participants, and the role of the facilitator, attributed to the third participant. It is known that facilitator style has an impact on dialogue outcomes (e.g. “supportive” vs. “oppositional” on qualities of reflection [16], “task oriented” vs. “socially oriented” on perceptions of efficacy [17]). We explore dialogue data that arise in a context in which the task definition (which is to construct and agree upon hypotheses of subjective opinions of independently surveyed anonymous groups) supports a fusion of task and social orientation to facilitation. Our goal is to assess whether measures of repetition in the players’ speech can be linked to occurrences and kinds of facilitators’ speech, which mainly consists in providing feedback to the players. Given the playful nature of the task which is modelled on that of a television game show, *Family Feud*, and the approach to facilitation provided, we do not actually expect substantial quantities of patently negative feedback from

---

<sup>2</sup>The extent of repetition across levels of linguistic repetition remains a subject of exploration: [12], for example, did not find repetition of structure to exceed chance, while repetition of lexis did exceed chance; [13] refine this further noting self-repetition of structure to exceed chance, but not repetition of others’ structures.

facilitators.<sup>3</sup> Rather, in the context of the task, we imagine that contributions from the facilitator will either tend towards introducing participants to discrete phases of the interaction (and therefore be deemed neutral) or will be positively encouraging. However, positive encouragement for a task/social style of facilitation is a natural response to a perception of communication gone awry or task failure in some other sense. Therefore, for this data set, we expect fewer positive and more neutral facilitator contributions in contexts where interlocutors experience success, and more (encouraging) positive contributions in contexts where interlocutors experience difficulty. The success of an interaction depends on many factors and is also determined by the point of view adopted during the success assessment. Obtaining objective measures of success regarding human communication is delicate. We believe that the feedback given by the facilitator represents a continuous assessment of the ongoing success of the interaction and the success of the task the two players were given to achieve. We therefore observed if repetitions happened above chance within each dialogue section (defined in a manner described below)—we deem amounts of repetitions that are above chance to signal alignment among players. We also investigate whether the degree of alignment among the two players in a dialogue is reflected in the facilitator's feedback.

## 32.2 Data set

The study presented here exploits the MULTISIMO corpus [18], a multimodal corpus consisting of 23 sessions of collaborative group interactions, where two players work together to provide answers to a quiz and are guided by a facilitator, who monitors their progress and provides feedback or hints when needed. This data set addresses multiparty collaborative interactions and aims to provide tools for measuring collaboration and task success based on the integration of the related multimodal information and at informing the creation of behavioural models. The sessions were carried out in English, and the task of the players was to converse with each other with the aim of estimating and agreeing on the three most popular answers to each of three questions, and rank their answers from the most to the least popular.<sup>4</sup> The corpus consists of synchronised audio and video recordings, and its overall duration is approximately 4 h, with an average session duration of 10 min. The average age of the participants is 30 years old, and their gender is balanced (25 females, 24 males). Eighteen nationalities are represented among participants, one-third of them being native English speakers.

---

<sup>3</sup>Perhaps with a substantially larger data set (than the 23 dialogues collected), task-diverting interpersonal conflict or enchantment might have emerged.

<sup>4</sup>Correctness of the answers and their rankings is determined by responses to an independent survey of 100 people (see <http://familyfeudfriends.arjdesigns.com/> last accessed 11.05.2018).

### 32.2.1 *Conversational Roles*

Each group consists of three members, who collaborate with each other in a quiz: two players and one facilitator. Out of the 49 corpus participants, 3 were designated as facilitators, and 46 were assigned the role of players and were randomly paired in 23 groups. Facilitators coordinated those discussions, i.e. provided the instructions of the game and confirmed participants' answers, but also assisted participants throughout the session and encouraged them to collaborate. Facilitators were briefly trained before the session recordings, i.e. they were given the quiz questions and answers and they were instructed to monitor the flow of the discussion.

The facilitator role is critical in the set-up design, considering that it is a role to be modelled for an embodied conversational agent that would coordinate group interaction and would help participants achieve their goals. In this respect, the facilitator role was designed to enable the extraction of behavioural cues for the development of an agent responsible for managing the interaction and choosing actions that maximise the collaboration effort and the performance of the group participants.

### 32.2.2 *Annotation Process*

All corpus sessions were fully transcribed by two annotators using Transcriber.<sup>5</sup> The transcription consists in the segmentation of the audio signal in speaker turns, the transcription of speech and the segmentation of dialogue in 11 sections, i.e. introduction, questions 1, 2 and 3, categorisation of each question in 2 parts (namely answering phase and ranking phase) and closing. Transcripts were then imported into the ELAN annotation editor,<sup>6</sup> so that all the information recorded in the transcript was visible and further editable.

For the purpose of the present study, we disregard the introductory and closing parts of each session, and we focus on the following section types:

- *Full*: consisting of the three questions of the quiz as a whole (23 sections in total)
- *Question*: each of the three questions, cutting *Full* into 3 parts (69 sections in total)
- *Answer*: the answering phase within each *Question* (69 sections in total)
- *Ranking*: the ranking phase within each *Question* (69 sections in total).

The *Full* section embeds the three *Questions*; the *Question* embeds the *Answer* and *Ranking* phases, while those last two are mutually exclusive. The facilitator's turns occurring during the question-answer sequences were further annotated for their feedback type. To this end, two annotation layers were introduced: the first annotation layer includes the values of *positive*, *neutral* and *negative* feedback. Table 32.1 presents the mean and median duration of each section type as well as the number

---

<sup>5</sup><http://trans.sourceforge.net/> last accessed 27.04.2018.

<sup>6</sup><https://tla.mpi.nl/tools/tla-tools/elan/> last accessed 27.04.2018.

**Table 32.1** Section type mean ( $\mu$ ) and median ( $M$ ) duration (in minutes); and number ( $n$ ) of turns, turn mean ( $\mu$ ) and median ( $M$ ) duration (in minutes) and mean ( $\mu$ ) and median ( $M$ ) number of words per feedback type

Sections		Feedback						
Section type	Duration ( $\mu$ )	Duration ( $M$ )	Feedback type	Turns ( $n$ )	Turn length ( $\mu$ )	Turn length ( $M$ )	Words ( $\mu$ )	Words ( $M$ )
Full	8.50	8.51	Positive	1062	1.01	0.48	141	108
Questions	2.59	2.45	Negative	360	1.24	1.06	69	70
Answer	1.58	1.31	Neutral	1154	1.40	1.16	391	298
Ranking	1.20	0.47						

**Table 32.2** Annotation values for the facilitator's feedback type and subtype

Type	Subtype	Subtype description	Subtype example
Positive	General	Positive feedback that the participants are doing well	Great job, well done!
	Confirmation	Confirms the correctness of the answers	That was the right ranking
Negative	General	Negative feedback while they discuss possible answers	It doesn't have to do with food
	Disconfirmation	Disconfirms replies that are not correct	Unfortunately you didn't get this one
Neutral	Elaboration	Provides helping cues	It is related to food, but think of a different category of food
	Feedback elicitation	Poses direct or indirect questions	Is that your final decision?
	Topic change	Manages the sequence of questions	Now let's move on to the second question

of turns per feedback type, the turn mean and median duration, and the mean and median number of words encountered in each feedback type.

Feedback values are further refined at a secondary level, that is, feedback subtypes. The annotation values of feedback type and subtype are listed in Table 32.2, together with a brief description and an example for each value from the corpus.

The facilitator's feedback was coded by one annotator, and annotations were edited for validity and consistency issues by a second annotator. The annotation task resulted in 2576 annotations, and their distribution per feedback type is presented in Table 32.3. The distribution is detailed per Question section (Q1, 2, 3), per answer [A] and per ranking [R] phase.

The most frequent value is that of *neutral* feedback, indicating that the facilitator often intervenes to help the participants by providing hints and examples. Almost equal to the number of the *neutral* values are the occurrences of *positive* feedback,

**Table 32.3** Distribution of feedback type values in section types full, question (Q), answer, ranking

	Positive		Negative		Neutral	
	n	%	n	%	n	%
Full	1062	41	360	14	1154	45
Q1	380	36	89	25	402	35
Q2	345	32	123	34	377	33
Q3	337	32	148	41	375	32
Answer	766	72	205	57	838	73
Ranking	296	28	155	43	316	27

implying that the facilitator not only confirms the correct answers, but also has a positive disposition towards participants, aiming at their successful results. This positive disposition created delicate cases to annotate, and the annotators often exploited multimodal information to disambiguate certain instances, that is, the speech prosody and facial expressions of the facilitator. For example, cases such as “They are all very good answers but they’re not the popular answers.” were considered as negative feedback, even if the facilitator’s words are positive in the first clause, because the audio and visual information indicated otherwise. Moreover, there is no significant difference in the quantity of expressed feedback among the three questions. However, it seems that the majority of feedback responses for all positive, negative and neutral types are occurring in the answering phase, where players need to identify the three most popular answers.

### 32.3 Method

We counted the repetition of tokens of a contribution and the immediately preceding contributions [19, 20]. A REGISTER was created for each participant, containing her or his most recent contribution. For each dialogue turn, the speaker’s content is compared with the speaker’s own register and with those of the other participants in order to count self-repetition (SELFSHARED) and other-repetition (OTHERSHARED), respectively. We note that among OTHERSHARED, the repetitions counted are those of the other player and of the facilitator.

The dialogues were cut by sections: *Full*, *Question*, *Answer* and *Ranking*, as mentioned before, to observe if the section type, by their nature and length, show variations. In each dialogue section, the turns were then randomly re-ordered ten times. This resulted in ten randomly ordered sections where other-repetitions and self-repetitions were counted again. A preprocess labelling, designed to measure five different levels of linguistic repetition types, was applied: (i) Token, (ii) Lemma, (iii) Part-Of-Speech (POS), (iv) a combination of Lemma with POS, and (v) a com-

bination of Token with POS. Data from the MULTISIMO corpus were labelled with the TreeTagger [21]. We wish to see how the repetitions differ at various linguistic representation levels. Tokens were counted as  $n$ -grams, up to  $n = 5$ .

We determine whether significantly more repetition appears in the actual dialogue sections than in the randomised dialogue sections. A single-step Tukey's HSD multiple comparison test was performed using a generalized linear model with a binomial error family [22]. The statistical null hypothesis  $H_0$  for the tests was as follows:

$$H_0 : \text{Random.Speaker.Level} - \text{Actual.Speaker.Level} \geq 0$$

To observe if differences appeared for sequences of  $n$ -grams, we also tested for  $n$ -grams with  $n > 1$ , (N2+). The null hypothesis ( $H_0$ ) states that the difference between the amount of repetitions in the randomised dialogue sections and the actual dialogue sections should equal (or exceed) zero if repetitions are due to chance. From a natural Zipfian distribution of linguistic forms, repetitions will happen by chance. Repetitions in randomised dialogues might exceed that of actual turn orderings because of the frequency of closed-class lexical types. However, if the null hypothesis is rejected, a communicatively meaningful role of repetitions could be assumed. We categorise a dialogue or a dialogue section that leads to null hypothesis rejection as having *Above chance* repetitions.

Our hypothesis is that the alignment detected by the method should be reflected in the facilitator's feedback as follows: a lack of mutual understanding signalled by a lack of alignment between the players is expected to correspond to a high number of occurrences of negative or neutral feedback from the facilitator; conversely, we expect a dearth of positive feedback where there is ample evidence of participant alignment. Although we have annotated three categories of feedback, we group two together to form a binary opposition: positive versus non-positive feedback. This approach towards neutral feedback (which consists of the facilitator's elaboration or elicitation of players' contribution to the dialogue) is consistent with a view of the content of neutral feedback as guidance from the facilitator: if participants are seen as needing guidance, then something is likely not proceeding perfectly. The following section (Sect. 32.4) focuses on the repetition behaviours of the two participants and the possible interaction with the facilitators' feedback.

## 32.4 Results and Discussion

### 32.4.1 Overview of Repetition Types and Dialogue Sections

For the *Full* dialogues, at the Level Token, following a threshold of ( $p \leq 0.05$ ), the Null Hypothesis was rejected 30 times over 46 for OTHERSHARED and 27 over 46 for SELFSHARED, for all  $n$ -grams (1–5) for the two players, which shows that there was a slightly higher proportion of significant OTHERSHARED repetitions in the corpus. The detail of the rejections of  $H_0$ , per dialogue section, linguistic representation

**Table 32.4** Rejections of  $H_0$  for OTHERSHARED and SELFSHARED, at all  $n$ -grams and N2+, (P. Rej.: Possible Rejections per cell), see Sect. 32.4

All $n$ -grams						N2+ ( $n$ -grams, $n > 1$ )				Total	
Level	Tok	Lem	LemP	POS	TokP					Total	Total
						Tok	Lem	LemP	POS		
<i>OtherShared</i>											
Full	30	31	29	14	30	134	19	23	19	21	96
Question	37	42	47	20	51	197	20	17	20	7	46
Answer	19	21	22	5	24	91	8	9	7	20	138
Ranking	10	6	13	3	13	45	3	2	1	0	138
<i>SelfShared</i>											
Full	27	25	28	12	29	121	27	27	28	12	123
Question	20	17	20	7	20	84	15	19	15	11	46
Answer	6	4	4	2	4	20	6	4	4	2	138
Ranking	34	29	31	13	31	138	3	2	1	0	7

level and repetition types can be found in Table 32.4. For the *Full* dialogues, in each case the Null Hypothesis can potentially be rejected 46 times, as there are 2 speakers in 23 dialogues. For the other dialogue sections, the Null Hypothesis can potentially be rejected 138 times, as each section is repeated 3 times. We observe that for OTHERSHARED repetitions, the rate of rejection of the null hypothesis is the highest in the *Full* dialogues and decreases as the dialogue sections shorten (see Table 32.1). For SELFSHARED repetitions, the rate of rejection is also the highest for the *Full* dialogues; however, we note that the section *Ranking* contains a higher rate of rejections despite being the shortest dialogue section. Since this pattern is not present in longer sequences of  $n$ -grams (N2+), we can conclude a high rate of lexical unigram repetitions in those sections.

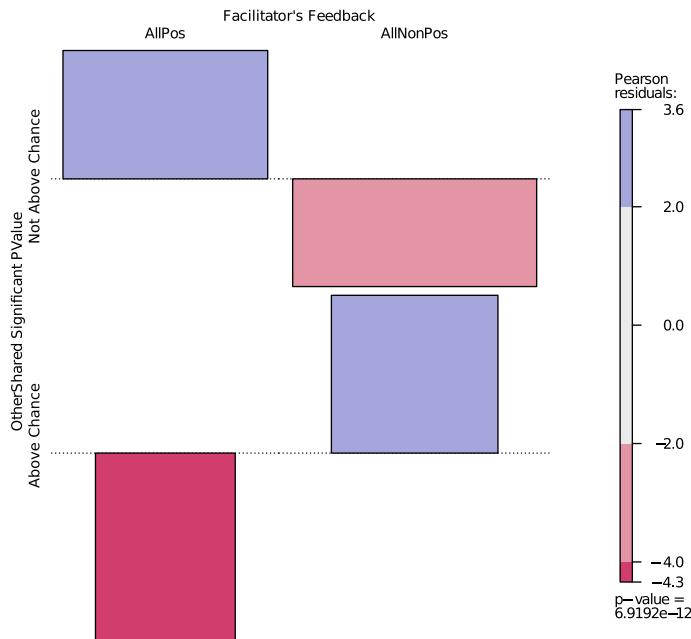
### 32.4.2 *Above Chance Repetitions and Facilitators' Feedback*

We adopted a binary classification of the facilitators' feedback: positive and non-positive (negative and neutral), as described in Sect. 32.3. Figure 32.1 shows that when there are *Above chance* OTHERSHARED repetitions, the amount of positive feedback is less than one would expect and non-positive feedback is in greater amount than one would expect if there were no interaction between the categories of facilitator feedback and the degree of repetition.<sup>7</sup> Conversely, where OTHERSHARED repetitions are at a level that is *Not Above chance*, there is more positive feedback than one would expect and less non-positive feedback than one would expect if there were no interaction between feedback type and degree of repetition.

Using Mann–Whitney–Wilcoxon tests, we observed the following for the *Full* dialogues: the amount of positive feedback found in the dialogues categorised as *Above chance* OTHERSHARED repetitions was significantly different from the amount found in *Not Above chance* ( $W = 4092, p = 2.487e-06$ ), with *Above chance* accompanied by less positive feedback ( $\bar{x} = 43.11$ ) and *Not Above chance* accompanied by more positive feedback ( $\bar{x} = 50.44$ ). The Mann–Whitney–Wilcoxon test applied to the amount of non-positive feedback between *Above chance* and *Not Above chance* *Full* dialogues did not return a significant result. The same observations were made for the *Question* sections: the amount of positive feedback found in the dialogues categorised as *Above chance* OTHERSHARED repetitions was significantly different from the amount found in *Not Above chance* ( $W = 39046, p = 5.33e-05$ ), relating *Above chance* to on average less positive feedback ( $\bar{x} = 14.21$ ) and *Not Above chance* to on average more positive feedback ( $\bar{x} = 15.86$ ). No significant difference was found for the amount of non-

---

<sup>7</sup>Figure 32.1 presents an association plot of residuals, determined by the difference between observed and expected values, using a loglinear model [23]: the magnitude of a box corresponds to the magnitude of residuals; shading intensity encodes significance (residuals between 2 and 4 are significant at the  $p < 0.05$  level); boxes projecting up from the horizontal line correspond to divergences in excess of expectations and boxes projecting down from the horizontal convey the extent to which observations are fewer than expected, where expectations are those of the null hypothesis, which is that there is no interaction among the categories examined.



**Fig. 32.1** Association plot of significant OTHERSHARED *p*-values (*Above chance* | *Not above chance*) and facilitator's feedback (All positive | All non-positive) across the *Full dialogues*

positive feedback between *Above chance* and *Not Above chance*. With respect to the *Ranking* sections, the amount of positive feedback found in the dialogues categorised as *Above chance* OTHERSHARED repetitions was **not** significantly different from the amount found in *Not Above chance* ( $W = 13602, p = *0.4768*$ ), relating *Above chance* to on average less positive feedback ( $\bar{x} = 3.95$ ) and *Not Above chance* to on average more positive feedback ( $\bar{x} = 4.31$ ). No significant difference was found for the amount of non-positive feedback between *Above chance* and *Not Above chance*. In the *Answer* section type, the amount of positive feedback was not found significantly different depending on *Above chance* repetitions, while the amount of non-positive feedback was ( $W = 32320, p = 0.004$ ). The *Answer* sections with *Above chance* levels of repetition have more non-positive feedback ( $\bar{x} = 17.62$ ), and the sections with *Not Above chance* repetition have less non-positive feedback ( $\bar{x} = 14.56$ ).

A null hypothesis expects no interaction between facilitator's feedback types and the degree of OTHERSHARED repetitions by the dialogue participants. The results breach this expectation. Facilitators respond with more non-positive feedback where *Above chance* OTHERSHARED repetitions are observed, and more positive feedback where *Not Above chance* OTHERSHARED repetitions are observed, than one would expect in either case with no interaction. If significant OTHERSHARED repetitions signal mutual understanding, facilitators are more likely to respond with non-positive than with positive feedback to signals of mutual understanding. Those results suggest that the

facilitators provide more encouragement where interactions are seen as difficult, and less encouragement when interactions are perceived as successful.

That the results observed for the full dialogues and for each of the three question sections (as separate parts of each full dialogue) do not apply identically within each of the components of the question section (answers and rankings, as separate parts within each question) is noteworthy and may follow from the task-related differences between idea generation and idea ranking. These task-related differences may be anticipated to be different both in the level of guidance needed from facilitators in order to achieve the task and in the linguistic-pragmatic need for interlocutors to repeat each other. Without reporting the full results of analysing self-repetition here, we note anecdotally that the propensity for significant self-repetition also varies between sections (see Table 32.4).

## 32.5 Conclusion

This paper described the use of a method of interaction analysis based on repetitions and its use in a newly created task-oriented corpus. A specific set of annotations of the facilitator feedback was designed to provide an estimation of the participants' social and task success. The relation appearing between above chance repetitions and positive and non-positive feedback confirms our general hypothesis that repetitions detected from this method reflect a degree of interaction success, as it is echoed in the verbal and non-verbal behaviours of an interactional facilitator. Where participants' repetition of others contributions is significant, which we interpret as providing evidence of having secured mutual understanding, there is significantly less facilitator encouragement; where participants' repetitions of others is not so pronounced, which we interpret as failing to show evidence of mutual understanding, there is significantly more facilitator feedback that encourages engagement.

**Acknowledgements** The research leading to these results has received funding from (a) the ADAPT Centre for Digital Content Technology, funded under the SFI Research Centres Programme (Grant 13/RC/2106) and co-funded under the European Regional Development Fund, and (b) the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 701621 (MULTISIMO).

## References

1. Clark, H.H., Brennan, S.E., et al.: Grounding in communication. *Perspect. Socially Shared Cogn.* **13**(1991), 127–149 (1991)
2. Tannen, D.: Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse, vol. 26. Cambridge University Press, Cambridge (2007)
3. Colman, M., Healey, P.: The distribution of repair in dialogue. *Proc. Ann. Meet. Cogn. Sci. Soc.* **33**, 1563–1568 (2011)

4. Cushing, S.: Fatal words: Communication Clashes and Aircraft Crashes. University of Chicago Press, Chicago (1994)
5. Anderson, A.H., Bader, M., Bard, E.G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H.S., Weinert, R.: The HCRC map task corpus. *Lang. Speech* **34**(4), 351–366 (1991)
6. Reverdy, J., Vogel, C.: Linguistic repetitions, task-based experience and a proxy measure of mutual understanding. In: Proceedings of CogInfoCom 2017, pp. 395–400. IEEE, Debrecen, Hungary (2017)
7. Reverdy, J., Vogel, C.: Measuring synchrony in task-based dialogues. In: Proceedings of INTERSPEECH'17, pp. 1701–1705. ISCA, Stockholm, Sweden (2017). <https://doi.org/10.21437/Interspeech.2017-1604>
8. Branigan, H.P., Pickering, M.J., Cleland, A.A.: Syntactic co-ordination in dialogue. *Cognition* **75**(2), B13–B25 (2000)
9. Garrod, S., Anderson, A.: Saying what you mean in dialogue: a study in conceptual and semantic co-ordination. *Cognition* **27**(2), 181–218 (1987)
10. Reitter, D., Moore, J.D.: Predicting success in dialogue. In: Proceedings of ACL 2007, pp. 808–815. Association for Computational Linguistics, Prague, Czech Republic (2007)
11. Giles, H., Coupland, J., Coupland, N.: Contexts of Accommodation: Developments in Applied Sociolinguistics. Cambridge University Press, Cambridge (1991)
12. Howes, C., Healey, P.G.T., Purver, M.: Tracking lexical and syntactic alignment in conversation. In: Proceedings of the 32nd Annual Conference of the Cognitive Science Society, pp. 2004–2009 (2010). <http://mindmodeling.org/cogsci2010/papers/0484/>
13. Healey, P.G.T., Purver, M., Howes, C.: Divergence in dialogue. *PLOS one* **9**(6), e98,598 (2014). <https://doi.org/10.1371/journal.pone.0098598>
14. Taylor, T.J.: Mutual Misunderstanding: Scepticism and the Theorizing of Language and Interpretation. Duke University Press, Durham (1992)
15. Brown, G., Anderson, A., Shillcock, R., Yule, G.: Teaching Talk: Strategies for Production and Assessment. Cambridge University Press, Cambridge (1985)
16. Cacciamani, S., Cesareni, D., Martini, F., Ferrini, T., Fujita, N.: Influence of participation, facilitator styles, and metacognitive reflection on knowledge building in online university courses. *Comput. Educ.* **58**(3), 874–884 (2012)
17. van Dolen, W., de Ruyter, K., Carman, J.: The role of self- and group-efficacy in moderated group chat. *J. Econ. Psychol.* **27**(3), 324–343 (2006)
18. Koutsombogera, M., Vogel, C.: Modeling collaborative multimodal behavior in group dialogues: The MULTISIMO corpus. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA), Paris, France (2018)
19. Vogel, C.: Attribution of mutual understanding. *J. Law Policy* **21**(2), 377–420 (2013)
20. Vogel, C., Behan, L.: Measuring synchrony in dialog transcripts. Cognitive behavioural systems. In: Lecture Notes in Computer Science, vol. 7403, pp. 73–88. Berlin, Heidelberg: Springer (2012)
21. Schmid, H.: Probabilistic part-of-speech tagging using decision trees. In: Proceedings of the International Conference on New Methods in Language Processing, pp. 154–164. Manchester, UK (1994)
22. Bretz, F., Hothorn, T., Westfall, P.: Multiple Comparisons Using R. CRC Press, Boca Raton (2016)
23. Meyer, D., Zeileis, A., Hornik, K.: The strucplot framework: visualizing multi-way contingency tables with VCD. *J. Stat. Softw.* **17**(3), 1–48 (2006)

# Chapter 33

## An Experiment on How Naïve People Evaluate Interruptions as Effective, Unpleasant and Influential



Ida Sergi, Augusto Gnisci, Vincenzo Paolo Senese and Angelo Di Gennaro

**Abstract** We conducted an experimental study on 144 participants to evaluate how naïve people evaluate different kinds of interruptions. We manipulated the point in which interruption occurs (early, late and no interruption) and the type of interruption (change subject, disagreement, clarification and agreement) on pre-built, acted and audio-recorded dialogues. Then, participants evaluated how much each interruption was effective, unpleasant and influential. The main results show that (a) with some exceptions, early and late interruptions were evaluated as more influential and unpleasant than control, and early interruption more unpleasant than late interruption; (b) change subject was more unpleasant and less effective than disagreement, and disagreement than clarification but only in not interruptive turn-taking while there was no difference between clarification and disagreement for effectiveness and unpleasantness; (c) the point was more important than type in determining the evaluation of interruption; (d) the perception of a turn-taking as an interruption was correlated with unpleasantness and only partially with influence; (e) all over the study, the effectiveness went in the opposite direction with respect to our hypothesis. Results are interpreted at the light of the literature on the effects of interruption and of politeness theory applied to interruption.

### 33.1 Introduction

The literature on interruption shows that many different criteria may coexist to generate different kinds of interruptions [4, 9, 25, 27]. Of course, one of the basic criteria is the *completeness of the sentence* that has been interrupted [5] with respect to the transition relevance point [26]: if interruption intrudes early, the interruption is deep, if late, shallow [27]. Recent studies provide further support to this evidence, sustaining that there is an incremental semantic update [6, 7]: each new piece of syntax in a sentence provides a new piece of semantics, and therefore, the more an interruption is

---

I. Sergi (✉) · A. Gnisci · V. P. Senese · A. Di Gennaro

Department of Psychology, University of Campania “L. Vanvitelli”, Caserta, Italy  
e-mail: [ida.sergi@unicampania.it](mailto:ida.sergi@unicampania.it)

early, the tougher will be evaluated [14, 15]. Another important criterion established by the literature is the *semantic link* of the interruptive turn with the interrupted turn [11], which distinguishes among different *types of interruptions* [21, 22]. Change subject is the strongest and toughest interruption because it intrudes in the turn of the interlocutor and changes the topic (absent or marginal semantic link). Instead, disagreement and agreement interruptions maintain a central semantic link with the interrupted turn. However, the two are very different: the first rejects the contents of the preceding turn, while the other supports them. Finally, clarification is interlocutory because the interlocutor interrupts to have more information on what the interrupted conversant was saying [19].

The two criteria above described can be regarded as *objective* (together with others as unsuccessful or simplex/complex interruptions; see [25]) as they are established by researchers and analysts. However, many studies in the last 30 years have studied how interruptions are perceived and evaluated by the naïve people who endure them or who listen to them when other people converse [13]. These studies focus on the effect of interruptions, usually using the toughest interruption, the deep and/or intrusive [3–5, 8, 10, 11, 16, 17, 20, 23, 24; cf. 15]. In terms of our vocabulary, it is the early, change subject interruption [11]. Even if with alternate results, these studies demonstrated that who interrupts is regarded as higher in status, dominant, competent, controlling, influent and effective than those interrupted or who do not interrupt [8]; as well, he/she is regarded as less likeable, friendly, pleasant [8, 24]. Opposite consideration for one who is interrupted or who does not interrupt: he/she is regarded as having few power and status but is more likeable.

Politeness theory claims that whatever act can be judged on the base of how it threatens the “face” of others or themselves in conversation [1, 2, 12]: disagreement, refusal, evasion, etc., and even interruption are, at different degree, threats to the face [10, 11]. In general, the face is the image that one wants to give to others, in terms of appreciated and approved social characteristics [12]. When applied to interruption, politeness theory demonstrated powerful representing a general framework able to predict all the results presented in the literature on the effect of interruptions [11, 18]. Indeed, the tougher is an interruption, the stronger the threat to the face. Therefore, it can be expected that change subject and disagreement, on a side, and early rather than late interruptions will be regarded as negative, impolite, unpleasant because they threaten more the face of the interrupted conversant [10, 11, 18].

Different studies have used different dependent variables. In brief, some of them focus on how a determined turn-taking (among which interruption) is perceived subjectively as an interruption [4]. These studies refer to *perception* or *recognition of interruptions*. Others focus on how the different kinds of interruptions are valued in terms of positivity, politeness, etc., and they refer to *evaluation* or *valence* [3, 5, 8, 10, 11, 16, 17, 20, 23, 24; cf 15].

This study falls in this literature and complements another study on the same sample [i.e. 10]. Its results show that naïve observers perceived more as interruption and evaluated more negatively change subject rather than disagreement, and disagreement rather than clarification interruptions. Instead, clarification was similar to agreement for many respects. As well, naïve observer perceived more as inter-

ruption and evaluated more negatively early rather than late interruptions, and late interruptions rather than not interruptive turn transition. Naïve people used more the point than the type of interruption for identifying and evaluating the interruptions. Finally, that study, for the first time, provided evidence that perception and evaluation of a conversational transition were correlated (no causation): in the interruptive and even in not interruptive conditions, the more a turn-taking was perceived as an interruption, the more it was evaluated as negative. In that research, we asked people to evaluate also the interruptions in terms of effectiveness, impoliteness and influence but we never published the relative results. Compared to the results of the literature, this study adds new information about different kinds of interruptions, not only the deep and/or intrusive one. Further on, we used two criteria, point and type of interruption, and checked their main and interactive effects on perception and evaluation of interruptions which gave also the chance to weight which of them is more important in the subjective perception and evaluation of the interruption. Finally, a further development of this research is the joint study of perception and evaluation of interruptions so that their relationships can be explored.

We think that these novel results can profitably be integrated with the original results in order to give a wider and complete picture of how naïve people perceive interruptions.

This paper aims at establishing how naïve people evaluate different types of interruptions in terms of effectiveness, unpleasantness and influence. We combined two criteria identified by the literature, the point in which an interruption occurs and the type of interruption based on semantic link between the interrupted turn and the interrupting turn. We have built a dialogue between two conversants that was audio-recorded. Then, we have asked naïve people to listen to each dialogue and evaluate how each interruption was effective, unpleasant and influential.

### 33.1.1 *Hypotheses*

Combining the literature on the effect of interruptions [cf. 15], in particular the studies on the opposite evaluation of status and likeability [8], and the ones on politeness theory applied to interruption [11, 12], allows making the following predictions.

*H<sub>1</sub> on point in which interruption occurs:* the earlier an interruption will occur, the stronger its effect. Thus, early rather than late, and late rather than not interruptive turn-taking (control) will be evaluated as more unpleasant but also more effective and influential.

*H<sub>2</sub> on type of interruption:* the tougher an interruption will be, the stronger its effect. Thus, change subject with respect to disagreement interruption, disagreement with respect to clarification interruption, clarification with respect to agreement interruption will be evaluated as more unpleasant but also more effective and influential.

*H<sub>3</sub> on relation between perception and unpleasantness, effectiveness and influence:* we will expect that the more a turn transition will be perceived as an interruption, even if it is not, the more it will be evaluated as unpleasant, effective and influential.

## 33.2 Study

### 33.2.1 Method

**Sample.** One-hundred and forty-four Italian first-year undergraduate students (86.1% female), from the University of Campania “Luigi Vanvitelli” (mean age = 21.58, SD = 5.96), took part in the experiment. They were first informed and then invited by the teacher of one of their courses to participate in the experiment as part of a practical exercise.

**Procedure.** We built as stimulus a dialogue that contained six turns between two conversants A and B on daily topic (trip to the park). The first four turns were always the same, and the fourth turn was a question (B: It was a sunny day, wasn’t it?). After a brief answer (A: Yes), the conversant A kept to talk producing two sentences linked by “and”. A could be interrupted by B at the end of the first sentence (early interruption) and at the end of the second (late) or his/her turn could be synchronized to end of the second sentence (no interruption—control condition). This way we manipulated the first independent variable, the point in which the interruption occurred. The second independent variable—type of interruption—was manipulated varying the semantic link of the A’s turn with the following interruptive B’s turn. It provided four types of interruptions: change subject, disagreement, clarification and agreement. The dialogues were therefore 12, as many as the combinations of the two independent variables ( $3 \times 4$ ). They were administered within subject by means of a balanced Latin square for controlling the effects of order and sequence. Finally, dialogues were built providing same-sex (M-M and F-F) and mixed-sex dyads (M-F and F-M) for controlling the effect of sex of interrupter and of interruptee that was equally distributed.

**Measures.** After each dialogue, five questions were asked to each participant. In brief, the first question was “How much did it seem to you that the last conversant took over, preventing the other conversant from concluding what he/she was saying?” and the other four questions were “How much did the last turn exchange seem positive/effective/unpleasant/influential to you? (In Italian: positivo, efficace, sgradevole, influente)” (1 = Not at all–7 = Completely). While the first two questions were addressed in a preceding paper [10], here we want to address the remaining ones.

**Data analysis.** A within-subject  $3 \times 4$  point-by-type ANCOVA was performed with sex of interrupted and sex of interrupter as between-subject covariates. In case of significant effects, planned comparisons were executed with Bonferroni correction (paired *t*-test).

### 33.3 Results

#### 33.3.1 *The Effect of Interruptions in Terms of Effectiveness, Pleasantness and Influence*

The ANCOVA  $3 * 4$  on effectiveness of interruption provided significant main effect of point ( $F(2, 140) = 19.86, p < 0.001, \eta^2 = 0.22$ ), type ( $F(3, 139) = 7.78, p < 0.001, \eta^2 = 0.14$ ) and of the interaction point\*type ( $F(6, 136) = 5.99, p < 0.001, \eta^2 = 0.21$ ).<sup>1</sup>

The ANCOVA  $3 * 4$  on unpleasantness of interruption provided significant main effect of point ( $F(2, 140) = 145.43, p < 0.001, \eta^2 = 0.68$ ), type ( $F(3, 139) = 31.88, p < 0.001, \eta^2 = 0.1$ ) and of the interaction point\*type ( $F(6, 136) = 7.19, p < 0.001, \eta^2 = 0.24$ ).<sup>2</sup>

The ANCOVA  $3 * 4$  on influence of interruption provided significant main effect of point ( $F(2, 140) = 3.62, p < 0.05, \eta^2 = 0.05$ ) and type ( $F(3, 139) = 3.80, p < 0.05, \eta^2 = 0.08$ ). The interaction point\*type was not significant ( $F(6, 136) = 0.27, p = 0.95, \eta^2 = 0.01$ ).<sup>3</sup>

Planned comparisons for the main effect of point (Table 33.1) show that early interruption was more unpleasant than late interruption but as well effective and influential; late is more unpleasant but more influential and less effective than control.

Planned comparisons for the main effect of type (Table 33.2) show that change subject was regarded as less effective and pleasant than disagreement interruption but as well influent; disagreement was regarded as less effective and pleasant than clarification interruption but as well influent; clarification was regarded as more influential than agreement but as well as effective and pleasant.

Planned comparisons for the point\*type interaction effect were executed among different types of interruptions within each point (Table 33.3). First, no significant difference arose between clarification and agreement interruptions: clarification is effective and pleasant as well as agreement. Second, no significant difference arose when comparing the effectiveness of the remaining types of interruptions when the interruption was early or late. A similar pattern was observed for unpleasantness. No

**Table 33.1** Planned comparisons, means and standard deviations for the main effects of point on effectiveness, unpleasantness and influence of interruptions ( $\alpha_c = 0.025$ )

	Early		Late		Control
Effectiveness	3.17 (0.12)	=	3.27 (0.11)	<	4.77 (0.08)
Unpleasantness	5.73 (0.08)	>	5.45 (0.09)	>	2.69 (0.09)
Influence	4.87 (0.11)	=	4.76 (0.11)	>	4.25 (0.09)

Note > or < means significantly higher or lower at  $\alpha_c = 0.025$ ; = means non-significantly different

<sup>1</sup>Type\*sex of interruptee ( $F(3, 139) = 3.51, p < 0.05, \eta^2 = 0.07$ ) and type\*sex of interrupter ( $F(3, 139) = 2.92, p < 0.05, \eta^2 = 0.06$ ) were also significant.

<sup>2</sup>Type\*sex of interrupter ( $F(3, 139) = 5.71, p = 0.001, \eta^2 = 0.11$ ) was also significant.

<sup>3</sup>Type\*sex of interrupter ( $F(3, 139) = 2.85, p < 0.05, \eta^2 = 0.05$ ) was also significant.

**Table 33.2** Planned comparisons, means and standard deviations for the main effects of type on effectiveness, unpleasantness and influence of interruptions ( $\alpha_c = 0.017$ )

	Change subject		Disagreement		Clarification		Agreement
Effectiveness	3.33 (0.12)	<	3.63 (0.10)	<	3.92 (0.09)	=	4.07 (0.09)
Unpleasantness	5.43 (0.09)	>	4.83 (0.09)	>	4.21 (0.08)	=	4.01 (0.09)
Influence	4.73 (0.12)	=	4.69 (0.10)	=	4.68 (0.09)	>	4.40 (0.09)

Note > or < means significantly higher or lower at  $\alpha_c = 0.017$ ; = means non-significantly different

**Table 33.3** Planned comparisons, means and standard deviations for the interaction effects of point\*type on effectiveness and unpleasantness of interruptions ( $\alpha_c = 0.006$ )

Effectiveness	Change subject		Disagreement		Clarification		Agreement
Early	3.12 (0.17)	=	3.08 (0.15)	=	3.13 (0.15)	=	3.35 (0.13)
Late	3.10 (0.16)	=	2.22 (0.13)	=	3.40 (0.13)	=	3.38 (0.13)
Control	3.77 (0.13)	<	4.58 (0.13)	<	5.22 (0.13)	=	5.48 (0.12)
<i>Unpleasantness</i>							
Early	6.24 (0.11)	=*	5.83 (0.12)	=	5.59 (0.12)	=	5.26 (0.14)
Late	6.01 (0.11)	>	5.51 (0.12)	=	5.16 (0.14)	=	5.02 (0.14)
Control	3.94 (0.17)	>	3.15 (0.15)	>	1.90 (0.12)	=	1.76 (0.12)

Note > or < means significantly higher or lower; = means non-significantly different at  $\alpha_c = 0.006$   
\* $p = 0.008$

significant difference arose when comparing unpleasantness of the remaining types of interruptions when the interruption was early and late, with the exception of change subject being less pleasant in late interruption (and perhaps on early interruption where  $p = 0.008$  just up to  $\alpha_c = 0.006$ ). Third, significant differences between types arose in the not interruption condition (control). In particular, change subject was less effective and pleasant than disagreement interruption; disagreement was less effective and pleasant than clarification. In sum, the effect of type of interruption, and therefore of the semantic link between adjacent turns, seems more effective when the turn is not interrupted.

### 33.3.2 Correlations of Perception of Interruption with Effectiveness, Unpleasantness and Influence

Pearson correlations within each combination of conditions of perception of interruption with evaluation of effectiveness, unpleasantness and influence of interruption are given in Table 33.4. First, all the correlations between perception and unpleasantness of interruption were significant and positive. Therefore, the more naïve observers per-

**Table 33.4** Pearson correlations between perception of interruption and evaluation of effectiveness, unpleasantness and influence of an interruption ( $\alpha_c = 0.0014$ )

Point	Type	Effectiveness	Unpleasantness	Influence
Early	Change subject	-0.09	0.44*	0.14
Early	Disagreement	-0.14	0.41*	0.01
Early	Clarification	-0.18	0.38*	0.11
Early	Agreement	-0.17	0.47*	0.22
Late	Change subject	-0.06	0.52*	0.264*
Late	Disagreement	-0.13	0.27*	0.255
Late	Clarification	-0.05	0.46*	0.41*
Late	Agreement	-0.33*	0.43*	0.12
Control	Change subject	-0.31*	0.46*	0.18
Control	Disagreement	-0.33*	0.55*	0.20
Control	Clarification	0.25	0.72*	0.09
Control	Agreement	-0.40*	0.69*	0.11

\* $p < 0.0014$  (when  $p = 0.0014$ ,  $r = 0.264$ )

ceived a turn transition as an interruption, the more they evaluated it as unpleasant, even in control where there was not an interruption. Second, only in two cases, perception significantly correlated with influence: only in case of late, change subject and clarification interruption, the more naïve observers perceived it as an interruption, the more they evaluated it as influential. Third, in three cases of not interruptive turn transitions and in the late agreement interruption, there was a significant and negative correlation between perception and effectiveness: in these cases, the more naïve observers perceived a turn transition as an interruption, the less they evaluated it as effective.

### 33.4 Discussion and Conclusions

This experiment aimed at understanding if and how naïve people “subjectively” evaluate different kinds of interruptions “objectively” defined by the literature [4, 5, 8–11, 25; cf. 15]. In particular, we have identified two criteria for generating interruptions: the point in which the interruption occurs (early and late, more a no interruption condition as control) and the semantic link among turns (change subject, disagreement, clarification and agreement). We have built some dialogues on a daily topic between the conversants A and B. The last exchange of each dialogue presented a particular combination of the above-discussed interruptions. Then, we have presented the acted and recorded dialogues to naïve people and asked them to evaluate the effectiveness, the unpleasantness and the influence of each interruption.

As regards the point of interruption, early and late interruptions were evaluated as well effective and influential, even if, as expected, early interruptions were more unpleasant than late interruptions. However, late interruptions were evaluated as more unpleasant and influential but less effective than control. In conclusion, early and late interruptions did not differ if not for unpleasantness (partial verification of  $H_1$ ) and interruptions were more unpleasant and influential than control, as expected by  $H_1$ , but less effective, evidence opposite to what we expected.

As regards the main effect of type of interruption on influence, clarification was judged as more influent than agreement, while change subject, disagreement and clarification interruptions were evaluated as having the same influence.  $H_2$  is not supported for influence.

We will discuss directly the interactive point\*sex on effectiveness and unpleasantness effect rather than the main effect of type because the planned comparisons of the interactive effect provide more detailed information. When there were both an early interruption and late interruption, four comparisons out of four for effectiveness and three comparisons out of four for unpleasantness were not significantly different. Instead, uniquely when there was not an interruption (control condition), the above-mentioned differences between change subject, disagreement and clarification arose. *In sum, the differences between types of interruptions, just emerged in the main effect, remain valid but limited to the not interruptive condition.* Uniquely when there was not an interruption, clarification was evaluated as more pleasant and effective than disagreement, and disagreement more than change subject interruption. This is an interesting piece of evidence. When the interlocutor was able to finish his/her sentence (control), then the different types of semantic links among turns had a strong impact; on the contrary, when there were an early interruption or late interruption (but in particular early), the semantic link lost its force. Therefore, interrupting sooner or later a turn is a more important criterion for defining an interruption in terms of pleasantness and effectiveness than the semantic link among adjacent turns and the typologies of interruptions. In conclusion, as regards  $H_2$ , agreement and clarification did not differ in pleasantness and efficacy (contrary to what expected). For the other types of interruptions (change subject, disagreement and agreement), the effect expected by  $H_2$  exists, but only when there is not an interruption, because, as described above, when there is an interruption the criterion of precocity prevails on the semantic one. Finally, for these types of interruptions, while pleasantness goes in the direction expected by  $H_2$ , the effectiveness goes in the opposite direction. We will return on this after having discussed the results on correlations.

While the coefficients of explained variance for the main effects of point and type on influence were low (5 and 8%), the ones on effectiveness (22 and 14%) and, particularly, unpleasantness (68 and 41%) were more substantial. Note also that the indexes show that the point had a major role in explaining the effectiveness and the unpleasantness of the interruptions than the type.

We will discuss now the correlations of perception of interruption with the three different aspects of evaluation. We warn not to interpret them as causations.

The set of significant correlations between perception of interruption and unpleasantness shows that when naïve people perceived a turn transition as an interruption,

they also evaluated it as unpleasant. There was only a not significant case (the Bonferroni correction was strongly conservative because we had 36 correlations). In sum,  $H_3$  was supported, even when there was not an interruption!

$H_3$  finds only partial confirmation for the set of correlations between perception of interruption and their perceived influence. Only two significant out of twelve correlations, always when interruption is late, showed that if naïve people perceived the interruption as interruption, they also evaluated it as influential.

Only four correlations between perception and effectiveness out of twelve are significant, three of them regard control, and only one late agreement interruption. Any case, all the significant correlations went in the opposite direction with respect to what expected in  $H_3$ : when a turn transition was perceived as an interruption, it was evaluated as less effective, even in the control condition. Again, effectiveness went in the opposite direction with respect to what expected.

If we align the results with the ones of the associated study [10] and with the literature [8, 10, 11, 24], we find that perception of interruption, positivity, unpleasantness and, partially, influence behave as predicted in  $H_1$ ,  $H_2$  and  $H_3$ , with some exceptions and limitations, while effectiveness at the contrary. The literature presented in the Introduction sustains that when an interruption is objectively “stronger” (e.g. the deep and/or intrusive interruption), naïve people tend to see the interrupter has having more status but less likeability [8, 24]. Apart from effectiveness, we find general confirmation of this hypothesis. “Stronger” (e.g. early or change subject) interruptions were recognized more as interruption, and evaluated more negatively and unpleasant but more influential. As well, correlations show that perception of interruption correlates positively with negativity and unpleasantness, and with influence, as expected, but negatively with effectiveness. Actually, we expected that effectiveness follows the same pattern of influence. Why does it go in the opposite direction? We can speculate that in this case naïve people could have provided to the different labels of the questions, particularly to influence and effectiveness, an opposite meaning. They recognize that the interruption has an influence because it indeed impedes the other from speaking, but they could have evaluated it as ineffective because, although influential, it does not provide a positive effect. Given that it is negative and unpleasant, it is as well ineffective in supporting the conversation.

From a theoretical point of view, this research has many important advances. First, it studies different types of interruptions, positive and negative ones, together with the kind usually studied in the literature, the deep and/or intrusive one. Along with few recent studies [10, 11], the effect of the interruption studied depends also, for example, by expressing agreement or disagreement, by asking a clarification, etc. As demonstrated by this research, many types of interruptions produce very different effects on the interruptee. Second, we were able to study and manipulate separately two objective criteria—type and point—involved in the formation of the interruption and identifying their main and interactive effects. In this way, we were able to establish, for example, that the point in which an interruption occurs is more important and stronger than the type of interruption to obtain a determined effect. Third, this research studies, at the same time, the perceptive and the evaluative aspects of the interruptions. Therefore, as in [10], we were able to check if and how they

correlated. Finally, the politeness theory [1, 2, 18] applied to the issue of the effects of interruptions [10, 11] is supported by the results of this research. In general, when interruptions threaten the “face” of interrupted people, they are evaluated as negative, unpleasant, etc., and recognized more as interruptions if compared to when they don’t. Thus, the level of face threat is ingrained in the different types of interruptions and, more, in the precocity of the interruption; their effects on recognition and evaluation by naïve people depend, ultimately, by the level of face threat of each interruption.

This study highlights the pros and cons of using different kinds of interruptions. For example, change subject interruptions, although effective, were evaluated as unpleasant. Therefore, there is always a dilemma between controlling the interaction, demonstrating power and status, and appearing impolite, unfriendly or unpleasant. Our results have different possible applications, for example, in TV debates, news interviews, legal examinations, as well as in informal conversation, as family or friends interaction. In the light of our results, dominating the interaction with tough interruptions may reveal counterproductive; the interrupter risks to be seen as negative, unfriendly and unpleasant. However, we think that our results may have an impact also on whatever situation people interact, for example, in teacher–pupil interaction in classroom, in doctor–patient interaction in health settings, in psychotherapeutic interaction in clinical contexts.

This research has some limitations. Two of them regard the way we operationalized the point and the type of interruption. First, although the four categories we have chosen by the criterion of the semantic link among turns are very important, they are not exhaustive. There are many other possible semantic links and, consequently, types of interruptions, which we have not considered. Second, in the Method, we have described how we operationalized the point of interruption. B asks a question, A provides an elaborated answer containing the reply and two other sentences linked by “and” (Yes, 1° sentence “and” 2° sentence), B may interrupt A’s turn at the end of the first sentence (early), at the end of the second (late), or does not interrupt (synchronization at the end of the sentence). The fact that the answer has just been given (Yes) makes the interruption more justified and less “early”. Some scholars assert that a real early interruption should impede the enunciation of the main content of the answer turn [4], that is, in our case the reply to the question. Therefore, other studies should investigate earlier interruptions than the one chosen in this study. A third limitation is methodological. While we balanced the sex of interruptee and of interrupter, we could not balance the sex of participants because females constituted the main part of the sample.

Future researches should solve the limitations above mentioned, in particular the one on the point in which interruption occurs, trying to substitute the types of interruptions (change subject, agreement, etc.), that are generated by a semantic criterion, with the interactional strategies among interlocutors, which are based on the accommodation of the interruptee to the interrupted person.

## References

1. Brown, P., Levinson, S.C.: Universals in language usage: politeness phenomena. In: Goody, E.N. (ed.) *Questions and Politeness: Strategies in Social Interaction*, pp. 56–311. Cambridge University Press, Cambridge (1978)
2. Brown, P., Levinson, S.C.: *Politeness: Some Universals in Language Usage*. Cambridge University Press, Cambridge (1987)
3. Chambliss, C.A., Feeny, N.: Effects of sex of subject, sex of interrupter, and topic of conversation on the perceptions of interruptions. *Percept. Mot. Skills* **75**, 1235–1241 (1992)
4. Coon, C.A., Schwanenflugel, P.J.: Evaluation of interruption behavior by naïve encoders. *Discourse Processes* **22**, 1–24 (1996)
5. Crown, C.L., Cummins, D.A.: Objective versus perceived vocal interruptions in the dialogues of unacquainted pairs, friends, and couples. *J. Lang. Soc. Psychol.* **17**, 372–389 (1998)
6. De Ruiter, J.P., Mitterer, H., Enfield, N.J.: Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* **82**, 515–535 (2006)
7. Eshghi, A., Howes, C., Gregoromichelaki, E., Hough, J., Purver, M.: Feedback in conversation as incremental semantic update. In: *Proceedings of the 11th International Conference on Computational Semantics (IWCS 2015)*. Association for Computational Linguistics, pp. 261–271 (2015)
8. Farley, S.D.: Attaining status at the expense of likeability: pilfering power through conversational interruption. *J. Nonverbal Behav.* **32**, 241–260 (2008)
9. Gnisci, A., Bull, P., Graziano, E., Ciancia, M.R., Errico, D.: Un sistema di codifica delle interruzioni per l'intervista politica italiana. *Psicologia Sociale* **1**, 107–128 (2011)
10. Gnisci, A., Graziano, E., Sergi, I., Pace, A.: Which criteria do naïve people use for identifying and evaluating different kinds of interruption? *J. Pragmat.* **138**, 119–130 (2018)
11. Gnisci, A., Sergi, I., De Luca, E., Errico, V.: Does frequency of interruptions amplify the effect of various types of interruptions? Experimental evidence. *J. Nonverbal Behav.* **36**, 39–57 (2012)
12. Goffman, E.: On face-work: an analysis of ritual elements in social interaction. *Psychiatry* **18**, 213–231 (1955)
13. Goldberg, J.A.: Interrupting the discourse on interruptions: an analysis in terms of relationally neutral, power- and rapport-oriented acts. *J. Pragmat.* **14**, 883–903 (1990)
14. Graziano, E., Gnisci, A.: The partiality in Italian political interviews: stereotype or reality? In: Esposito, A., Esposito, A.M., Martone, R., Müller, V.C., Scarpetta, G. (eds.) *Analysis of Verbal and Nonverbal Communication and Enactment: The Processing Issues*, pp. 363–375. Springer, Berlin (2011)
15. Graziano, E., Gnisci, A., Pace, A.: Gli effetti delle interruzioni nella conversazione diadica: Una rassegna degli studi sperimentali. *Giornale Italiano di Psicologia* **41**, 747–772 (2014)
16. Hawkins, K.: Interruptions in task-oriented conversations: effects of violations of expectations by males and females. *Women's Stud. Commun.* **11**, 1–20 (1988)
17. Hawkins, K.: Some consequences of deep interruption in task-oriented communication. *J. Lang. Soc. Psychol.* **10**, 185–203 (1991)
18. Holtgraves, T.: Politeness. In: Robinson, W.P., Giles, H. (eds.) *The New Handbook of Language and Social Psychology*, pp. 341–355. Wiley, New York (2001)
19. Kempson, R., Gargett, A., Gregoromichelaki, E.: Clarification requests: An incremental account. In: *Proceedings of the 11th Workshop on the Semantics and Pragmatics of Dialogue (DECALOG)*, pp. 62–75 (2007)
20. LaFrance, M.: Gender and interruptions: individual infraction or violation of the social order? *Psychol. Women Q.* **16**, 497–512 (1992)
21. Makri-Tsilipakou, M.: Interruption revisited: affiliative vs. disaffiliative intervention. *J. Pragmat.* **21**, 401–426 (1994)
22. Murata, K.: Intrusive or cooperative? A cross-cultural study of interruption. *J. Pragmat.* **21**, 385–400 (1994)
23. Orcutt, J.D., Mennella, D.L.: Gender and perception of interruption as intrusive talk: an experimental analysis and reply to criticism. *Symbolic Interact.* **18**, 59–72 (1995)

24. Robinson, L.F., Reis, H.T.: The effects of interruption, gender, and status on interpersonal perceptions. *J. Nonverbal Behav.* **13**, 141–153 (1989)
25. Roger, D., Bull, P., Smith, S.: The development of a comprehensive system for classifying interruptions. *J. Lang. Soc. Psychol.* **7**, 27–34 (1988)
26. Sacks, H., Schegloff, E.A., Jefferson, G.: A simplest systematics for the organization of turn-taking for conversation. *Language* **50**, 696–735 (1974)
27. West, C., Zimmerman, D.H.: Small insults: a study of interruptions in cross-sex conversations between unacquainted persons. In: Thorne, B., Kramarae, C., Henley, N. (eds.) *Language, Gender, and Society*, pp. 102–117. Newbury, Rowley (1983)

# Chapter 34

## Analyzing Likert Scale Inter-annotator Disagreement



Carl Vogel, Maria Koutsombogera and Rachel Costello

**Abstract** Assessment of annotation reliability is typically undertaken as a quality assurance measure in order to provide a sound fulcrum for establishing the answers to research questions that require the annotated data. We argue that the assessment of inter-rater reliability can provide a source of information more directly related to the background research. The discussion is anchored in the analysis of conversational dominance in the MULTISIMO corpus. Other research has explored factors in dialogue (e.g. big-five personality traits and conversational style of participants) as predictors of independently perceived dominance. Rather than assessing the contributions of experimental factors to perceived dominance as a unitary aggregated response variable following verification of an acceptable level of inter-rater reliability, we use the variability in inter-annotator agreement as a response variable. We argue the general applicability of this in exploring research hypotheses that focus on qualities assessed with multiple annotations.

### 34.1 Introduction

Assessment of inter-annotator agreement has long been a topic of research in the cognitive sciences. Normally, this assessment is undertaken in order to justify the use of an aggregation of raters' judgements as a unitary quantity which may then be used as a response variable in the exploration of hypotheses about factors that independently or in interaction influence the response variable. The subsequent exploration is deemed warranted only if a sufficient level of agreement is achieved in

---

C. Vogel · M. Koutsombogera (✉) · R. Costello  
School of Computer Science and Statistics, Trinity College Dublin, Dublin 2, Ireland  
e-mail: [koutsomm@scss.tcd.ie](mailto:koutsomm@scss.tcd.ie)

C. Vogel  
e-mail: [vogel@cs.tcd.ie](mailto:vogel@cs.tcd.ie)

R. Costello  
e-mail: [costelra@tcd.ie](mailto:costelra@tcd.ie)

the initial assessment. Debate accompanies the question of what counts as a sufficient level of agreement.

Carletta [3] has been influential in computational linguistics in advocating for the measurement of agreement in applying analytic labels to data with the use of the Kappa statistic as deployed within the content analysis literature [4]. Krippendorff [9] provides an analysis that relates a number of agreement coefficients, and Artstein and Poesio [1] provide a comparable analysis, but in developing an argument that Krippendorff's Alpha [8] is more suitable for many tasks in computational linguistics, partly because of allowing different weightings for disagreements. Veronis [12] analyses disagreement in the annotation of word senses and provides a method of clustering annotations in order to compensate for the inevitability of annotation disagreement that accompanies finely grained distinctions among annotations. Geertzen and Bunt [5] also work with a highly articulated labelling system (for annotation of dialogue acts) and develop a method of rating agreement that is sensitive to the multi-dimensional hierarchical relationships among the categories of labels available, offsetting the vast potential for disagreement that could arise with large ( $n = 86$ ) inventory of labels without hierarchical relationships. Beigman Klebanov and Beigman [2] suggest that "hard instances", those for which disagreement among annotators is rife, should be eliminated from "gold-standard" data sets. Reidsma and Carletta [11] reach the same conclusion that we do, that it is important to investigate patterns that emerge in annotation disagreement—default use of particular label types is likely among human annotators and likely to give rise to systematic disagreements. They too emphasize the importance of this in the context of using annotations to inform follow-on processes, such as training machine learning to label additional data sets.

The argument that we make is that while the aggregated annotations themselves may well be extremely useful for follow-on processes, the labels for data instances that arise from classifying data in terms of annotation difficulty are useful as well. Here, we focus on Likert scale annotations, thus, ordered categorical labels. We propose a means of aggregating response that does not assume that the Likert units are evenly separated on a linear scale. The aggregation method involves categorizing the degree of disagreement. Here, we perform such an aggregation that produces an ordered categorical labelling, but a continuous approach is, naturally, also possible.

The analysis here is conducted in the context of the MULTISIMO corpus of group dialogues [7], particularly, with reference to a recent analysis of dominance among interlocutors in those dialogues [6]. The MULTISIMO corpus records 23 group triadic dialogues among individuals engaged in a playful task akin to the premise of the television game show, *Family Feud*.<sup>1</sup> Two of the participants in each dialogue were randomly partnered with each other, and the third participant serves as the facilitator for the discussion. Individuals who participate as facilitators are involved in a number of such discussions (there were three facilitators, in total), while the other participants participate in only one dialogue. Each dialogue has the

---

<sup>1</sup>With complete compliance with the terms of consent provided by the participants, 18 of these dialogues are represented in the version of the corpus publicly available.

same overall structure: an introduction, three successive instances of the game, and a closing. Each instance of the game consists of the facilitator presenting a question like “What are three instruments in a symphony orchestra?” followed by a phase in which the participants have to propose and agree to three answers, and that followed in turn by a phase in which the participants must agree to a ranking of the three answers. The rankings are based on perceptions of what 100 randomly chosen people would propose as answers,<sup>2</sup> and therefore the “correct” ranking is in the order of popularity determined by this independent ranking.

Participants in the dialogues engaged in a big-five personality trait assessment, although partner matching did not take these scores into account. One of the purposes of the corpus construction was to assess perceptions of conversational dominance in relation to both features of conversational style and the personality profiles of the participants. In support of this, the multi-modal corpus was annotated by five annotators (see Sect. 34.2), with an aggregate annotation score for each participant derived from the median of annotators’ labels, given a five-point Likert scale. Koutsombogera et al. [6] discuss the application of usual inter-coder consistency scores to the data set. Here, we develop a coding of disagreement and use this factor as a response variable in order to determine whether features that have been analysed as contributing to dominance scores or explained by dominance scores also interact in explaining disagreement regarding dominance scores.

The next sections describe the data set used for the experiment, the annotation task and its results, and method of analysing disagreement in annotation, and then the application of the resulting disagreement classification to analysing factors that contribute to disagreement regarding dominance and ultimately to an aggregated dominance score.

## 34.2 Data Set

The study presented here exploits the MULTISIMO corpus [7], a multi-modal corpus consisting of 23 sessions of collaborative group interactions, where two players work together to provide answers to a quiz and are guided by a facilitator, who monitors their progress and provides feedback or hints when needed. The sessions were carried out in English and the task of the players was to converse with each other with the aim of estimating and agreeing on the three most popular answers to each of three questions, and rank their answers from the most to the least popular.<sup>3</sup> The corpus consists of synchronized audio and video recordings, and its overall duration is approximately 4 h, with an average session duration of 10 min. The average age of the participants is 30 years old, and gender is balanced (25 female, 24 male participants).

---

<sup>2</sup>This survey was conducted independently, and rankings are reported in a database related to the game, <http://familyfeudfriends.arjdesigns.com/>, last accessed 11.05.2018.

<sup>3</sup>Correctness of the answers and their rankings is determined by responses to an independent survey of sample of 100 people.

Eighteen nationalities are represented among participants, one-third of them being native English speakers.

The facilitator role was designed to enable the extraction of behavioural cues for the development of an agent responsible for managing the interaction and choosing actions that maximize the collaboration effort and the performance of the group participants. Because the facilitator had a greater level of prescription guiding their interactions in the dialogue, we do not address perceptions of facilitator dominance in the conversation, only perceived dominance of the players. However, because of the part of the responsibilities of the facilitator includes balancing participation of the players, a series of interventions by a facilitator could in principle form a signal regarding the relative dominance of one of the players.

A perception experiment was conducted to collect ratings from a group of annotators who observed the dialogue videos and assessed the level of dominance of the involved players. Five annotators were recruited and were asked to watch the corpus videos twice, observe the behaviour of participants and, for each video file, rate the two quiz players (not the facilitators) on a scale from 5 (highest) to 1 (lowest), depending on how they perceive the level of their dominance. Raters were given concise guidelines that included a definition of conversational dominance as “a person’s tendency to control the behaviour of others when interacting with them”. Although the raters were not provided with concrete examples of observable behavioural cues that are related to dominance estimation, they were informed that, in general, more dominant participants tend to be more active in terms of verbal and non-verbal manifestations, and that conversational dominance can be identified by observing people’s behaviours and expressive acts without prior knowledge of the social and personal relationships between them.

The annotators completed the ratings after watching the full videos, and they did not spend more than one hour per day on the task. The experimental implementation was supervised by the Trinity College Dublin, School of Computer Science and Statistics Research Ethics Committee, and all annotators signed the related consent forms. The intraclass correlation coefficient, calculated with the presumption of random row (subject) and column (annotator) effects ( $ICC(A,5) = 0.776$ ) is significant ( $F(35, 24.2) = 6.68, p = 3.57 \times 10^{-6}$ ), suggesting a reasonable level of agreement among annotators, a level sufficient to warrant use of the aggregation of annotations (using the median annotation score for each individual) as a response variable [6], even though annotation disagreements exist. Table 34.1 shows the correlation coefficient (and  $p$ -value) between each annotator’s classification of each participant and the median score for each participant. The main results of the study in [6] report the association of high dominance scores to the large number of words a speaker utters per minute, as well as to high extroversion and high openness personality traits, and the association of low dominance scores with high agreeableness.

**Table 34.1** Pearson product-moment correlation between the aggregated annotation scores (the median of the five annotators' scores for each item) and the scores of each of the annotators individually

	Annotator				
	1	2	3	4	5
Correlation coefficient	0.749	0.855	0.847	0.598	0.655
<i>p</i> -value	$1.47 \times 10^{-7}$	$3.02 \times 10^{-11}$	$7.75 \times 10^{-11}$	$1.17 \times 10^{-4}$	$1.46 \times 10^{-5}$

Significance levels as *p*-values are reported in the second row

### 34.3 Method

We propose to measure response disagreement in units of the same magnitude as that of the Likert scale. One issue with Likert response data is that an arithmetic mean response is not manifestly meaningful, given that annotators may not perceive the separation between points as equal in all cases. Using the median of responses to items as a measure of the central tendency for an item makes fewer assumptions. Therefore, for each item, we subtract each annotator's score from the median response, and sum the absolute value of those differences, and then divide by the number of annotators (34.1). We then sum the item scores and divide by the number of items (34.2). In this definition, we are relying on an arithmetic mean, but in the difference between the median of annotators' response and each annotator's response—a natural alternative to consider the median of these differences, but this will yield a value no larger than the mean deviation [10] (see Table 34.2, whose terms are described in Table 34.3).

**Table 34.2** Possible Likert category scales and maximal disagreement ( $\delta_i$ ) per item

$C = k$	$\tilde{k}_i$	$\sum_{c=1}^C  (\tilde{k}_i - k_i^c) $	$\delta_i$
1	1	0	0
2	1.5	1	0.5
3	2	2	0.67
4	2.5	4	1
5	3	6	1.2
6	3.5	9	1.5
7	4	12	1.71
...	...	...	...
$C$	$\frac{C+1}{2}$	$\left\lfloor \left( \frac{C}{2} \right)^2 \right\rfloor$	$\left\lfloor \left( \frac{C}{2} \right)^2 \right\rfloor / C$

**Table 34.3** Term definitions

Quantity name	Definition
$n$	Number of items
$k$	Number of Likert categories
$k_i^c$	Category supplied by annotator $c$ to item $i$
$\tilde{k}_i$	Median response to item $i$
$C$	Number of annotators
$\delta_i$	Item score for item $i$ [see (34.1)]

$$\delta_i = \frac{\sum_{c=1}^C |(\tilde{k}_i - k_i^c)|}{C} \quad (34.1)$$

$$\delta = \frac{\sum_{i=1}^n \delta_i}{n} \quad (34.2)$$

The result ( $\delta$ ) is a measure of average Likert unit disagreements per item. Perfect agreement on each item over a set yields a score of zero. Supposing that there are as many annotators as there are Likert categories (and supposing that the Likert categories are identified by the integers [1:k]), then the maximal disagreement situation is for each annotator to choose a distinct label, and in this situation, the maximal disagreement item score may be calculated as in (34.3). A sketch of the progression of possible values for ( $\delta_i$ ) for an item is provided in Table 34.2. It follows that the maximum disagreement across a set of items obtains if the scores for each item involve permutations of  $C$  annotators each selecting a Likert rating for each item that none of the other annotators have chosen, and in this case, the average across items ( $\delta$ ) will be identical to each  $\delta_i$ .<sup>4</sup>

$$\frac{\left\lfloor \left( \frac{k}{2} \right)^2 \right\rfloor}{C} \quad (34.3)$$

In any experimental data developed from an instrument that elicits responses to a Likert scale will fix values for  $C$  and  $k$ , participants are likely to agree on some responses. The values that  $\delta_i$  may take across all of the items form discrete, but ordered, labels for disagreement on each of the items, and thus, this categorization of items can be used within analysis as an ordinal (or as a nominal) value. That is, follow-on analysis may make use of aggregated annotations (such as the median of annotator responses to the Likert scale) and such analysis may also make use of the additional labelling that arises from the disagreement scores. Here, we focus on the latter ([6] focus on the former).

---

<sup>4</sup>The first two rows of this table are provided for sake of completeness—it does not appear rational to propose a Likert scale with only one point, and if the experimental question required only two points, it seems unlikely that one would approach the binary judgement using a Likert scale.

In Sect. 34.4, we apply this method to the data described in Sect. 34.2: first, we illustrate the disagreement scores that emerge for the data set discussed by [6], and then we explore the role of variables analysed in the work [6] in relation to their being able to predict the disagreement scores (as evaluated using ordinal logistic regression).

## 34.4 Results

Suppose four of our annotators had identical judgements (those of Annotator1) and that the fifth had the judgements of Annotator2, as shown in Fig. 34.1. In that situation, there are  $\delta = 0.22$  Likert points disagreement, on average, per item, across all the judgements in the data set. As noted in Table 34.2, the “worst case” for this data set is 1.2.

When the method is applied to the actual annotations, we obtain  $\delta = 0.56$  (across items, min. = 0, 1st. Q = 0.4, median = 0.6, 3rd. Q = 0.8, max = 1). A summary of the number of times each  $\delta_i$  value was applied is provided in Table 34.4. Thus, it can be seen that within the data, only one item elicited no disagreement. Using this classification of the data, it is possible to identify the single item that obtained a disagreement score of 1. The annotators assigned dominance ratings to this item of 5, 2, 3, 3 and 5, respectively. This sequence has a median of 3, so  $(2 + 1 + 0 + 0 + 2)/5$  determines the score of 1. There is interesting agreement within this disagreement (two annotators choosing a maximal dominance label and two choosing a neutral

```
MedDomComplement11112 <-
  apply((cbind(Annotator1,
  Annotator1,
  Annotator1,
  Annotator1,
  Annotator2)), 1,
  FUN <- function(x) {m <- median(x);
  d <- function(y) {abs(m-y)} ;
  return(sum(d(x))/length(x))})

MedDomComplement11112f <- factor(MedDomComplement11112)

> mean(MedDomComplement11112)
[1] 0.2166667
```

**Fig. 34.1** Hypothetical agreement among annotators (as processed using R)

**Table 34.4** Counts of items within each level of possible disagreement score ( $\delta_i$ )

$\delta_i$	0	0.2	0.4	0.6	0.8	1
Item count	1	2	11	12	9	1

label). Thus, the disagreement categories highlight both items and point to possible differences that may be explored according to the demographic variables available in connection with the annotators.

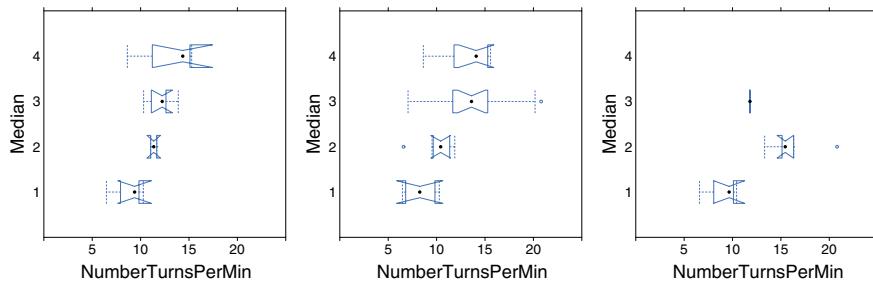
Our annotators are closer to perfect agreement than they are to perfect disagreement—at 0.56, they are just to the left of the midway point (using the usual left–right ordering of the number line). One can see from Table 34.4 that there are more items that are to the left of the midpoint between perfect agreement and perfect disagreement than there are to the right of that midpoint (and more to the left that are at the midpoint).

It becomes possible to explore the data further using the index of disagreement in order to learn what factors in the interactions (or their meta-data) predict these categories of disagreement scores. This, in turn, might provide some leverages in lifting the lid on answers to the question about what gives rise to the annotation disagreements. Because  $\delta_i$  is an ordinal variable, it is possible to use ordinal logistic regression on the quantities available. Here, we explore variables that are studied by [6]: Number of Turns Per Minute, Number of Words Per Minute and Average Turn Duration.

An ordinal logistic model that predicts disagreement on the basis of those three factors as independent influences may be rejected as providing an adequate account of the variation within the disagreement ( $p = 1.435946 \times 10^{-09}$ ); nonetheless, within this model, the number of turns per minute and number of words per minute are significant ( $p < 0.05$ ). With a unit increase in turns per minute, the odds of a high level of disagreement (vs. the combined adjacent disagreement categories) are 11.9 times greater, assuming the other variables are held constant. Similarly, a unit increase in number of words per minute gives odds of a high level of disagreement 0.9 times greater than adjacent combined categories), with the other variables held constant. That is, that increase in the number of turns per minute is significant in leading to increase in disagreement about dominance among annotators, while decrease in number of words per minute leads to decrease in disagreement. Of course, these observations pertain to a model that can be rejected as adequate to explain the categories of disagreement, even if that model is itself better than one that also paid attention to the personality traits of EXTROVERSION, AGREEABLENESS and OPENNESS as independent classifications of the individuals rated by the annotators.

## 34.5 Discussion

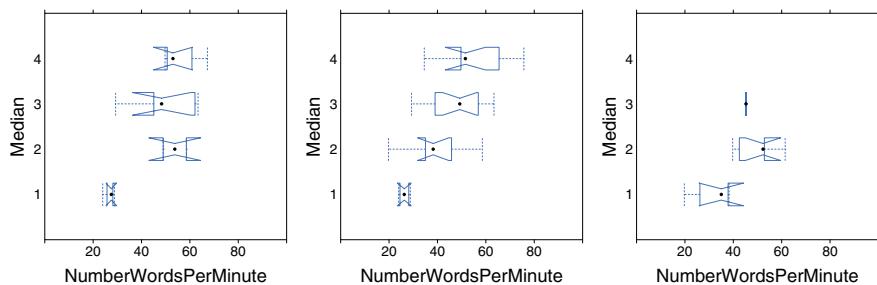
A point here is that reasoning about the data in relation to disagreement may shed further light on the relationships underlying the data. In relation to this data set, a background hypothesis that duration of the conversation spent holding the floor correlates positively with dominance. This in mind, two quantities, both relativized to the total number of minutes of conversation, are relevant: number of words per minute and number of turns per minute. It was noted above that disagreement regarding dominance labelling increases with increase in the number of turns per minute. In



**Fig. 34.2** Median dominance given number of turns per minute: on the left, the restriction to where  $\delta < 0.6$ ; on the right, the restriction to where  $\delta > 0.6$ ; in the middle, over the whole data set (including the points where  $\delta = 0.6$ , which are not represented on the left or right)

Fig. 34.2, the plots depict how the median dominance score (the aggregate of the five raters' annotations) relates to turns per minute: the plot on the left indicates the relationship where there is least disagreement ( $\delta < 0.6$ ); the plot in the middle shows the relationship over the whole data set; and the plot on the right shows the relationship where there is greatest disagreement ( $\delta > 0.6$ ). From the middle plot, it can be discerned that over the whole data set, there is not a significant difference between neutral scores (i.e., a median dominance score of 3) and indications of greater dominance, as a function of the number of turns per minute, but that there is a difference between neutral scores and lesser levels of dominance (overall the Pearson correlation coefficient is 0.47,  $p < 0.005$ ). The plot on the left in Fig. 34.2 shows that where there is most agreement ( $\delta < 0.6$ ), a linear relationship is evident, with an increasing number of turns per minute correlating with an increasing median dominance score ( $0.65$ ,  $p < 0.05$ ). The plot on the right depicts the cases where there is least agreement ( $\delta > 0.6$ ), and here there is no clear trend (the correlation is not significant): the greatest number of turns per minute yields a less than neutral median dominance score.

In comparison, the plots of Fig. 34.3 show that the variable with a lesser but still significant impact on disagreement than number of turns per minute, number of words per minute, also has the effect of disagreement impinge mostly on the categories below the neutral point of the Likert scale. Where there is most agreement about dominance ( $\delta < 0.6$ , shown in the plot on the left), increase in the number of words spoken relative to the total duration of the conversation in minutes does not reliably lead to a higher agreed dominance rating [although the correlation coefficient of 0.67 is significant ( $p < 0.05$ )]. A trend of increase in this measure of ownership of the floor and dominance does appear over the whole data set (middle plot, correlation coefficient = 0.62,  $p = 5.226 \times 10^{-5}$ ). However, where disagreement is greatest ( $\delta > 0.6$ ), depicted in the plot on the right, the trend is partial (and the correlation is not significant).



**Fig. 34.3** Median dominance given number of words per minute: on the left, the restriction to where  $\delta < 0.6$ ; on the right, the restriction to where  $\delta > 0.6$ ; in the middle, over the whole data set (including the points where  $\delta = 0.6$ , which are not represented on the left or right)

The analysis of disagreement in dominance annotation emphasizes that the quantities addressed here as measures of ownership of the floor do not have univocal explanatory value in the determination of dominance annotations.

## 34.6 Conclusion

Our analysis of disagreement in Likert scale annotations suggests using this as a response variable in itself in order to determine the factors that contribute to disagreement about annotations. While the primary research questions may be directed towards an aggregation of Likert annotations as a response variable, the disagreement variable provides a principle means of separating data points where the hypothesized model may be behaving differently than expected. This seems advisable in avoidance of a version of Simpson's paradox in which trends visible on the underlying aggregated response variable are reversed when exploring areas of agreement and disagreement in the scores that contribute to the response variable. In particular, while it is customary to report annotator agreement metrics in relation to data that is to be put to further use, such as serving in the role of a response variable. Where models of such an aggregated response variable are built and tested from theoretical assumptions about experimental variables that determine the response, it is advisable to additionally test the extent to which those same experimental variables have significant impacts on a variable that encodes disagreement among the annotations that aggregate into the underlying response variable.

**Acknowledgements** The research leading to these results has received funding from (a) the ADAPT Centre for Digital Content Technology, funded under the SFI Research Centres Programme (Grant 13/RC/2106) and co-funded under the European Regional Development Fund, and (b) the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 701621 (MULTISIMO).

## References

1. Artstein, R., Poesio, M.: Inter-coder agreement for computational linguistics. *Comput. Linguist.* **34**(4), 555–596 (2008)
2. Beigman Klebanov, B., Beigman, E.: From annotator agreement to noise models. *Comput. Linguist.* **35**(4), 495–503 (2009)
3. Carletta, J.: Assessing agreement on classification tasks: the kappa statistic. *Comput. Linguist.* **22**(2), 249–254 (1996)
4. Cohen, J.: A coefficient of agreement for nominal scales. *Educ. Psychol. Measur.* **20**(1), 37–46 (1960)
5. Geertzen, J., Bunt, H.: Measuring annotator agreement in a complex hierarchical dialogue act annotation scheme. In: Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue, pp. 126–133. Association for Computational Linguistics (2006)
6. Koutsombogera, M., Costello, R., Vogel, C.: Quantifying dominance in the multisimo corpus. In: Baranyi P., Esposito A., Földesi P., Mihálydeák T. (eds.) 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2018), pp. 147–152. IEEE (2018)
7. Koutsombogera, M., Vogel, C.: Modeling collaborative multimodal behavior in group dialogues: The MULTISIMO corpus. In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA), Paris, France (2018)
8. Krippendorff, K.: Content Analysis: An Introduction to its Methodology. Sage, Thousand Oaks, CA (2004)
9. Krippendorff, K.: Reliability in content analysis: some common misconceptions and recommendations. *Hum. Commun. Res.* **30**(3), 411–433 (2004)
10. Pham-Gia, T., Hung, T.L.: The mean and median absolute deviations. *Math. Comput. Model.* **34**, 921–936 (2001)
11. Reidsma, D., Carletta, J.: Reliability measurement without limits. *Comput. Linguist.* **34**(3), 319–326 (2008)
12. Veronis, J.: A study of polysemy judgements and inter-annotator agreement. In: Programme and Advanced Papers of the Senseval Workshop, Herstmonceux (1998). <http://www.itri.brighton.ac.uk/events/senseval/ARCHIVE/PROCEEDINGS/interannotator.ps>. URL last verified Feb 2019

# Chapter 35

## PySiology: A Python Package for Physiological Feature Extraction



Giulio Gabrieli Atiqah Azhari and Gianluca Esposito

**Abstract** Physiological signals have been widely used to measure continuous data from the autonomic nervous system in the fields of computer science, psychology, and human–computer interaction. Signal processing and feature estimation of physiological measurements can be performed with several commercial tools. Unfortunately, those tools possess a steep learning curve and do not usually allow for complete customization of estimation parameters. For these reasons, we designed *PySiology*, an open-source package for the estimation of features from physiological signals, suitable for both novice and expert users. This package provides clear documentation of utilized methodology, guided functionalities for semi-automatic feature estimation, and options for extensive customization. In this article, a brief introduction to the features of the package, and to its design workflow, are presented. To demonstrate the usage of the package in a real-world context, an advanced example of image valence estimation from physiological measurements (ECG, EMG, and EDA) is described. Preliminary tests have shown high reliability of feature estimated using *PySiology*.

### 35.1 Introduction

Physiology, the science of the functions of living organisms and their parts, is a broad scientific discipline that encompasses elements of psychology, computer sciences, and human–computer interaction. As a subject, physiology spans across molecular and cellular components through to the levels of tissues, organs, and whole systems. Physiology provides the bridge between scientific discoveries and their application to medical science.

Analysis of physiological signals can bring advantages to several different fields. First, physiological signals are directly controlled by the autonomic nervous

---

G. Gabrieli · A. Azhari · G. Esposito

Psychology Program, School of Social Sciences, Nanyang Technological University,  
Singapore, Singapore

e-mail: [GIULIO001@e.ntu.edu.sg](mailto:GIULIO001@e.ntu.edu.sg)

G. Esposito

Department of Psychology and Cognitive Science, University of Trento, Trento, Italy

system (ANS), and therefore, analysis based on physiological features allows for the exclusion of potentially biased social inferences from collected data. Second, using wearable sensors and devices, it is possible to collect continuous measurements that can be used in pervasive computing and continuous medical investigation [1]. On the downside, physiological measurements are highly influenced by the presence of artifacts and noise. In order to be effectively used, hardware setup for data collection, signal processing, and feature extraction have to be carefully executed in order to maximize the signal-to-noise ratio.

From the variety of available physiological signals, electrocardiography (ECG), electromyography (EMG), and electrodermal activity (EDA) play an exceptionally important role in clinical application and have been widely adopted in emotion recognition, affective computing, and neuroscientific research [2–9].

Various software have been developed in recent years for automatic or semi-automatic analysis of physiological signals. These tools provide support for signal preprocessing and feature extraction, as well as data analysis and data reporting. Unfortunately, these programs are usually proprietary (closed-source), limiting access to the methodology used in the analysis, which subsequently makes comparison across different methods impossible. Furthermore, some of them do not allow for configuration of estimation parameters, precluding the possibility of testing different setups and tunings during feature estimation.

For these reasons, we have decided to develop a new tool for physiological feature estimation, in the form of a free and open-source Python package, released under the name *PySiology*.

The aim of *PySiology* is to provide researchers with a tool that can be easily configured for feature estimation of physiological signals, suitable for both novice and expert users. The open nature of the project allows users to assess and modify the source code, share new features, and make comparisons across new and old techniques. Also, because of the high degree of customization and ease of use, *PySiology* can be easily integrated into complex analysis structures from other software. *PySiology* (version 0.0.9) provides modules for the estimation of features from three different types of signals: ECG, EMG, and EDA.

In this paper, a brief introduction to the development workflow and status of our library is provided.

## 35.2 Development Workflow

The workflow of *PySiology* has been specifically developed to enhance collaboration between users. The source code of the project is provided through GitHub, a public repository that allows for external contributions from users with regard to bug tracking, the proposal of novel physiological features, and general discussion. This platform also allows users to enter sample data into the package, upload tutorials, improve the precision of estimated features, and enhance the quality of modules documentation.

Documentation is automatically generated using Sphinx, and provided in both an online and printable format, on the GitHub Repository<sup>1</sup> and on Read the Docs.<sup>2</sup>

Tutorials and examples are preferably uploaded in the form of *Jupyter notebooks*, while no specific format is suggested for the sample data. This allows for the presentation of different experimental situations (e.g., data collected from different devices) while maintaining high readability and reproducibility of example codes. Feature estimation is conducted using functions, classes, and objects. Prior to being added to the project, parameters should be customizable and have predefined values, methods of feature estimation should be referenced, and all functions must be documented. Data are preferred in the form of Numpy array.

## 35.3 Modules, Features, and Installation

### 35.3.1 Package Structure

*PySiology*'s code is organized into subpackages, each one referring to a specific signal (e.g., ECG) or functionality (e.g., loading of sample data). Version 0.0.9 is structured as follows:

- `physiology.electrocardiography`: ECG signal processing and feature estimation
- `physiology.electrodermalactivity`: EDA signal processing and feature extraction
- `physiology.electromyography`: EMG signal processing and feature estimation
- `physiology.sampledata`: sample data for testing and educational purposes.

### 35.3.2 Feature Estimation

*PySiology* allows for the estimation of different physiological features, both in time and frequency domains. For each feature, expert users can customize all the used parameters (e.g., thresholds and cutoff frequencies), while standard values are provided for quick analysis and to assist novice users. A brief list of estimable features for each signal is reported in Table 35.1. More detailed information on estimable features, references, and methodology used for estimation is available in the document.

Functions and methods for signal preprocessing are provided. Basic preprocessing pipelines, including bandpass filters, phasic filters, and downsampling, are suggested.

---

<sup>1</sup><https://gabrock94.github.io/PySiology/html/index.html>.

<sup>2</sup><https://physiology.readthedocs.io>.

**Table 35.1** List of features estimable using *PySiology* (v. 0.0.9), divided by signal

Signal	Features
ECG	IBI, BPM, sdnn, sdsd, rmssd, ppn50, ppn20, high frequency, low frequency, very low frequency
EMG	IEMG, MAV, MAV1, MAV2, SSI, VAR, TM, LOG, RMS, WL, AAC, DASDV, AFB, ZC, MYOP, WAMP, SSC, MAVSLPk, HIST, MNF, MDF, peak frequency, TTP, SM, FR, PSR, VCF
EDA	Rise time, amplitude, EDA at apex, decay time, SCR width

### 35.3.3 Installation

*PySiology* is available via the Python package index PyPi or from the GitHub repository. Detailed instruction for installation, including required packages and their versions, is available in the document. Installation using a package manager, such as *pip* or *conda*, is recommended.

## 35.4 Advanced Example: Predicting Valence of an Image from Physiological Features

There is a longstanding tradition for emotion recognition to be investigated from physiological signals. For example, in a study on musically induced aesthetic pleasure by Ray et al. [10], morphological variation in ECG and EEG signatures was found, suggesting a correlation between a mild increase in sympathetic activity of the autonomic nervous system with the valence of the stimuli. Similar results were obtained by de Jong et al. [11, 12] in which ECG and EDA signals were collected during experiments on music and painting perception.

In this example, physiological recordings of viewers in order to estimate the valence of emotional images selected from the International Affective Picture System (IAPS) dataset are used. IAPS is a dataset consisting of 1180 emotionally evocative pictures, developed by NIMH Center for the Study of Emotion and Attention, University of Florida [13]. Affective norms (ratings of valence, arousal, and dominance) for IAPS pictures were obtained from 18 separate studies and are provided together with the dataset.

Because of the high arousal level of selected images, we expect a high correlation between physiological activity and emotional experience [14]. To estimate the valence of an image, *PySiology* has been used, paired with classifiers provided in *scikit-learn*, a python package for machine learning and neural networks implementation.

### 35.4.1 Data Collection

From the IAPS dataset, 50 stimuli were selected for presentation and divided into two groups: low valence ( $N = 25$ , mean valence =  $1.79 \pm 0.14$ ) and high valence ( $N = 25$ , mean valence =  $8.06 \pm 0.13$ ). Stimuli were presented to university students ( $N = 58$ , 26 males, 33 females, mean age  $21.5 \pm 2.3$ ) for 8 s each, with an interval of 6 s between two stimuli. ECG, EDA, and EMG (*corrugator supercilii*) signals were collected on a Bitalino Revolution BT board, a low-cost device designed for physiological data collection (sampling rate: 1000 Hz, Wireless Biosignals S.A.) [15], at a resolution of  $1279 \times 800$ . Data used in this manuscript are available online [16].

### 35.4.2 Preprocessing

ECG signal was first preprocessed for noise removal, using a fifth-order bandpass filter with a high-pass cutoff at 0.5 Hz and a low-pass cutoff at 2.5 Hz. Baseline correction was conducted with reference to the 20 s of recording preceding the presentation of the first stimulus.

GSR signal was preprocessed for noise remotion. First, a second-order bandpass filter, with a high-pass cutoff at 0.05 Hz and a low-pass cutoff at 1 Hz, was applied. Then, the signal was downsampled using a scaling factor of 100–10 Hz. Using the median value, a median filter was then applied to remove the tonic component from the signal.

Raw EMG signal was first preprocessed for noise correction, using a second-order bandpass filter with a high-pass cutoff at 20 Hz and a low-pass cutoff at 50 Hz. A basic median filter was applied to separate the phasic from the tonic components, subtracting the median amplitude calculated four seconds before and after each time point.

### 35.4.3 Feature Extraction

For feature estimations, preprocessed raw signals were first segmented into epochs. Functionalities for feature estimation were then applied to each epoch.

**ECG:** ECG features were estimated using the submodule *electrocardiography*. From the ECG signals, both time-domain and frequency-domain features were estimated. For time-domain frequencies, peak detection was done by utilizing a minimum distance between peaks of 500 ms. Frequency-domain features were estimated using cutoff frequency values adapted from blood [17]. Used values are reported in Table 35.2.

**Table 35.2** Cutoff values for ECG frequency analysis

Frequency	Lower cutoff (Hz)	Upper cutoff (Hz)
Very low frequency	0.0033	0.04
Low frequency	0.04	0.15
High frequency	0.153	0.4

**GSR:** GSR features were estimated using the submodule *electrodermalactivity*. The first peak with an onset of at least 1 s from stimuli presentation was used for feature estimation.

**EMG:** EMG features were estimated using the submodule *electromyography*. For the evaluation of zero crossing (ZC), average myopulse output (MYOP), Willison amplitude (WAMP), and slope sign changes (SSC), the threshold value was set to 0.01 (*\*\*threshold*). While estimating the mean absolute value slope (MAVLSPk), the number of segments used (*\*\*nseg*) was 3.

### 35.4.4 Classification

Recursive partitioning (decision tree) and multi-layer perceptron classifier (MLP) have been used for classification. Estimated physiological features were first reduced using principal component analysis (PCA) to six components, while IAPS valence values have been used as targets. Before being fed to the neural network classifier, input values were standardized by removing the mean and values were scaled to unit variance.

For both the classifiers, standard scikit-learn (version 0.19.1) parameter was used.

For each classifier, 100 tests have been done, from which 45 pictures were chosen for training and 5 used as test cases.

### 35.4.5 Results and Discussion

Average accuracy results, reported in Table 35.3, show that both classifiers were able to achieve an accuracy above chance with a small number of features and samples. Classifier parameters were not tuned according to the input data, suggesting that enhancement of accuracy level can be achieved. Also, having used data recorded from low-cost devices, feature extraction from clinical-grade instrumentation seems promising for high-quality estimation from raw signals.

**Table 35.3** Average accuracy results of decision tree and MLP

Classifier	Avg. accuracy (%)
Decision tree	66
MLP classifier	65

## 35.5 Future Development

*PySiology* is still under continuous development, despite alpha and beta versions having already been released and tested. Future development for the package will include improvement in computational duration and accuracy, implementation of novel methodologies for physiological feature estimation, modules for guided statistical analysis, and data modeling with report generation. Furthermore, we aim to add more open-source sample datasets and tutorials, with detailed instructions for every step of the conducted analysis.

## 35.6 Conclusion

*PySiology* is an actively developing open-source package for the estimation of physiological features. Its primary goal is to provide researchers with a tool that can easily be used for feature estimation and that allows for the customization of every parameter for ad hoc tuning of used algorithms. Furthermore, the package is designed to be employed by non-expert users.

Due to the open nature of this software, *PySiology* allows for comparison across different methodologies. Aside from its user-friendly features, such as tutorials, examples, suggested parameters, and analysis procedures, this package has also been developed to promote customization, high-quality documentation, and sharing of open data. In this paper, a brief introduction to the package, demonstrated within the context of an example case, has been provided, thus showcasing the accuracy of estimated features.

## References

1. Wagner, J., Kim, J., André, E.: From physiological signals to emotions: implementing and comparing selected methods for feature extraction and classification. In: IEEE International Conference on Multimedia and Expo. ICME 2005, pp. 940–943. IEEE (2005)
2. Pitt, M.: The use of stimulated EMG in the diagnosis of neuromuscular junction abnormality. In: Pediatric Electromyography, pp. 123–136. Springer, Berlin (2017)
3. Ahuja, N.D., Agarwal, A.K., Mahajan, N.M., Mehta, N.H., Kapadia, H.N.: GSR and HRV: its application in clinical diagnosis. In: 16th IEEE Symposium on Computer-Based Medical Systems, 2003. Proceedings, pp. 279–283. IEEE (2003)

4. Xun, L., Zheng, G.: ECG signal feature selection for emotion recognition. *Indonesian J. Electr. Eng. Comput. Sci.* **11**(3), 1363–1370 (2013)
5. Esposito, G., Yoshida, S., Ohnishi, R., Tsuneoka, Y., del Carmen Rostagno, M., Yokota, S., Okabe, S., Kamiya, K., Hoshino, M., Shimizu, M., et al.: Infant calming responses during maternal carrying in humans and mice. *Curr. Biol.* **23**(9), 739–745 (2013)
6. Castroflorio, T., Bargellini, A., Rossini, G., Cugliari, G., Deregibus, A., Manfredini, D.: Agreement between clinical and portable EMG/ECG diagnosis of sleep bruxism. *J. Oral Rehabil.* **42**(10), 759–764 (2015)
7. Truzzi, A., Setoh, P., Shinohara, K., Esposito, G.: Physiological responses to dyadic interactions are influenced by neurotypical adults' levels of autistic and empathy traits. *Physiol. Behav.* **165**, 7–14 (2016)
8. Truzzi, A., Bornstein, M.H., Senese, V.P., Shinohara, K., Setoh, P., Esposito, G.: Serotonin transporter gene polymorphisms and early parent-infant interactions are related to adult male heart rate response to female crying. *Front. Physiol.* **8**, 111 (2017)
9. Truzzi, A., Poquérusse, J., Setoh, P., Shinohara, K., Bornstein, M.H., Esposito, G.: Oxytocin receptor gene polymorphisms (rs53576) and early paternal care sensitize males to distressing female vocalizations. *Dev. Psychobiol.* **60**(3), 333–339 (2018)
10. Ray, G., Kaplan, A.Y., Jovanov, E.: Morphological variations in ECG during music-induced change in consciousness. In: Proceedings of the 19th Annual International Conference of the Engineering in Medicine and Biology Society, 1997, vol. 1, pp. 227–230. IEEE (1997)
11. de Jong, M.A.: A physiological approach to aesthetic preference—I. Paintings. *Psychother. Psychosom.* **20**(6), 360–365 (1972)
12. de Jong, M.A., Van Mourik, K., Schellekens, H.: A physiological approach to aesthetic preference. *Psychother. Psychosom.* **22**(1), 46–51 (1973)
13. Lang, P., Bradley, M.M.: The international affective picture system (IAPS) in the study of emotion and attention. In: *Handbook of Emotion Elicitation and Assessment*, 29 (2007)
14. Mauss, I.B., Levenson, R.W., McCarter, L., Wilhelm, F.H., Gross, J.J.: The tie that binds? coherence among emotion experience, behavior, and physiology. *Emotion* **5**(2), 175 (2005)
15. da Silva, H.P., Guerreiro, J., Lourenço, A., Fred, A.L., Martins, R.: Bitalino: a novel hardware framework for physiological computing. In: *PhyCS*, pp. 246–253. Citeseer (2014)
16. Esposito, G., Gabrieli, G.: Replication data for: physiology: a python package for physiological feature extraction. Dataset Published on DR-NTU (Data) (2019). <https://doi.org/10.21979/N9/GZSQT7>
17. Blood, J.D., Wu, J., Chaplin, T.M., Hommer, R., Vazquez, L., Rutherford, H.J., Mayes, L.C., Crowley, M.J.: The variable heart: high frequency and very low frequency correlates of depressive symptoms in children and adolescents. *J. Affect. Disord.* **186**, 119–126 (2015)

# Chapter 36

## Effect of Sensor Density on eLORETA Source Localization Accuracy



**Serena Dattola, Fabio La Foresta, Lilla Bonanno, Simona De Salvo, Nadia Mammone, Silvia Marino and Francesco Carlo Morabito**

**Abstract** The EEG source localization is an interesting area of research because it provides a better understanding of brain physiology and pathologies. The source localization accuracy depends on the head model, the technique for solving the inverse problem, and the number of electrodes used for detecting the EEG. The purpose of this study is to examine the localization accuracy of the eLORETA method applied to high-density EEGs (HDEEGs). Starting from the 256-channel EEGs, three different configurations were extracted. They consist of 18, 64, and 173 electrodes. The comparison of the results obtained from the different configurations shows that an increasing number of electrodes improve eLORETA source localization accuracy. It is also proved that a few number of electrodes could be not sufficient to detect all active sources. Finally, some sources resulting significant when few electrodes are used could turn out to be less significant when EEG is detected by a greater number of sensors.

### 36.1 Introduction

Electroencephalography (EEG) is a noninvasive diagnostic technique for recording brain electric activity by means of electrodes placed on the head surface. The potentials measured at the scalp originate from synchronous activity of populations of cortical pyramidal neurons. Because of the conductivity of the scalp and the underlying tissues, each electrode measures the potentials from all active sources, superposed as a function of their distance and orientation. Therefore, a realistic head model is essential for an accurate analysis of EEG signal [1].

---

S. Dattola (✉) · F. La Foresta · F. C. Morabito  
DICEAM - Università degli Studi Mediterranea di Reggio Calabria Feo di Vito, 89100 Reggio Calabria, Italy  
e-mail: [serena.dattola@unirc.it](mailto:serena.dattola@unirc.it)

L. Bonanno · S. De Salvo · N. Mammone · S. Marino  
IRCCS Centro Neurolesi Bonino-Pulejo, Via Palermo c/da Casazza, SS. 113, Messina, Italy

EEG source localization is a very important matter to understand brain physiology and highlight any alterations in the presence of pathological conditions. Source localization consists in solving the so-called EEG inverse problem. Many different source configurations can generate the same electric field on the scalp so there is more than one possible solution. In order to obtain a unique solution, mathematical, biophysical, and anatomical constraints are required. Over the years, many methods have been developed to solve the EEG inverse problem [2]. Among the electromagnetic tomographies, the minimum norm estimate was the first algorithm to provide an instantaneous, distributed, discrete, linear solution [3], but it showed a misplacement of deep sources onto the outermost cortex. The problem of large localization errors was solved in 1994 with the introduction of low-resolution electromagnetic tomography (LORETA) [4]. It is characterized by a good localization accuracy even for deep sources, and its average localization error is smaller than one grid unit.

A few years later, the development of LORETA led to a new version of it called sLORETA, based on the standardization of the current density [5]. Unlike the Dale method [6], sLORETA takes into account not only the noise due to the EEG measurements but also the biological noise due to the actual sources. sLORETA shows zero localization error under ideal (no-noise) conditions and no localization bias in the presence of measurement and biological noise.

The last version of LORETA, eLORETA, introduced a weight matrix which considers the deeper sources in a more adequate way [7]. eLORETA has zero error localization in the presence of measurement and structured biological noise. It also shows a better ability of suppressing less significant sources and provides less blurred images as compared with sLORETA [8].

Over the years, several studies proved the validity of LORETA, combining it with functional MRI [9, 10]. It was also successfully employed in researches about Alzheimer's disease [11–13] and epilepsy [14, 15].

The determination of the minimum number of electrodes to avoid undersampling the scalp potential has always been a very important issue. Several studies have been conducted to set the minimum interelectrode spacing. They obtained a range varying from 1 to 3 cm [16–18]. For years, EEG has been recorded by a few number of electrodes according to the international 10–20 system, with an interelectrode distance of 7 cm. Nowadays, scalp potentials can be detected by a greater number of electrodes, up to 256 (high-density EEG). Several studies have shown that an increasing number of recording sensors improve the limited spatial resolution of standard EEG [19, 20].

Song et al. [21] proved that the source localization accuracy of epileptiform activity is inadequate with less than 64 electrodes and improves with higher densities, becoming reasonable for 128 electrodes. The improvement from 128 to 256 electrodes is modest. They also demonstrate that the accuracy of source localization depends on sensor density as well as an adequate coverage of the whole head surface (inferior and superior head regions).

The purpose of this paper is to compare the source localization accuracy using a 256-channel EEG, the corresponding 10–10 system 64-channel EEG and the corresponding 10–20 system 18-channel EEG. Both 64-channel EEG and 18-channel

EEG are extracted from the 256-electrode configuration. The method used for the solution of the inverse problem is eLORETA [7]. The data analyzed are from two patients with gliotic lesions. Our expected result is that an increasing number of electrodes improve the source localization accuracy.

## 36.2 Materials and Methods

### 36.2.1 The EEG Inverse Problem

The EEG inverse problem is defined as the estimation of the source location from scalp potential measurements. This is an undetermined ill-posed problem because the number of unknown parameters (active sources) is greater than the number of known parameters (electrodes). In order to choose the most likely solution among all possible solutions, mathematical, neuroanatomical, and neurophysiological constraints are used.

There are two main groups of inverse algorithm based on two different approaches:

1. parametric methods, which assume that activated areas are generated by a few number of point sources (equivalent current dipole approach);
2. nonparametric methods, which assume that all brain points (a discrete search space) can be simultaneously active (linear distributed approach).

In particular, in this paper, a discrete, three-dimensional distributed tomographic method is used. For the EEG inverse problem, the solution space is made up of a distribution of voxels (points in three-dimensional space). A single point source is placed on each voxel. Every source is defined by a current density vector with unknown moment components.

The mathematical relationship between scalp potentials and current density can be expressed by the following equation [22]:

$$\Phi = KJ + c1 \quad (36.1)$$

where  $\Phi \in R^{N_{Ex}x1}$  is a vector containing the measurements of scalp potential differences at  $N_E$  electrodes with respect to a reference electrode;  $K \in R^{N_{Ex}(3N_V)}$  is the lead field matrix corresponding to  $N_V$  voxels;  $J \in R^{(3N_V)x1}$  is the current density;  $c$  is an arbitrary constant; and  $1 \in R^{N_{Ex}x1}$ . The solution of the EEG inverse problem is given by

$$J = T\Phi \quad (36.2)$$

where  $T$  is the generalized inverse of  $K$ .

### 36.2.2 LORETA

Active source reconstruction is conducted by the LORETA-KEY software v20170220 ([www.uzh.ch/keyinst/loreta.htm](http://www.uzh.ch/keyinst/loreta.htm)) using the eLORETA algorithm [7].

LORETA [4] is a linear inverse method which provides a three-dimensional reconstruction of the brain electrical activity distribution. LORETA uses a digitized model of the human brain, created by the Brain Imaging Center of the Montreal Neurological Institute, called “MNI brain.” The brain volume is discretized into a three-dimensional grid, and the sources are placed on each grid point. The solution space is limited to gray matter and hippocampus.

LORETA assumes that neighboring grid points are more likely to have similar orientation and activation strength than distant grid points. From all possible solutions that could explain a given set of data, LORETA selects the smoothest. Mathematically, this is achieved by introducing a discrete spatial Laplacian operator whose inverse matrix implements a smoothing operator able to blur abrupt discontinuities in the solution. But this smoothest solution also implies that although the location of the maximal activity is preserved, the amount of dispersion around it increases (*blurring*).

The general inverse problem is stated as [22]

$$\min_J F_W \quad (36.3)$$

with

$$F_W = \|\Phi - K J\|^2 + \alpha J^T W J \quad (36.4)$$

where the vector  $\Phi \in R^{N_E \times 1}$  contains the scalp electric potential differences measured at  $N_E$  electrodes with respect to a single common reference electrode;  $K \in R^{N_E \times (3N_V)}$  is the lead field matrix corresponding to  $N_V$  voxels;  $J \in R^{(3N_V) \times 1}$  is the current density; and  $\alpha \geq 0$  is a regularization parameter.

The solution is

$$\hat{J}_W = T_W \Phi \quad (36.5)$$

with the pseudoinverse given by

$$T_W = W^{-1} K^T (K W^{-1} K^T + \alpha H)^+ \quad (36.6)$$

The centering matrix  $H$  in Eq. 36.6 is the average reference operator. The weight matrix  $W$  is block diagonal, with subblocks of dimension  $3 \times 3$  for each voxel.

In the case of LORETA, the matrix  $W$  represents the squared spatial discrete Laplacian operator.

The eLORETA weights satisfy the following equation:

$$W_v = \left[ K_v^T (K W^{-1} K^T + \alpha H)^+ K_v \right]^{1/2} \quad (36.7)$$

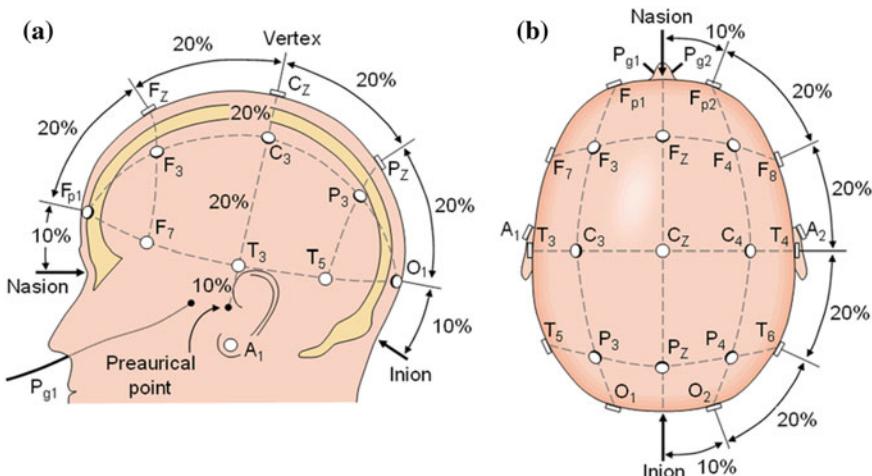
where  $W_v \in R^{3x3}$  is the v-th diagonal subblock of  $W$ .

The eLORETA method is standardized so its theoretical expected variance is unity. eLORETA presents no localization bias in the presence of measurement and structured biological noise [7].

### 36.2.3 Acquisition System

The international 10–20 system [23] has been considered the de facto standard for electrode placement for a long time. According to this system, introduced in 1958, the distance among the electrodes is set on the basis of the following landmarks over the head surface: the nasion, the inion, and the left and right preauricular points. The nomenclature “10” and “20” refers to 10 and 20% of the distance between inion and nasion and between left and right preauricular points. The electrode number placed on the scalp is up to 21. Each electrode location is identified by letters and numbers. Fp, F, C, P, and O stand for the frontopolar, frontal, central, parietal, and occipital areas, respectively. Even numbers are used for the right hemisphere and odd numbers for the left one. “z” refers to the midline (Fig. 36.1).

The need for improving the spatial resolution of the EEG led to some extensions of the 10–20 system:



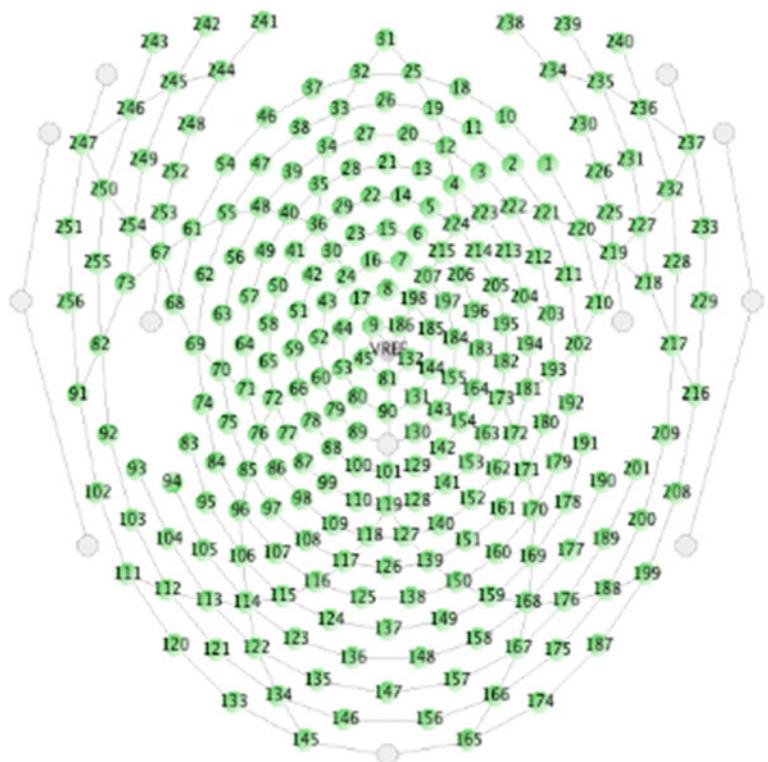
**Fig. 36.1** Electrode position on the scalp according to the 10–20 system. **a** Lateral view of skull; **b** superior view of skull

- the 10–10 system, first proposed by Chatrian [24] with a channel density of 81, later modified and accepted as a standard of the American Clinical Neurophysiology Society and the International Federation of Clinical Neurophysiology [25];
- the 10–5 system [26], which describes more than 300 electrode positions.

Nowadays, EEG systems up to 256 channels are commercially available.

The acquisition system used in this study is the 256-channel HydroCel Geodesic Sensor Net (Fig. 36.2).

Each line between sensor pairs is a geodesic, the shortest distance between two points on the surface of a sphere. The accurate geodesic tessellation of the head surface optimizes the sampling of the electrical field [27].



**Fig. 36.2** 256-channel HydroCel Geodesic Sensor Net (HCGSN)

## 36.3 Results

### 36.3.1 Data Description

The dataset consists of two EEGs recorded by the 256-channel HydroCel Geodesic Sensor Net from two patients with gliotic lesions. The EEGs were recorded during eyes-closed resting conditions. The data have been cleaned from artifacts using a preprocessing. Only 173 channels from the starting 256 were selected because the recordings from electrodes placed on the face and the neck were too noisy. The EEGs were filtered at 0.5 Hz low cutoff (high-pass) and at 48 Hz high cutoff (low-pass). The sampling rate was set at 100 Hz. The EEG recordings were segmented into artifact-free non-overlapping epochs of 256 samples [28]. The EEGs were transformed to a common average reference montage.

The 18-channel and 64-channel configurations were obtained on the basis of the 10–10 position equivalence for the HydroCel GSN [29].

The data were recorded at IRCCS Centro Neurolesi Bonino-Pulejo of Messina.

### 36.3.2 Analysis of Result

The brain activity of each patient was computed by LORETA-KEY software for the following frequency bands: delta (0.5–4 Hz), theta (4–8 Hz), alpha1 (8–10.5 Hz), alpha2 (10.5–13 Hz), beta1 (13–18 Hz), beta2 (18–30 Hz), gamma (30–40 Hz), and total band (0.5–40 Hz). The sequence of operations performed by the software is summarized as follows (Fig. 36.3):

The three-dimensional current density distribution was evaluated for three different electrode configurations.

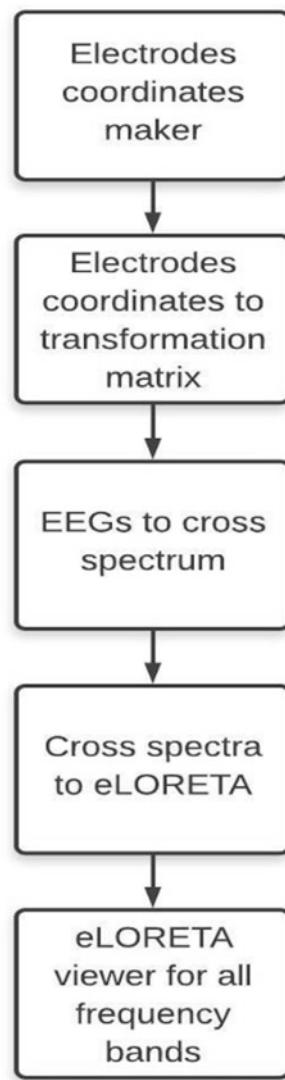
Figure 36.4 shows the current density field of patient 1 in beta1 band.

All the images reveal that the most activated area is the frontotemporal region. In particular, the images obtained in the case of the 173-channel EEG (Fig. 36.4c) best match the structural maps generated by the magnetic resonance imaging (MRI) of patient 1. The area near the gliotic lesions shows a greater activation. This could be caused by the mechanisms of neuroplasticity, according to which new neural connections could be created near the lesion and generate a growth of brain activity. These results suggest that the source localization accuracy improves significantly with the increase of the number of electrodes.

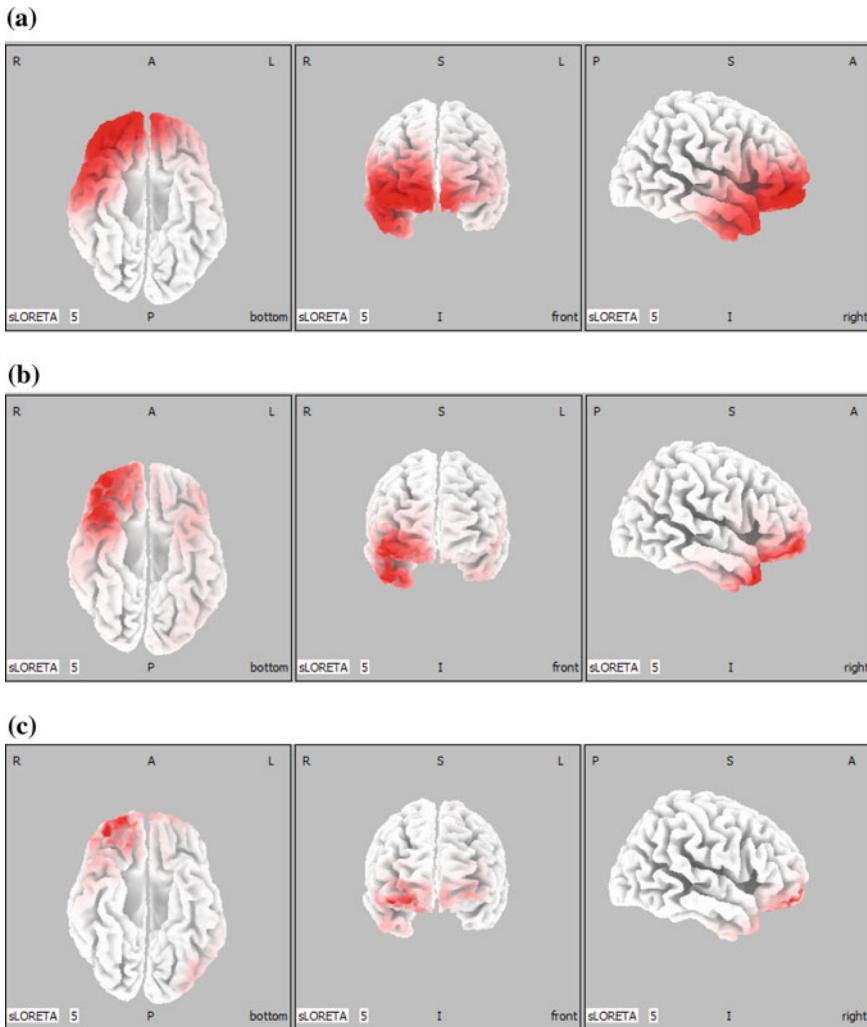
Figure 36.5 displays the current density field of patient 2 in delta band. The pictures show that the application of LORETA to a HDEEG allows to detect sources that are not revealed by the 10–20 configuration.

In Fig. 36.5a, only the frontotemporal area is active. In Fig. 36.5b, a frontoparietal activity in both hemispheres appears. Finally, Fig. 36.5c highlights a greater activation in the left frontoparietal area and the sources of the frontotemporal regions previously detected by the 10–20 electrode configuration become barely visible. Also in this

**Fig. 36.3** Flowchart of computing eLORETA source reconstruction



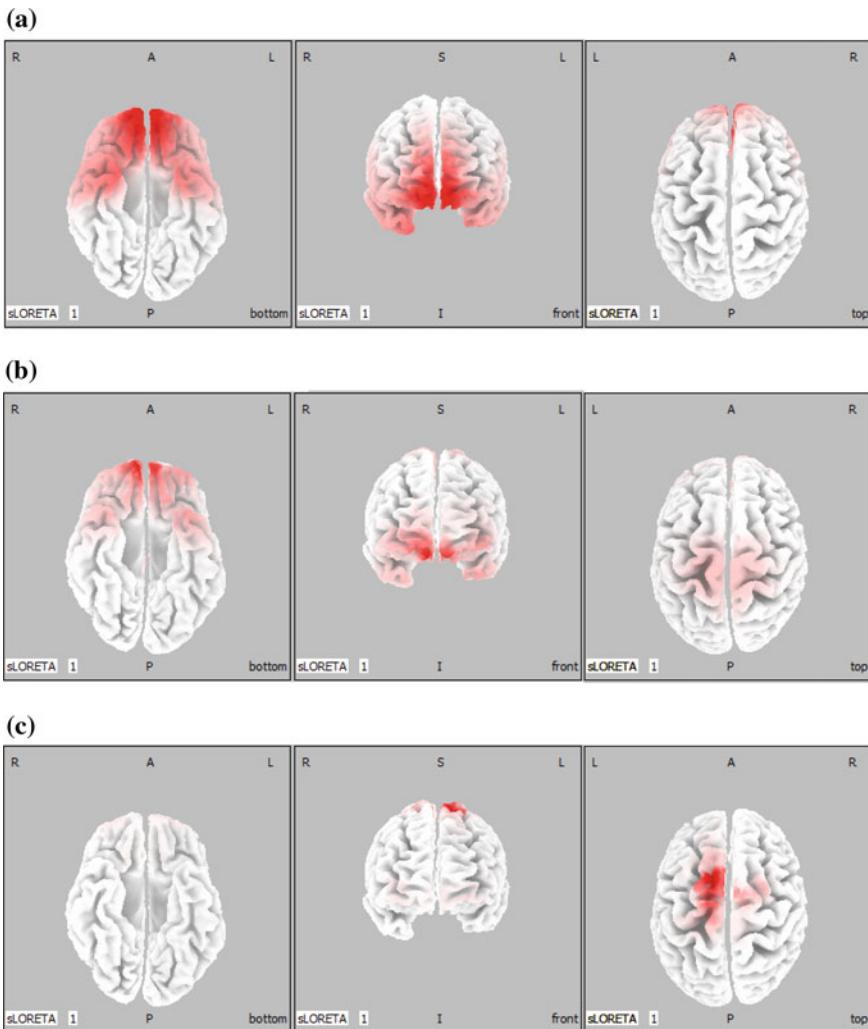
case, the images obtained with the 173-channel EEG (Fig. 36.5c) agree with the structural maps generated by the magnetic resonance imaging (MRI) of patient 2, whose gliotic lesions are placed near the left frontoparietal region.



**Fig. 36.4** Brain source distribution of patient 1 in beta1 band for **a** 18, **b** 64, and **c** 173 electrodes

## 36.4 Discussion

EEG is a tool widely used in the study of brain activity. The main drawback of the standard EEG is its low spatial resolution, which could be improved employing a greater number of electrodes (HDEEG). The high density of sensors is useful to enhance accuracy in the localization of the brain active sources, regardless of the method for solving the inverse problem. In particular, this paper deals with the effect of sensor density on eLORETA source localization accuracy. Previous research analyzed the difference between low and high resolution, focusing on EEG of epileptic



**Fig. 36.5** Brain source distribution of patient 2 in delta band for **a** 18, **b** 64, and **c** 173 electrodes

patients. The increasing number of sensors has shown a decrease in the localization error of seizure onset zones [20, 21, 30].

In our paper, we study if and how the localization accuracy of the most activated brain regions changes, depending on the number of electrodes. The results reveal that the edges of the area with a more intense brain activity are better defined when sensor number increases. Then, some sources marked as active when the 10–20 system is used appear to be less intense when they are detected by 173 electrodes. Finally, there are sources detected only when the high-density configuration of electrodes is used.

## 36.5 Conclusions

The simulations show that LORETA improves its performance when the sensor density increases. In particular, three important aspects arise from this study:

1. The regions with stronger activity detected by the 10–20 system electrode are defined in a better way when a greater number of electrodes are used.
2. A few number of sensors could be not sufficient to detect all active sources.
3. Sources marked as significant when the 10–20 system configuration is used could turn out to be less significant when high-density sensor configurations are employed.

The limitation of the current study is that only two EEGs were analyzed. The matter of eLORETA source reconstruction accuracy needs further development using a larger database.

So, it could be determined the minimum number of electrodes to reconstruct the active sources in an accurate way not only in the case of epilepsy but for brain pathologies in general.

**Acknowledgements** The present work was funded by the Italian Ministry of Health, Project Code GR-2011-02351397.

## References

1. Wang, G., Ren, D.: Effect of brain-to-skull conductivity ratio on EEG source localization accuracy. In: BioMed Research International 2013 (2013)
2. Jatoi, M.A., Kamel, N., Malik, A.S., Faye, I., Begum, T.: A survey of methods used for source localization using EEG signals. Biomed. Sig. Process. Control **11**, 42–52 (2014)
3. Hämäläinen, M.S., Ilmoniemi, R.J.: Interpreting Measured Magnetic Fields of the Brain: Estimates of Current Distributions. Helsinki University of Technology, Department of Technical Physics, Espoo (1984)
4. Pascual-Marqui, R.D., Michel, C.M., Lehmann, D.: Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. Int. J. Psychophysiol. **18**(1), 49–65 (1994)
5. Pascual-Marqui, R.D.: Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. Methods Find. Exp. Clin. Pharmacol. **24**(Suppl D), 5–12 (2002)
6. Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E.: Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron **26**(1), 55–67 (2000)
7. Pascual-Marqui, R.D.: Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization. arXiv preprint [arXiv:0710.3341](https://arxiv.org/abs/0710.3341) (2007)
8. Jatoi, M.A., Kamel, N., Malik, A.S., Faye, I.: EEG based brain source localization comparison of sLORETA and eLORETA. Australas. Phys. Eng. Sci. Med. **37**(4), 713–721 (2014)
9. Mulert, C., Jäger, L., Schmitt, R., Bussfeld, P., Pogarell, O., Möller, H.J., et al.: Integration of fMRI and simultaneous EEG: towards a comprehensive understanding of localization and time-course of brain activity in target detection. Neuroimage **22**(1), 83–94 (2004)

10. Vitacco, D., Brandeis, D., Pascual-Marqui, R., Martin, E.: Correspondence of event-related potential tomography and functional magnetic resonance imaging during language processing. *Hum. Brain Mapp.* **17**(1), 4–12 (2002)
11. Rossini, P.M., Del Percio, C., Pasqualetti, P., Cassetta, E., Binetti, G., Dal Forno, G., et al.: Conversion from mild cognitive impairment to Alzheimer's disease is predicted by sources and coherence of brain electroencephalography rhythms. *Neuroscience* **143**(3), 793–803 (2006)
12. Gianotti, L.R., Küng, G., Lehmann, D., Faber, P.L., Pascual-Marqui, R.D., Kochi, K., Schreiter-Gasser, U.: Correlation between disease severity and brain electric LORETA tomography in Alzheimer's disease. *Clin. Neurophysiol.* **118**(1), 186–196 (2007)
13. Babiloni, C., Frisoni, G.B., Pievani, M., Toscano, L., Del Percio, C., Geroldi, C., et al.: White-matter vascular lesions correlate with alpha EEG sources in mild cognitive impairment. *Neuropsychologia* **46**(6), 1707–1720 (2008)
14. Worrell, G.A., Lagerlund, T.D., Sharbrough, F.W., Brinkmann, B.H., Busacker, N.E., Cicora, K.M., O'Brien, T.J.: Localization of the epileptic focus by low-resolution electromagnetic tomography in patients with a lesion demonstrated by MRI. *Brain topography*, **12**(4), 273–282 (2000)
15. Clemens, B., Bessenyei, M., Fekete, I., Puskás, S., Kondákor, I., Tóth, M., Hollódy, K.: Theta EEG source localization using LORETA in partial epilepsy patients with and without medication. *Clin. Neurophysiol.* **121**(6), 848–858 (2010)
16. Spitzer, A.R., Cohen, L.G., Fabrikant, J., Hallett, M.: A method for determining optimal inter-electrode spacing for cerebral topographic mapping. *Electroencephalogr. Clin. Neurophysiol.* **72**(4), 355–361 (1989)
17. Tucker, D.M.: Spatial sampling of head electrical fields: the geodesic sensor net. *Electroencephalogr. Clin. Neurophysiol.* **87**(3), 154–163 (1993)
18. Freeman, W.J., Holmes, M.D., Burke, B.C., Vanhatalo, S.: Spatial spectra of scalp EEG and EMG from awake humans. *Clin. Neurophysiol.* **114**(6), 1053–1068 (2003)
19. Srinivasan, R., Tucker, D.M., Murias, M.: Estimating the spatial Nyquist of the human EEG. *Behav. Res. Methods* **30**(1), 8–19 (1998)
20. Sohrabpour, A., Lu, Y., Kankirawatana, P., Blount, J., Kim, H., He, B.: Effect of EEG electrode number on epileptic source localization in pediatric patients. *Clin. Neurophysiol.* **126**(3), 472–480 (2015)
21. Song, J., Davey, C., Poulsen, C., Luu, P., Turovets, S., Anderson, E., et al.: EEG source localization: sensor density and head surface coverage. *J. Neurosci. Meth.* **256**, 9–21 (2015)
22. Tong, S., Thakor, N.V.: Quantitative EEG Analysis Methods and Clinical Applications. Artech House, Norwood (2009)
23. Jasper, H.H.: The ten twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.* **10**, 371–375 (1958)
24. Chatrian, G.E., Lettich, E., Nelson, P.L.: Ten percent electrode system for topographic studies of spontaneous and evoked EEG activities. *American Journal of EEG Technology* **25**(2), 83–92 (1985)
25. Jurcak, V., Tsuzuki, D., Dan, I.: 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage* **34**(4), 1600–1611 (2007)
26. Oostenveld, R., Praamstra, P.: The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.* **112**(4), 713–719 (2001)
27. Electrical Geodesics, I.: Geodesic sensor net technical manual. Electrical Geodesics, Inc (2007)
28. Nunez, P.L., Srinivasan, R., Westdorp, A.F., Wijesinghe, R.S., Tucker, D.M., Silberstein, R.B., Cadusch, P.J.: EEG coherency: I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalogr. Clin. Neurophysiol.* **103**(5), 499–515 (1997)
29. Luu, P., Ferree, T.: Determination of the HydroCel Geodesic Sensor Nets' average electrode positions and their 10-10 international equivalents. Inc, Technical Note (2005)
30. Yang, L., et al.: Dynamic imaging of ictal oscillations using non-invasive high-resolution EEG. *Neuroimage* **56**(4), 1908–1917 (2011)

## Chapter 37

# To the Roots of the Sense of Self: Proposals for a Study on the Emergence of Body Awareness in Early Infancy Through a Deep Learning Method



Alfonso Davide Di Sarno, Raffaele Sperandeo , Giuseppina Di Leva,  
Irene Fabbricino, Enrico Moretto , Silvia Dell'Orco   
and Mauro N. Maldonato

**Abstract** In the psychological field, the study of the interaction between the caregiver and the child is considered the context of primary development. Observing how infants interact with their mothers, in the first year of their life, has allowed researchers to identify behavioural indicators useful to highlight the crucial role of their synchronisation, within the behavioural and cognitive domain. The aim of this study is to employ a new observational method, mediated by an artificial neural network model, to systematically study the complex bodily gestural interaction between child and caregiver, which concurs in the development of the child's bodily awareness, and how the quality of early interactions affects the child's relationship with his body. The first part will deal with qualitative observational methods, their limits and the object of the study; the second part will describe the reference theory; the third part will describe the neural network model, the details of the proposed study and the

---

A. D. Di Sarno · R. Sperandeo · G. Di Leva · I. Fabbricino · E. Moretto · S. Dell'Orco  
SiPGI—Postgraduate School of Integrated Gestalt Psychotherapy, Torre Annunziata, Naples, Italy  
e-mail: [a.davidedisarno@gmail.com](mailto:a.davidedisarno@gmail.com)

R. Sperandeo  
e-mail: [raffaele.sperandeo@gmail.com](mailto:raffaele.sperandeo@gmail.com)

G. Di Leva  
e-mail: [giuseppinadileva@hotmail.com](mailto:giuseppinadileva@hotmail.com)

I. Fabbricino  
e-mail: [irenefabbricino@hotmail.com](mailto:irenefabbricino@hotmail.com)

E. Moretto  
e-mail: [enrico.more@gmail.com](mailto:enrico.more@gmail.com)

S. Dell'Orco  
e-mail: [silviadellorco@gmail.com](mailto:silviadellorco@gmail.com)

M. N. Maldonato  
Department of Neuroscience and Reproductive Sciences and Odontostomatology,  
University of Naples Federico II, Naples, Italy  
e-mail: [nelsonmauro.maldonato@unina.it](mailto:nelsonmauro.maldonato@unina.it)

expected results. We hypothesise that this model could be used for the construction of a tool that could compensate for some of the critical shortcomings that lie within observational methods in the field of psychology.

### 37.1 Introduction

The process through which one's own body's representation is structured in humans has progressively become a subject of growing interest in the scientific community [1]. In literature with the term "body awareness", we refer to a complex multidimensional construct, an emergent process that involves sensory awareness, defined as the ability to identify inner sensations, physiological and emotional state of one own body [2].

Daniel Stern, following an interpersonal approach, referred to a complex global psychic organisation; according to his perspective, the construction of a sense of self is delineated when the individual begins to experience himself as a physical entity with defined boundaries. The self would thus progressively acquire the status of "existential entity", which is defined through actions that stem from personal corporeity, in relation to the space-time coordinates that it perceives in the external world and in the movements of other individuals [3–5]. In the same way, Gallese identifies within body awareness the most likely nucleus of the various forms of self-consciousness [6, 7].

The neural correlate for body representation has been identified within two brain networks: the fronto-parietal network and the one responsible for sensory motor control [8]. The sensory motor control network, whose structure includes motor and somatosensory cortex, basal ganglia, thalamus and cerebellum, is peculiar with regard to the fact that it is largely developed from the age of 2 and is implicated in the formation of body representation and in the correction of movements as they occur. The fronto-parietal network extends from the inferior frontal gyrus to the posterior parietal cortex, which, integrating environmental information and physical information from the body, provides a representation of one's own body [9, 10] in relation to the personal domain as well as the external environment. The fronto-parietal network, conversely to the somatosensory network, has a significantly longer developmental course. Through the use of fMRI, it has in fact been possible to demonstrate that its structuring process extends well beyond adolescence, into adult life [11].

Our existence in the world as corporeal beings is tightly linked to movement and to our potential ability to manipulate our environment, motivated by motor intents. The study of body awareness is of particular interest as it concerns the clinical domain, as evident from the study of phenomena related to a lack of body awareness, such as psychotic spectrum disorders [12] and dissociative phenomena [13, 14].

In this context, empirical research has allowed us to structure a neurocognitive model that highlights how the feeling of owning a body derives from the interaction

between multisensory inputs and the internal representation model of how one's body is structured [15, 16].

All aforementioned research put movement and intent at the base of the emergence of the Self, so, translating these concepts to the particular dynamic that is observed within the mother–infant dyad, it is possible to define their movement in space as a dynamic-relational act that allows for the development of body awareness. The emergence of body awareness in the infant is necessarily mediated by the awareness and the relationship that the caregiver entertains with their own body; thus, focusing our observation on the harmony (or lack thereof) within the movement of the dyad is, within this framework, of crucial importance to define the development of self-awareness, not as a phenomenon exclusive to the infant, but as a relational construction related to the mother–infant–environment relationship.

In this work, we deal with this topic through different steps: in the first place the features, the critical issues and the limits of the observation method in psychology will be discussed, giving particular attention to the infant–caregiver observation; it will follow a description of the theory of Frank [17] that theorises a phenomenological model of dyadic observation centred on six fundamental movements; lastly, our research proposal for the creation and the implementation of a deep learning video analysis model for infant–caregiver video registration will be presented.

## 37.2 The Observational Methods in Psychology

The main aims of qualitative observation are to understand the complexities that characterise human and social phenomena, as well as to highlight the role of a particular context, in order to produce knowledge grounded in time and space [18].

Observation can be an appropriate method for the study of human behaviour [19, 20]. Within interpersonal relationships, along with an overt content constituted by verbal elements, meta-communicative elements are essential to the correct attribution of meaning, but are often hard to identify in a superficial analysis of any interaction [21–23].

One of the main tasks for those pursuing observation is thus to establish a direct contact with what is observed, to create a vision as complete and accurate as possible [24]; to this end, naturalistic observation is often employed, defined as a type of observation where the researcher avoids influencing subjects' behaviours, or to project their ideas and preconceptions on what is observed [25].

Nevertheless, observation has substantial limitations: one is lack of control over the environment, where undesirable variables which could influence data cannot be manipulated, sometimes hindering or nullifying the results of the study; another limitation is the inherent difficulty classifying data in a quantitative way, as often the data gathered is plentiful but hard to systematically classify; additionally, the presence of an observer can nevertheless have an effect on the subjects of the observation, and the data-gathering process can be affected as well; finally, observational research is oftentimes based on a sample of reduced size compared to quantitative research [26].

### 37.3 The Observation of the Dyad

The interaction between meaningful adult and child (usually intended like mother and child as systems open to one another) is considered the context of primary development, in which to perceive oneself becomes perceiving the other, in a continuous interexchange where body and environment become indissoluble [27–29], an interactive process in which both partners, with different levels of competence, reciprocally influence each other [30].

Observing how infants in the first moments of life relate to their mothers, how the pair develops the ability to coordinate, or to react to the dissonances that can arise in the interaction system, has contributed to the subject of what infants understand of others, their affective states and their interactions [31]. Such observations have allowed researchers to identify categories and behavioural indicators useful when describing these exchanges, which have been proven to be extremely effective to clarify, on an empirical level, not only the profile of the infant, but also that of the adult involved in the interaction, and to highlight the crucial role of their synchronisation, within the behavioural, affective and cognitive domain.

The main difference between observation and other data-gathering methods is how data is recorded by the observer, whose judgement acts as a filter for all information. Observation is a complex activity, requiring time, intellectual freedom, distension and self-awareness [32].

Although observation can be influenced by subjectivity, and thus to errors and gaps in information, it becomes objective when it is conducted through systematic, replicable and communicable procedures [33].

Our choice to objectify an observational method through the use of an artificial neural network answers to our need to study phenomena and behaviours without compromising on scientific rigour and internal validity and, by guaranteeing the relevance and generalisability of the results, on external validity. Normally, the relationship between these two types of validity is inversely proportional (as internal validity increases in a controlled experiment, external validity decreases) [34].

As such, researchers' involvement must not create any confusion between the observer and the object of observation, preventing the development of the uniqueness and authenticity of the dyad observed. The mother–infant pair which we observe is not stable, but in continuous evolution. To observe it therefore implies accepting such changes, yet still establishing bonds and giving a unified meaning to behaviours that progressively emerge within the relationship between mother and infant [35].

We could state that neutrality and objectivity are a useful goal for the observer, but seem in fact an abstract concept that, by definition, will never become concrete [36, 37].

It is of extreme importance to know how social relationships are structured, from their origins in early infancy, and to be equipped with valid instruments to observe interactions as micro-exchanges upon which any relationship is based.

## 37.4 Theory of Reference

### 37.4.1 *Movements: The Child First Language*

The analysis of fundamental movements is at the base of the somatic evolutionary approach theorised by Ruella Frank e Frances La Barre, who structure a conceptual framework that employs the phenomenological observation method aimed at understanding the interactive movement patterns of the mother–infant dyad [38]. Frank centres her investigation around movement as the starting point for observation of interaction modes and of contact with the external environment. The characteristic movements of an infant are reciprocated by the parent’s movements in a constant process of communication and co-regulation. Children receive feedback and information regarding their competences from their parents. The parents can also encourage or discourage the children’s efforts toward motor maturity and can influence how the children perceive themselves by producing for them opportunities to make new motor experiences (for example during play sessions) [39].

Movement, as theorised by Frank, is organised in terms of rhythmic contractions and muscular releases that push the limbs away from or towards the body, with particular reference to “evolutionary schemes of movement” [38, 40] defined as recurrent patterns of rhythmical movement and muscle released guided by the infant’s intention to reach towards what they need and to withdraw from what they deem undesirable or dangerous [16].

According to the author, these schemes are the origin of all types of learning related to the elaboration of experiences in early life; such learning processes progressively shape the schemas which infants employ to overcome obstacles or to satisfy their curiosity and needs. Within this model, the emerging schemas for breathing, gestures, posture and gait of the subject are observed through an evolutionary lens that focuses on movement under a structural/functional perspective.

Infants employ a non-verbal vocabulary made of micro-movements: they experiment on themselves when yielding with, pushing against, reaching for, grasping on to, pulling towards and releasing from [38, 41, 42].

This is the origin of “kinaesthetic knowledge”, a phenomenon that occurs when one experiences their body in relation to others: the experience of movement becomes the first manner of knowing the surrounding environment in a way where body movement allows not only to explore the environment, but at the same time to generate kinaesthetic data at the base of the awareness from which the infant can discriminate their own body parts and differentiate “Me—Other than Me” [43]. Any time big or small changes in situations occur—such as a shift in the infant’s position—a change in perceived movement quality and general muscular tone of the body also occurs. The postural aptitude is the perceived ground from which actions emerge and feelings develop. When a change in situation occurs, we can also observe a change in quality of movement, feelings and postural tone. In the same way, when feelings and actions emerge, postural attitudes can change [44, 45]. Movements, feelings and postural tones can thus be considered inextricably intertwined.

Such experiences do not require mediation from thought nor of concepts other than the immediate evaluation of the situation carried out by the person. This delicate interconnection, along with feedback from our sensory organs and from touch, is at the base of awareness. By moving in their environment, infants develop a prereflexive kinaesthetic orientation which allows them to be aware of the existence of others.

### ***37.4.2 The Six Fundamental Movements Theory***

In this context, Frank identifies six fundamental movements that can be considered “a child’s first language”: a language that is exclusively linked to body movements and that is present and essential for all human relationships. These fundamental movements are interdependent from one another, and they act as a base for all types of actions and emerge as a response to the surrounding environment: they work as a function of the relationship, so each parent–child relationship creates a personal movement vocabulary. In other terms, although all the fundamental movements are utilised, each relationship is characterised by the formation of specific patterns in which certain movements are prevalent, and are performed more often than others. Our understanding of the role these patterns have in creating relationships can be applied to recognise and treat the various interaction difficulties that arise during childhood and persist throughout one’s lifetime, given how as development proceeds these schemas do not disappear, but rather evolve into progressively more complex patterns. All healthy infants develop and employ these basic movements, but not in the same way nor at the same time during development. As a matter of fact, many variations in fundamental movements exist, as many as the possible relational domains. In each relational domain, the infant organises a personal movement vocabulary (an individual variation of the fundamental movements) that includes all parts of the body: head, coccyx, arms, hands, legs and feet.

To ensure a better understanding of the movement units we intend to observe, we will briefly describe the fundamental movements as originally theorised by Frank [38].

1. “Yielding with” corresponds to the experience of giving into another as well as taking from another. It is a full-body experience involving all fundamental physical forces (gravity, ground and space). The child experiences higher pressure on those parts of the body that come into contact with a surface: such an increase in pressure causes a slight compensation in muscle tone that creates a feeling of weight, flowing from the centre to the periphery experiencing balance and stability, which can be described as the child’s ability to maintain their barycentre in relation to a supporting surface [46].
2. When “Pushing against” the weight of the body is focused on the peripheral areas (head, hands, feet coccyx, pelvis) towards the centre of the body, inducing an experience of density. While the infant pushes, their experience of weight is condensed in the point of origins of the pushing (whether it be the head, coccyx,

arms or legs). When an infant pushes, he is able to simultaneously experience that he is separated from others while including that in their range of experience. If “Yielding with” allows to join others, “Pushing against” represents the act of differentiating oneself from others: such behaviours become essential in the relationship domain, influencing one’s caregivers, their responsiveness and the ways through which they take care of the child [47].

3. “Reaching for” is the act of extending oneself, within the environment, to explore the space around one’s body. The importance of the acquisition of this skill is highlighted by evidence demonstrating how a failure in reaching can substantially compromise the acquisition of superior neuromotor function [48]. In each extension movement aimed at reaching others, the infant expands himself through their environment, moving further away from their barycentre, hindering their balance, but becoming increasingly able to explore the surrounding space.
4. “Grasping on to” is the act of grasping and containing what has been reached. While they grab and hold an object with their hands or mouth, infants develop an understanding of the object itself, learning about its possible uses. A child’s exploration through their eyes, mouth and hands integrate and strengthen one another.
5. “Pulling Toward” corresponds to the act of taking what is around to make it one’s own. In this way, we experience others’ qualities, the attraction towards it, its resistance and flexibility. The co-action that is organised in this movement contributes to the experience of the self with others, of the concept of “us”. It is crucial within the development of the self, as it is at the base of a healthy autonomy, given how it represents the ability to differentiate oneself from others, as well as the ability to give up separation and join others.
6. When “Releasing from” the child leaves the object, they were holding on to and allow for the reorganisation of the whole body. When we release, the action is complete, and the “yielding with” phase becomes once again evident. In the behavioural repertoire of an infant, the action of letting go of an object is initially passive. From the fifth month of age, the infant holding on to an object can actively decide to release it.

### ***37.4.3 Effects of the Mismatches in the Relationship***

Observing mismatches between mother and child, that is to say when response patterns to each other’s movements are different and contrasting, can allow researchers to establish whether a certain pattern has developed in the infant, whether the mother is able to spot it and understand it, and whether this has evolved into a predictable routine within the dyad. For example, in certain occasions the infant reaches towards the mother (reaching) and her response is to push him away (pushing), misinterpreting the needs of the child, tapping into and responding through her own movement pattern learned throughout her lifetime: oftentimes, in cases such as these, parents’

faulty interpretation of their child's actions can originate from their unconscious, unanalyzed corporeal history [49].

These results are supported by consistent reports in previous literature that describe how a child's production of cortisol, a stress hormone, is coordinated with that of the mother, so, as parental stress increases, a child's susceptibility to stress becomes elevated, while it can be moderated if the caregiver-child relationship is characterised by behavioural reciprocity and synchronisation [50]. A synchronistic relationship would thus develop a dynamic process in which hormonal, physiological and behavioural signals are shared between parent and child during social contact [51]. The movements we make are not limited to reflecting a mental state, but rather give them shape [52–55].

## 37.5 The Deep Learning Model Applied to Human Behaviour

Deep learning is a specific form of machine learning, wherein elevated quantities of input are sent to artificial neural networks to improve their ability to “learn” as increasing numbers of data are elaborated. The term “Deep” refers to the levels that the neural network accumulates, with its performance improving as the depth of the network increases. The analysed signals are used in a process of weight adjustment, increasing its knowledge and progressively approximating to the set objective [56].

The use of neural networks for the analysis of human behaviours and movements is particularly complex, as these processes create time-dependant structures that possess inter- and intra-individual differences and are extremely susceptible to be influenced by the type of activity and context in which they occur [57–59]. This highlights the necessity to use complex models that operate motion recognition, its tracking and its analysis/interpretation. With the term “action”, we mean the simplest unit of meaning at the base of more complex behaviours, and this type of network is particularly suitable to handle input for which time is a relevant factor and where time between events is variable [60].

When people produce gestures, they move hands and arms in a particular sequence: starting from a resting position, they adjust their arms in an area in front of their body (preparation), they perform other movement that can be described as the signifying part of a gesture, the one that will be interpreted by their interlocutor (stroke), to return to a resting position (retraction/rest position). The single gestural units build a complex movement pattern that can be subject to variations and interruptions [61].

This model can be useful to develop a movement recognition system, given the necessity to go through a segmentation process for the moving body parts. Each movement has specific geometric coordinates that construct a complex configuration (pattern), a cluster of movements that can be simplified and automatically recognised

by the network, translating the data into a quantitative form that can be inserted in a database [57].

### 37.5.1 Long Short-Term Memory Recurrent Neural Network

Tsironi et al. [62] describe an interesting model called convolutional long short-term memory recurrent neural network (CNNLSTM). The architecture of this system employs two convolutional layers, a flattening layer a long short-term memory recurrent layer and a softmax output layer, in order to allow for the recognition of gestural dynamics. This model proposes the combination of two neural networks: the convolutional neural network (CNN), created to allow for the recognition of visual patterns directly from images represented by pixels [63] and the long short-term memory (LSTM) [64]. The most peculiar feature of the latter model is its inclusion of a memory unit that allows to save information regarding long-term temporal dependencies, making it ideal for the prediction and the classification of temporal sequences. The composition of this type of network features three types of gates: input gate, output gate, and forget gate. The presence of these gates allows cells (LSTM blocks) to retain information for an indeterminate amount of time, as well as allowing them to determine which information has the characteristics that make it suitable for being retained in their memory and which information should be “forgotten”, while the output gate decides whether the value of the cells must conclude its path through the subsequent units up until the exit [65]. The target movements that we are aiming to identify to provide the input for the network will be inserted into a dataset featuring the six fundamental movements. These will be recorded through a fixed RGB camera and will be isolated through a background subtraction technique.

This type of network is trained by using images of the sequence of movements which are temporally segmented with backpropagation through time (BPTT). Each differential image is inserted in the network with its own specific label, so that, when they are separately processed by the network, the output layer will feed back that specific class of movements [66].

This network model will be applied to video recordings of dyadic interactions, with the aim to analyse the video to derive patterns of movement and recursiveness.

## 37.6 Study Goals

For this study, we propose to train a neural network structured as long short-term memory (LSTM) to recognise the fundamental movement units identified in our Theory of Reference that, in spite of being based on empirical evidence and in line with recent neuroscientific discoveries, lacks a structured and standardised observation grid.

Our hypothesis is that it is possible to create an observation tool of the dyadic interaction that, using a neural network inspired model, can identify and recognise the fundamental movements that characterise the complex interactions between parent and child, allowing us to formulate inferences on the modalities through the dyad, structure the synchronicity, the reciprocity and the necessary responsiveness so that the child can develop bodily awareness, a good relationship with his own body and make experience about himself and the surrounding world.

Moreover, given the complexity of the subject, we want to render accessible an instrument for observation, as employable by researchers with varying degrees of competence.

### 37.7 Method

Our study can be classified within the domain of human activity recognition (HAR) experiments. HAR is the sector of machine learning whose aim is to build techniques to recognise and interpret human movement, through the use of sensors that allow a machine to automatically learn how to classify human activities [67, 68].

The main aim of the study is to test and validate an electronic system to observe the parent-child dyad, and to this end, the study takes into account the involvement of parents and children aged 12–24 months that attend a kindergarten within the geographical area of Naples.

The research itself will be comprised of the following phases: stipulation of a convention with the school; presentation of the proposed research and gathering of participation agreements and informed consent from the parents; gathering of the medical histories of parent and child, and administration of personality tests as well as evaluation of any possible affective disorders; video recording of free interactions between mother and child, which will then be analysed with the tool that we are proposing to create, and by a group of experts in observation of mother-child dyads that will evaluate on parallel with the tool to obtain data for comparison in order to evaluate if our movement recognition program is reliable.

This study complies with the Associazione Italiana di Psicologia (AIP) ethical code and will be submitted to the Federazione Italiana delle Scuole e Istituti di Gestalt (FISIG) ethical committee.

### 37.8 Conclusion and Expected Results

Numerous works are focused on the advancement regarding the understanding of the complex dynamics that intervenes between the caregiver and his/her own child [69, 70]; however, the number of experiments which have proposed mediate methods through technological tool is small [71–73].

Based on the scientific evidence considered, we identified the need for a tool that could compensate for some of the critical shortcomings that lie within observational methods in the field of psychology.

The use of an artificial neural network could provide researchers with a simple-to-use instrument, which would allow to gather and handle a great volume of qualitative data. The simplification and the hoped-for increase in coding speed for the interactions would also allow for an increase in sample size for the dyads being observed.

One of the main advantages is the possibility to directly observe a phenomenon and to analyse it in a phenomenological way, minimising the subjective interpretation factor through the application of an objective and replicable observation model. The theory for the construction of such tool, to which we refer, poses defined gestural units as a focus for observation, but considering how the model is based on phenomenology, it does not allow for inferences or interpretations; as such, it can be described as a purely observational model.

The use of such a tool would allow us to overcome another obstacle, as the observation of a phenomenon can be influenced not only by a researcher's subjectivity, but also by their level of competence: different psychologists, with different degrees of experience, will capture different aspects of the same relationship.

The analysis of fundamental movements provides a corporeal vocabulary that allows for early specific identification of problematic patterns in the mother-child interaction which can then become an object for therapy. It allows therapists and parents to visualise the complexity in the transitions of movement, which can provide both a support and a hindrance to the necessity of the child.

The systematic analysis of movements in the interaction between parent and child will allow us to widen our knowledge regarding movement patterns observed in parent-child exchanges, allowing us to highlight how the synchronisation and the construction of a shared movement vocabulary pattern have a part in the development of body awareness in the infant, and how they influence the relationship with their body and contribute to the formation of behaviours of progressively increased complexity that will be applied throughout the course of the infant's life.

Furthermore, it is possible to hypothesise that this model could be used as a diagnostic tool for early preventative identification of dysfunctional relationship patterns that can precede the development of psychopathologies by focusing on moments where the harmony in the dyad is lacking, which can create maladaptive behavioural and relational patterns [74].

## References

1. Kalckert, A., Ehrsson, H.H.: Moving a rubber hand that feels like your own: a dissociation of ownership and agency. *Front. Hum. Neurosci.* **6**, 40 (2012). <https://doi.org/10.3389/fnhum.2012.00040>
2. Mehling, W.E., Wrubel, J., Daubenmier, J.J., Price, C.J., Kerr, C.E., Silow, T., Gopisetty, V., Stewart, A.L.: Body awareness: a phenomenological inquiry into the common ground of mind-

- body therapies. *Philos. Ethics Hum. Med.* **6**(1), 6 (2011). <https://doi.org/10.1186/1747-5341-6-6>
3. Stern D.N.: Lo sviluppo precoce degli sé, dell'altro e si sé-con-l'altro. In: Stern D.N. (eds.) *Le interazioni madre-bambino*. Cortina, Milano (1982, 1998)
  4. Stern, D.N.: *The Interpersonal World of the Infant*. Basic Books, New York (1985); Trad. italiana a cura di A. Biocca e L. Marghieri Biocca, Il mondo interpersonale del bambino. Bollati Boringhieri, Torino (1987)
  5. Stern, D.N.: Prerequisiti evolutivi per il senso di un Sé narrativo. In: Stern D.N. (eds.) *Le interazioni madre-bambino nello sviluppo e nella clinica*. Cortina, Milano (1998, 1989)
  6. Gallese, V., Sinigaglia, C.: The bodily self as power for action. *Neuropsychol.* **48**(3), 746–755 (2010). <https://doi.org/10.1016/j.neuropsychologia.2009.09.038>
  7. Maldonato, M.: The birth of consciousness. The origin of temporality and the sense of self. *Hum. Evol.* **28**(1–2), 91–101 (2013)
  8. Fontan, A., Cignetti, F., Nazarian, B., Anton, J.-L., Vaugoyeau, M., Assaiante, C.: How does the body representation system develop in the human brain? *Dev. Cogn. Neurosci.* **24**, 118–128 (2017). <https://doi.org/10.1016/j.dcn.2017.02.010>
  9. Morita, T., Saito, D.N., Ban, M., Shimada, K., Okamoto, Y., Kosaka, H., Asada, M., Naito, E.: Self-face recognition shares brain regions active during proprioceptive illusion in the right inferior fronto-parietal superior longitudinal fasciculus III network. *Neurosci.* **348**, 288–301 (2017). <https://doi.org/10.1016/j.neuroscience.2017.02.031>
  10. Maldonato, M., Dell'Orco, S., Springer, M.: Rethinking consciousness: some hypothesis on the role of the ascending reticular activating system in the global workspace. In: WIRN, pp. 212–219 (2011)
  11. Zieliński, B.A., Gennatas, E.D., Zhou, J., Seeley, W.W.: Network-level structural covariance in the developing brain. *Proc Natl Acad Sci* **107**(42), 18191–18196 (2010). [https://doi.org/10.1073/pnas.101053-8119\(09\)70127-2](https://doi.org/10.1073/pnas.101053-8119(09)70127-2)
  12. Stanghellini, G.: Embodiment and schizophrenia. *World Psychiatry* **8**(1), 56–59 (2009)
  13. Sperandeo, R., Monda, V., Messina, G., Carotenuto, M., Maldonato, N.M., Moretto, E., Leone, E., De Luca, V., Monda, M., Messina, A.: Brain functional integration: an epidemiologic study on stress-producing dissociative phenomena. *Neuropsychiatric Dis. Treat.* **14**, 11 (2018). <https://doi.org/10.2147/ndt.s146250>
  14. Cantone, D., Sperandeo, R., Maldonato, M.N., Cozzolino, P., Perris, F.: Dissociative phenomena in a sample of outpatients. *Riv. Psichiatr.* **47**(3), 246–253 (2012)
  15. Filippetti, M.L., Lloyd-Fox, S., Longo, M.R., Farroni, T., Johnson, M.H.: Neural mechanisms of body awareness in infants. *Cereb. Cortex (N.Y., NY)* **25**(10), 3779–3787 (2015). <https://doi.org/10.1093/cercor/bhu261>
  16. Tsakiris, M.: My body in the brain: a neurocognitive model of body-ownership. *Neuropsychol.* **48**(3), 703–712 (2010). <https://doi.org/10.1016/j.neuropsychologia.2009.09.034>
  17. Frank, R.: Body of awareness: a somatic and developmental approach to psychotherapy. Taylor & Francis (2013). <https://doi.org/10.4324/9780203766989>
  18. Fabbri, L., Melacarne, C., Striano, M.: Emotional dimensions in transformative learning processes of novice teachers. A qualitative study. In: Paper to the ESREA Conference Emotionality in Learning and Research, Canterbury, March (2008)
  19. Bailey, K.D.: *Metodi della ricerca sociale*. Vol. I. I principi fondamentali. Il Mulino, Bologna (2006)
  20. Sempio Liverta, O., Cavalli, G.: Lo sguardo consapevole. L'osservazione psicologica in ambito educativo Unicopli (2005)
  21. Hinde, R.A. (ed.): *Non-verbal Communication*. Cambridge University Press, London (1972). <https://doi.org/10.1126/science.176.4035.625>
  22. Watzlawick, P., Beavin, J.H., Jackson, D.D.: *Pragmatica della comunicazione umana. Studio dei modelli interattivi delle patologie e dei paradossi*. Astrolabio, Roma (1967, 1971)
  23. Amplatz, C.: *Osservare la comunicazione educativa*. Pensa, Lecce (1999)
  24. Wright, H.F.: Observational child study. In: *Handbook of research methods in child development*, pp. 71–139 (1960)

25. Camaioni, L., Simion, F. (eds.): *Metodi di ricerca in psicologia dello sviluppo*. Il Mulino, Bologna (1990)
26. D'Isa, L., Foschini, F.: *Scienze umane. Psicologia e metodologia della ricerca per gli studi economici-sociali*. Hoepli, Milano (2011)
27. Schaffer, H.R.: *Il bambino e i suoi partner. Interazione e socialità* (1984). Trad. it. Franco Angeli, Milano (1990)
28. Schaffer, H.R.: *Lo sviluppo sociale* (1996). Trad. it. Cortina, Milano (1998)
29. Bertolini, P., Caronia, L.: *Dizionario di pedagogia e scienze dell'educazione*. Zanichelli (1996)
30. Ugazio, V.: Prefazione all'edizione italiana: oltre la teoria dell'attaccamento. In: Schaffer H.R. (ed.) *L'interazione madre-bambino: oltre la teoria dell'attaccamento* (1977). Trad. it. Franco Angeli, Milano, pp. 13–49 (1984)
31. Lavelli, M.: *Intersoggettività, origini e primi sviluppi*. Cortina, Milano (2007)
32. Camaioni, L., Bascetta, C., Aureli, T.: *L'osservazione del bambino nel contesto educativo*. Società Editrice il Mulino (1999)
33. Braga, P., Mauri, M., Tosi, P.: *Perché e come osservare nel contesto educativo: presentazione di alcuni strumenti*. Junior (1994)
34. Camaioni, L., Aureli, T., Peruccini, P.: *Osservare e valutare il comportamento infantile*. Il Mulino, Bologna (2004)
35. Bonichini, S., Giovanna Axia, G.: *L'osservazione dello sviluppo umano*. Carocci, Roma (2001)
36. Genovese, Bonichini, S., Giovanna Axia, G.: *L'osservazione dello sviluppo umano*, Carocci, Roma (1981, 2001)
37. Maldonato, M.: *Phenomenology of discovery: the cognition of complexity*. World Futures **67**(4–5), 372–379 (2011)
38. Frank, R., La Barre, F.: *The First Year and the Rest of Your Life. Movement, Development and Psychotherapeutic Change*. Routledge, New York, NY (2011). <https://doi.org/10.4324/9780203857472>
39. Haywood, K., Getchell, N.: *Life Span Motor Development*, 6th edn. Human Kinetics (2014)
40. Maggio, F., Il movimento in una prospettiva evolutiva: per una contestualizzazione dell'approccio di Ruella Frank. *Quaderni di Gestalt*. FrancoAngeli, Milano (2016)
41. Robine, J. M.: *Self: A Polyphony of Contemporary Gestalt Therapists*. St. Romaine de Virvée, L'Exprimerie, France (Original work published 2015 in French) (2016)
42. Frank, R.: *Self in motion*. In: Robine, J.-M. (ed.) *Self. A Polyphony of Contemporary Gestalt Therapists*. St. Romain-La-Virvée: L'Exprimerie (2015)
43. Maggio, F., Tosi, S.: *L'esperienza del movimento: la risonanza cinestetica come sentimento relazionale*. Intervista a Ruella Frank. *Quaderni di Gestalt*. FrancoAngeli, Milano (2016)
44. Dael, N., Mortillaro, M., Scherer, K.R.: Emotion expression in body action and posture. *Emotion* **12**(5), 1085 (2012). <https://doi.org/10.1037/a0025737>
45. Gea, J., Muñoz, M.A., Costa, I., Ciria, L.F., Miranda, J.G., Montoya, P.: Viewing pain and happy faces elicited similar changes in postural body sway. *PLoS ONE* **9**(8), e104381 (2014). <https://doi.org/10.1371/journal.pone.0104381>
46. Westcott, S.L., Lowes, L.P., Richardson, P.K.: Evaluation of postural stability in children: current theories and assessment tools. *Phys. Ther.* **77**(6), 629–645 (1997). <https://doi.org/10.1093/ptj/77.6.629>
47. Vallotton, C.D.: Do infants influence their quality of care? Infants' communicative gestures predict caregivers' responsiveness. *Infant Behav. Dev.* **32**(4), 351–365 (2009). <https://doi.org/10.1016/j.infbeh.2009.06.001>
48. Rönnqvist, L., Domellöf, E.: Quantitative assessment of right and left reaching movements in infants: a longitudinal study from 6 to 36 months. *Dev. Psychobiol.* **48**(6), 444–459 (2006). <https://doi.org/10.1002/dev.20160>
49. Neppi, T.K., Conger, R.D., Scaramella, L.V., Ontai, L.L.: Intergenerational continuity in parenting behavior: mediating pathways and child effects. *Dev. Psychol.* **45**(5), 1241 (2009). <https://doi.org/10.31274/etd-180810-3735>
50. Pratt, M., Apter-Levi, Y., Vakart, A., Kanat-Maymon, Y., Zagoory-Sharon, O., Feldman, R.: Mother-child adrenocortical synchrony; moderation by dyadic relational behavior. *Horm. Behav.* **89**, 167–175 (2017). <https://doi.org/10.1016/j.yhbeh.2017.01.003>

51. Rosenblatt, J.S.: The basis of synchrony in the behavioral interaction between the mother and her offspring in the laboratory rat. *Determin. Infant Behav.* **3**, 3–41 (1965)
52. Cook, S.W., Goldin-Meadow, S.: The role of gesture in learning: do children use their hands to change their minds? *J. Cogn. Dev.* **7**(2), 211–232 (2006). [https://doi.org/10.1207/s15327647jcd0702\\_4](https://doi.org/10.1207/s15327647jcd0702_4)
53. Cartmill, E.A., Beilock, S., Goldin-Meadow, S.: A word in the hand: action, gesture and mental representation in humans and non-human primates. *Phil. Trans. R. Soc. B* **367**(1585), 129–143 (2012). <https://doi.org/10.1098/rstb.2011.0162>
54. Feldman, R., Magori-Cohen, R., Galili, G., Singer, M., Louzoun, Y.: Mother and infant coordinate heart rhythms through episodes of interaction synchrony. *Infant Behav. Dev.* **34**(4), 569–577 (2011). <https://doi.org/10.1016/j.infbeh.2011.06.008>
55. Feldman, R.: Infant-mother and infant-father synchrony: the coregulation of positive arousal. *Infant Ment. Health J.* **24**(1), 1–23 (2003). <https://doi.org/10.1002/imhj.10041>
56. Sainath, T.N., Vinyals, O., Senior, A., Sak, H.: Convolutional, long short-term memory, fully connected deep neural networks. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4580–4584. IEEE (2015)
57. Perl, J.: A neural network approach to movement pattern analysis. *Hum. Mov. Sci.* **23**(5), 605–620 (2004). <https://doi.org/10.1016/j.humov.2004.10.010>
58. Robertson, N., Reid, I.: A general method for human activity recognition in video. *Comput. Vis. Image Underst.* **104**(2–3), 232–248 (2006). <https://doi.org/10.1016/j.cviu.2006.07.006>
59. Maldonato, M., Dell’Orco, S.: The natural logic of action. *World Futures* **69**(3), 174–183 (2013)
60. Almeida, A., Azkune, G.: Inter-activity behaviour modelling using long short-term memory networks. In: International Conference on Ubiquitous Computing and Ambient Intelligence, pp. 394–399. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-67585-5\\_41](https://doi.org/10.1007/978-3-319-67585-5_41)
61. Bressem, J., Ladewig, S.H.: Rethinking gesture phases: articulatory features of gestural movement? *Semiotica* **2011**(184), 53–91 (2011). <https://doi.org/10.1515/semi.2011.022>
62. Tsironi, E., Barros, P., Wermter, S.: Gesture Recognition with a Convolutional Long Short-Term Memory Recurrent Neural Network, p. 2. Bruges, Belgium (2016)
63. LeCun, Y., Kavukcuoglu, K., Farabet, C.: Convolutional networks and applications in vision. In: Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 253–256. IEEE (2010)
64. Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997). <https://doi.org/10.1162/neco.1997.9.8.1735>
65. Mundy, P., Jarrold, W.: Infant joint attention, neural networks and social cognition. *Neural Netw. Off. J. Int. Neural Netw. Soc.* **23**(8–9), 985–997 (2010). <https://doi.org/10.1016/j.neunet.2010.08.009>
66. Taylor-Colls, S., Pasco Fearon, R.M.: The effects of parental behavior on infants’ neural processing of emotion expressions. *Child Dev.* **86**(3), 877–888 (2015)
67. Yang, J., Lee, J., Choi, J.: Activity recognition based on RFID object usage for smart mobile devices. *J. Comput. Sci. Technol.* **26**(2), 239–246 (2011)
68. Lara, O.D., Labrador, M.A.: A survey on human activity recognition using wearable sensors. *IEEE Commun. Surv. Tutor.* **15**(3), 1192–1209 (2013). <https://doi.org/10.1109/surv.2012.110112.00192>
69. Leclère, C., Avril, M., Viaux-Savelon, S., Bodeau, N., Achard, C., Missonnier, S., Keren, M., Feldman, R., Chetouani, M., Cohen, D.: Interaction and behaviour imaging: a novel method to measure mother–infant interaction using video 3D reconstruction. *Transl. Psychiatry* **6**(5), e816– (2016). <http://doi.org/10.1038/tp.2016.82>
70. Lotzin, A., Romer, G., Schiborr, J., Noga, B., Schulte-Markwort, M., Ramsauer, B.: Gaze synchrony between mothers with mood disorders and their infants: maternal emotion dysregulation matters. *PLoS ONE* **10**(12), e0144417 (2015). <https://doi.org/10.1371/journal.pone.0144417>
71. Zimmerman, P.H., Bolhuis, J.E., Willemse, A., Meyer, E.S., Noldus, L.P.: The observer XT: a tool for the integration and synchronization of multimodal signals. *Behav. Res. Methods* **41**(3), 731–735 (2009). <https://doi.org/10.3758/brm.41.3.731>

72. Greff, K., Srivastava, R.K., Koutrník, J., Steunebrink, B.R., Schmidhuber, J.: LSTM: a search space odyssey. *IEEE Trans. Neural Netw. Learn. Syst.* **28**(10), 2222–2232 (2017). <https://doi.org/10.1109/tnnls.2016.2582924>
73. Tsironi, E., Barros, P., Weber, C., Wermter, S.: An analysis of convolutional long short-term memory recurrent neural networks for gesture recognition. *Neurocomput.* **268**, 76–86 (2017). <https://doi.org/10.1016/j.neucom.2016.12.088>
74. Maldonato, N.M., Sperandeo, R., Moretto, E., Dell'Orco, S.: A non-linear predictive model of borderline personality disorder based on multilayer perceptron. *Front. Psychol.* **9**, 447 (2018)

## Chapter 38

# Performance of Articulation Kinetic Distributions Vs MFCCs in Parkinson's Detection from Vowel Utterances



Andrés Gómez-Rodellar , Agustín Álvarez-Marquina , Jiri Mekyska , Daniel Palacios-Alonso , Djamila Meghraoui and Pedro Gómez-Vilda

**Abstract** Speech is a vehicular tool to detect neurological degeneration using certain accepted biomarkers derived from sustained vowels, diadochokinetic exercises, or running speech. Classically, mel-frequency cepstral coefficients (MFCCs) have been used in the organic and neurologic characterization of pathologic phonation using sustained vowels. In the present paper, a comparative study has been carried on comparing Parkinson's disease detection results using MFCCs and vowel articulation kinematic distributions derived from the first two formants. Binary classification results using support vector machines avail the superior performance of articulation kinematic distributions with respect to MFCCs regarding sensitivity, specificity, and accuracy. The fusion of both types of features could lead to improve general performance in PD detection and monitoring from speech.

---

A. Gómez-Rodellar · A. Álvarez-Marquina · D. Palacios-Alonso · P. Gómez-Vilda (✉)  
Neuromorphic Speech Processing Lab, Center for Biomedical Technology, Universidad  
Politécnica de Madrid, Campus de Montegancedo, 28223 Pozuelo de Alarcón, Madrid, Spain  
e-mail: [pedro@fi.upm.es](mailto:pedro@fi.upm.es)

A. Gómez-Rodellar  
e-mail: [grodellar@ctb.upm.es](mailto:grodellar@ctb.upm.es)

A. Álvarez-Marquina  
e-mail: [agustin@juniper.datsi.fi.upm.es](mailto:agustin@juniper.datsi.fi.upm.es)

D. Palacios-Alonso  
e-mail: [daniel.palacios@ctb.upm.es](mailto:daniel.palacios@ctb.upm.es)

J. Mekyska  
Department of Telecommunications, Brno University of Technology, Brno, Czech Republic  
e-mail: [mekyska@feec.vutbr.cz](mailto:mekyska@feec.vutbr.cz)

D. Palacios-Alonso  
Escuela Técnica Superior de Ingeniería Informática—Universidad Rey Juan Carlos,  
Campus de Móstoles, Tulipán, s/n, 28933 Móstoles, Madrid, Spain

D. Meghraoui  
Laboratory of Spoken Communication and Signal Processing, University of Science &  
Technology Houari Boumediene, Algiers, Algeria  
e-mail: [djamila.meghraoui@gmail.com](mailto:djamila.meghraoui@gmail.com)

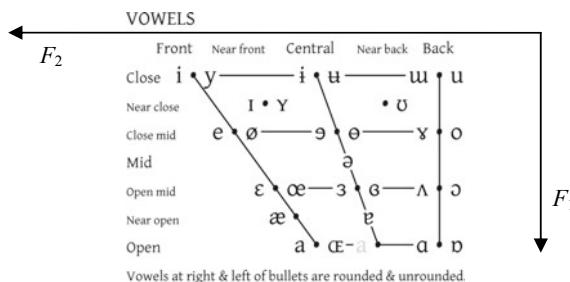
## 38.1 Introduction

Parkinson's disease (PD) is a neurodegenerative illness affecting neuromotor systems responsible for muscular control of most parts of the patient's body [1]. The muscles activating speech and phonation (chest, belly, larynx, pharynx, mouth, and face [2]) have shown to be highly affected by this disease [3]; therefore, speech is considered a vehicular tool to detect and grade neuromotor deterioration. Classically, articulation indices as the vowel space area (VSA) and formant centralization ratio (FCR) have been used as markers to detect deviations from a normative population to determine the presence of articulation pathological conditions by contrast tests [4], although these markers do not reflect speech dynamics properly. Other possible estimates with known representation power of speech dynamics are the mel-frequency cepstral coefficients (MFCCs). Their efficacy in speech recognition [5], speaker biometry [6], or voice pathology detection [7] is universally recognized. It seems natural to think about MFCCs as candidate features to detect and monitor speaker neurodegenerative conditions [8], although the way in which MFCCs encode speech articulation is not straight forward [9]. Usually, the first and second differences of MFCCs are added to MFCCs in an extended vector feature structure [5, 6], to represent the dynamic behavior of speech. In the present work, an alternative based on formant kinematics is used to encode articulation dynamics, having in mind that formant positions are dependent on specific articulation gestures. Therefore, a given relative 'speed' of articulators, defined on the first formant derivatives, could give a description of articulation dynamics [10]. The performance of this description has to be evaluated in comparison with other representations of speech dynamics, as extended MFCCs vectors. A model representing articulation dynamics by the first two formants is presented in Sect. 38.2. Section 38.3 gives a highly semantic articulation kinematic feature based on formant kinematics, as well as the databases and classifiers (based on support vector machines), used in detection experiments. Results are presented and discussed in Sect. 38.4. Conclusions are summarized in Sect. 38.5.

## 38.2 Kinematic Model of Speech Articulation

Speech articulation is determined by the movement of jaw, tongue, lips, and velopharyngeal tissues [11, 12]. As far as vowels are concerned, the main articulation gestures are the open–close, front–back, and round–oval qualities determining acoustic features perceived as formant positions. This mapping is usually represented on a vowel triangle (see Fig. 38.1).

The open–close gesture, mainly dominated by the jaw, is affecting mainly the first formant  $F_1$  (pulling up jaw is the dominant gesture in the phonation of [i] and [u], whereas relaxing down jaw is the gesture to phonate [a]). The front–back gesture is dominated by the tongue position, affecting mainly the second formant  $F_2$  (pushing the tongue forward is the gesture for [i]; pulling it back results in [u]). This is an

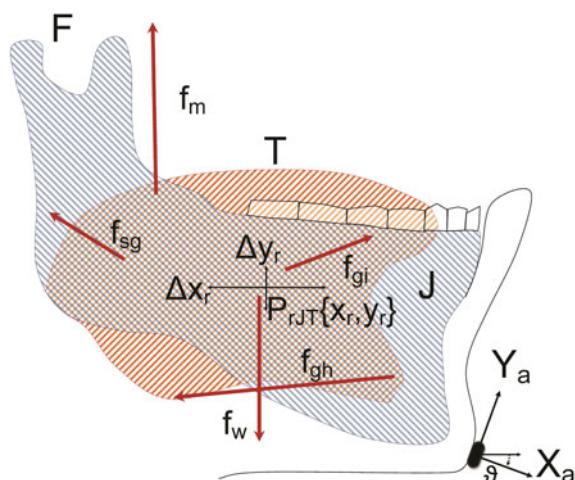


**Fig. 38.1** English vowel set (from IPA: <http://www.internationalphoneticassociation.org>). The vertical and horizontal axes represent the first ( $F_1$ ) and second ( $F_2$ ) formants, respectively (reversed). The feature *close/open* corresponds to the more precise term *high/low*

oversimplification of what is a more complicate relationship between articulation gestures and formant positions [11, 12], but it will be the starting point for defining kinematical correlates of phonation in the present work. In the present work, the articulation gesture of the jaw has been studied to relate articulatory gestures and acoustic features as the first two formants ( $F_1, F_2$ ). For such, an articulation kinematic model is proposed based on Fig. 38.2. The force exerted by the masseter  $f_m$  will pull up the low mandible acting as a third-order lever. Other acting muscles are the styloglossus, geniohyoid, and glosso-intrinsic, acting on the jaw–tongue with respect to the reference point.

Gravity acts as a constant acceleration downwards. The articulation gesture will determine the position of a hypothetical reference point in the jaw–tongue center of masses ( $P_{JT}$ ), moving on the jaw joint point  $F$ . The acoustic features  $\{F_1, F_2\}$  may be associated with the reference point coordinates  $\{x_r, y_r\}$  as in (38.1), assuming that the system is linear and time-invariant and that a one-to-one association between

**Fig. 38.2** Articulation kinematic model. F: fulcrum, T: tongue, J: jaw;  $f_m$ : masseter force,  $f_{sg}$ : styloglossus force,  $f_{gh}$ : geniohyoid force,  $f_{gi}$ : glosso-intrinsic forces,  $f_w$ : gravity;  $P_{JT}$ : jaw–tongue reference point,  $\{x_r, y_r\}$ : sagittal plane reference;  $\{X_a, Y_a\}$ : accelerometer reference



articulatory gestures and acoustic features is possible [11].

$$\begin{bmatrix} F_1(t) \\ F_2(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_r(t) \\ y_r(t) \end{bmatrix} \quad (38.1)$$

Under these assumptions, a relationship between the dynamic components of articulation and acoustics may be derived from (38.1)

$$\begin{bmatrix} \Delta F_1(t) \\ \Delta F_2(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \Delta x_r(t) \\ \Delta y_r(t) \end{bmatrix} \quad (38.2)$$

The utility of these relations is conditioned by the possibility of estimating the set of parameters  $a_{ij}$ . Facial accelerometry (fAcc), as indicated in Fig. 38.2 where a 3D accelerometer has been fixed to the chin of the subject under test, may be used in parameter estimation. The accelerometer reference axes are the chin-normal ( $X_a$ ) and tangential ( $Y_a$ ), which will be changing following jaw displacements. The accelerometry signals may be transformed to the reference coordinates  $\{x_r, y_r\}$  by means of a rotation in terms of  $\vartheta$ , the angle between the axes  $X_a$  and  $x_r$ .

### 38.3 Materials and Methods

#### 38.3.1 Validation of Articulatory Gesture to Acoustic Feature Mapping

To validate the model in (38.2), the set of parameters  $a_{ij}$  relating acoustic features and articulation dynamics has been estimated by regression-based methods. An example from a diadochokinetic exercise consisting in the repetition of the gliding sequence [...]aja...] from a male speaker is used to illustrate the methodology as follows:

- An accelerometer is fixed to the speaker's chin. Speech  $s(t)$  and fAcc  $\{a_{X_a}(t), a_{Y_a}(t)\}$  are recorded synchronously.
- The acceleration component means  $\{\bar{a}_{X_a}(t, W), \bar{a}_{Y_a}(t, W)\}$  are used to estimate the sensor angle  $\vartheta$  on short-time windows ( $W$ ) to preserve time invariance. The acceleration components are rotated to the reference axes to produce  $\{a_{xr}(t), a_{yr}(t)\}$

$$\vartheta = \arctan \left\{ \frac{\bar{a}_{X_a}(t, W)}{\bar{a}_{Y_a}(t, W)} \right\}; \quad \begin{bmatrix} a_{x_r}(t) \\ a_{y_r}(t) \end{bmatrix} = \begin{bmatrix} \cos \vartheta & \sin \vartheta \\ -\sin \vartheta & \cos \vartheta \end{bmatrix} \begin{bmatrix} a_{X_a}(t) \\ a_{Y_a}(t) \end{bmatrix} \quad (38.3)$$

- Speech  $s(t)$  is resampled to 8 kHz and 16 bits and inverse filtered [13] to obtain the eighth-order adaptive prediction vector  $\{b_i\}$  representing the inverse vocal tract.
- The first two formants  $F_1$  and  $F_2$  are estimated by a combined technique on the vocal tract transfer function

$$B(z) = 1 - \sum_{i=1}^k b_i z^{-i} \quad (38.4)$$

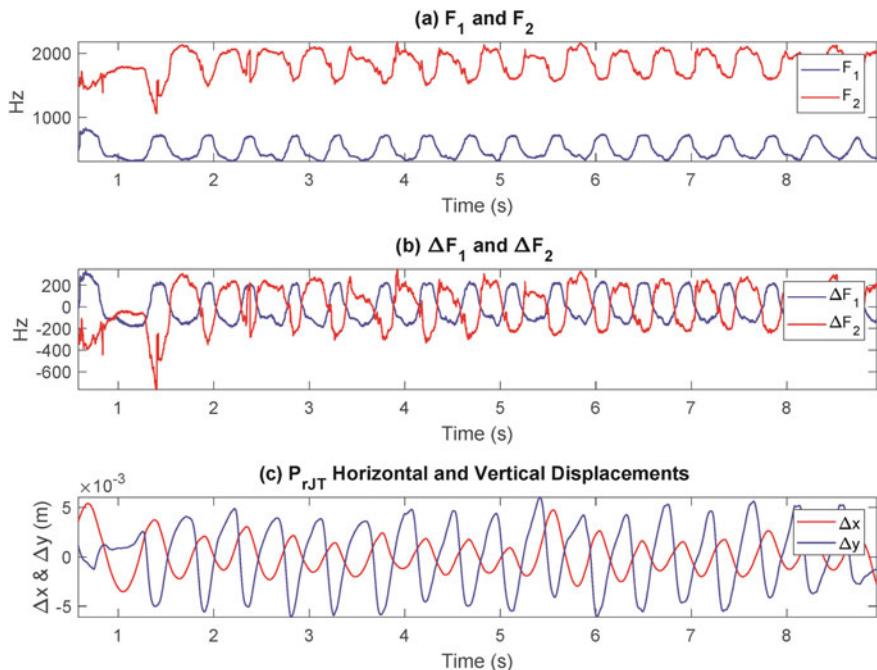
detecting the maxima of  $1/|B(z = e^{j\omega\tau})|$ , and the zeros of  $B(z)$  in the complex plane.

- The equivalent speed and displacement corrected to the reference point  $P_{rJT}$  may be estimated from the acceleration data as

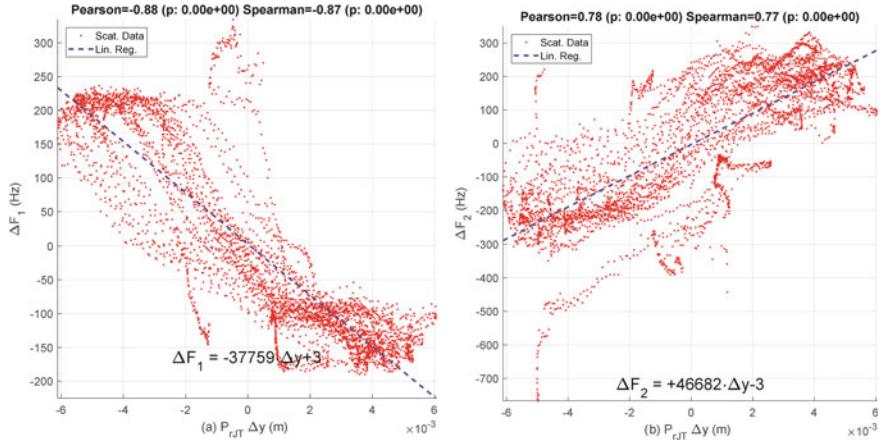
$$\begin{bmatrix} \Delta x_r(t) \\ \Delta y_r(t) \end{bmatrix} = \begin{bmatrix} \int_{t_1}^{t_2} v_{x_r}(t) dt \\ \int_{t_1}^{t_2} v_{y_r}(t) dt \end{bmatrix}; \quad \begin{bmatrix} v_{x_r}(t) \\ v_{y_r}(t) \end{bmatrix} = \begin{bmatrix} \int_{t_1}^{t_2} a_{x_r}(t) dt \\ \int_{t_1}^{t_2} a_{y_r}(t) dt \end{bmatrix} \quad (38.5)$$

As an example, speech and displacement data from the diadochokinetic exercise described above are given in Fig. 38.3.

Comparing the templates (b) and (c) in Fig. 38.3, it is clear that a relationship between  $\Delta F_1$  and  $\Delta F_2$  with respect to  $\Delta x_r$  and  $\Delta y_r$  may exist. This relationship has



**Fig. 38.3** Relationship between acoustical features ( $F_1$  and  $F_2$ ) and articulatory gestures (displacement of the reference point  $P_{rJT}$   $\{x_r, y_r\}$ ): **a**  $F_1$  and  $F_2$  absolute values. **b**  $F_1$  and  $F_2$  incremental values (smoothed and unbiased). **c** Displacements of  $P_{rJT}$   $\{\Delta x_r, \Delta y_r\}$



**Fig. 38.4** Regression results. Left: regression analysis of the dependence between the first formant and the vertical displacement with respect to the  $P_{rJT}$ . Right: dependence between the second formant and the vertical displacement

been explicitly estimated by linear regression, as shown in Fig. 38.4, focused only on  $\Delta y_r$ .

The regression results shown in Fig. 38.4 provide valuable estimates of the parameters  $a_{12}$  and  $a_{21}$  which may be formulated as

$$\begin{aligned}\Delta F_1(t) &\cong \tilde{a}_{12} \Delta y_r(t) + \tilde{a}_{120} \\ \Delta F_2(t) &\cong \tilde{a}_{22} \Delta y_r(t) + \tilde{a}_{220}\end{aligned}\quad (38.6)$$

where  $\tilde{a}_{12}$  and  $\tilde{a}_{22}$  may be identified with the regression line slopes; thus,  $\tilde{a}_{12} = -37,759 \text{ Hz/m} = -377.59 \text{ cm}^{-1} \text{ s}^{-1}$ , and  $\tilde{a}_{22} = 46,682 \text{ Hz/m} = 466.82 \text{ cm}^{-1} \text{ s}^{-1}$  where  $\tilde{a}_{120}$  and  $\tilde{a}_{220}$  can be neglected. Similar estimates of the remnant parameters in (38.2) may be produced ( $\tilde{a}_{11} = 585.69 \text{ cm}^{-1} \text{ s}^{-1}$ , and  $\tilde{a}_{21} = -802.94 \text{ cm}^{-1} \text{ s}^{-1}$ ). The results provided by the regression study avail the existence of a statistically relevant correlation between acoustic features and articulation gestures (see the very low  $p$  values in Fig. 38.4). This fact opens a possibility for using acoustic features to estimate articulation kinematic correlates as descriptors of articulation competence. Similar expressions as those in (38.2) may be found for  $\Delta F_1$  and  $\Delta F_2$  in terms of  $\Delta x_r$ . Having in mind the chain derivation rules

$$\begin{bmatrix} dx_r(t)/dt \\ dy_r(t)/dt \end{bmatrix} = \begin{bmatrix} \frac{\partial x_r}{\partial F_1} & \frac{\partial x_r}{\partial F_2} \\ \frac{\partial y_r}{\partial F_1} & \frac{\partial y_r}{\partial F_2} \end{bmatrix} \begin{bmatrix} \partial F_1(t)/\partial t \\ \partial F_2(t)/\partial t \end{bmatrix} \quad (38.7)$$

Defining the absolute kinematic velocity (AKV) for the reference point as

$$v_r(t) = \left[ \left( \frac{dx_r(t)}{dt} \right)^2 + \left( \frac{dy_r(t)}{dt} \right)^2 \right]^{1/2} \quad (38.8)$$

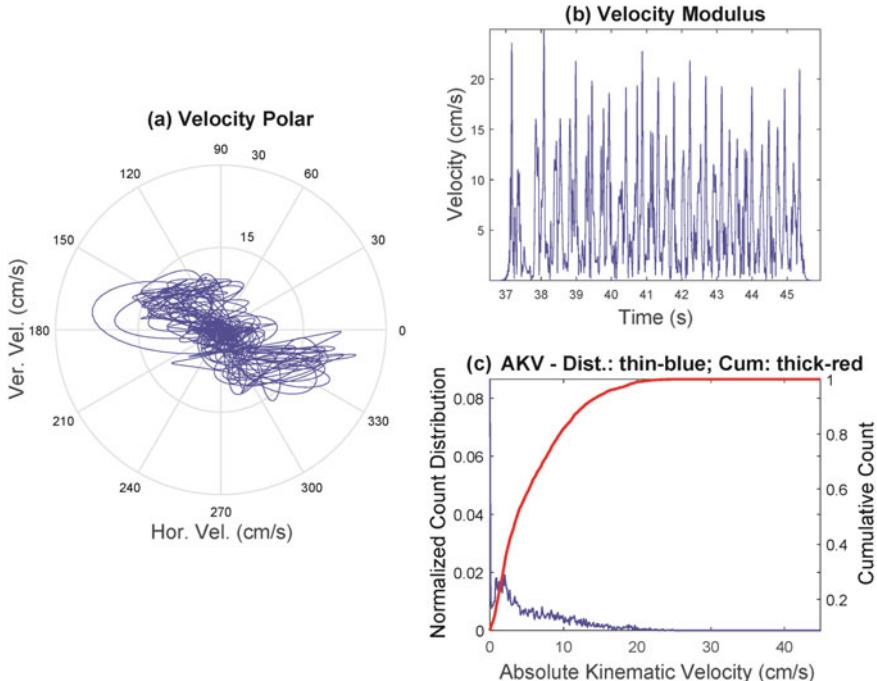
and taking (38.7) into account, this last expression may be rewritten as

$$v_r(t) = \left[ H_1 \left( \frac{dF_1(t)}{dt} \right)^2 + H_2 \left( \frac{dF_2(t)}{dt} \right)^2 + H_{12} \frac{dF_1(t)}{dt} \frac{dF_2(t)}{dt} \right]^{1/2} \quad (38.9)$$

where  $H_1$ ,  $H_2$ , and  $H_{12}$  are quadratic forms of  $(\tilde{a}_{ij})^{-1}$ . In this way, an estimation of the AKV may be produced exclusively in terms of formant dynamics. In the present study, expression (38.9) has been approximated as

$$\hat{v}_r(t) \cong \left[ \frac{\tilde{a}_{11}^2 + \tilde{a}_{12}^2}{\tilde{a}_{11}^2 \tilde{a}_{12}^2} \left( \frac{dF_1(t)}{dt} \right)^2 + \frac{\tilde{a}_{21}^2 + \tilde{a}_{22}^2}{\tilde{a}_{21}^2 \tilde{a}_{22}^2} \left( \frac{dF_2(t)}{dt} \right)^2 \right]^{1/2} \quad (38.10)$$

The AKV of the speech segment under study is given in Fig. 38.5. A probability



**Fig. 38.5** AKV from the diadochokinetic exercise under analysis: **a** trajectory of the kinetic velocity  $v_r$  in polar coordinates. **b** Absolute value of  $v_r$  (AKV). **c** Probability density function of the AKV (blue) and its cumulative function (red)

density distribution  $p(v_r)$  of the AKV may be produced from the normalized count histogram of  $v_r$  by amplitude levels as shown in Fig. 38.5c.

### 38.3.2 AKV-Based Parkinson Disease Detection

The use of AKV pdfs in the detection of irregular articulation from vowel utterances produced by PD patients with respect to healthy controls will be discussed. The Czech Parkinsonian Speech Database (PARCZ) [14] has been used in the experiments.

This database was recorded at St. Anne's University Hospital in Brno, the Czech Republic, and it contains samples of normative and pathological speech from healthy controls and PD patients including four sets of five Czech vowels ([a:, e:, i:, o:, u:]) pronounced in four different ways: short and long vowels uttered in a natural way, long vowels uttered with maximum loudness, and long vowels pronounced with minimum loudness, but not whispering. Recordings were sampled at 16 kHz and 16 bits. The samples selected corresponded to long [a:] vowels at maximum loudness by normative females and males, and PD females and males as described in Table 38.1.

The study compares the performance of AKV pdf-based features to classical MFCC features using the same classifier. Therefore, two sets of features have been generated:

- Feature Set A (FSA): Normalized distribution of the AKV associated with formants  $F_1$  and  $F_2$  as by (38.10). Distributions are represented by 500-bin histograms.  $F_1$  and  $F_2$  have been estimated on 5 ms windows with a window-sliding of 1 ms.
- Feature Set B (FSB): Mean and standard deviation of 20 MFCCs,  $c_0$  (zeroth-order cepstral coefficient) and  $\log E$  (logarithm of Energy), and their first and second differences,  $\Delta(\text{MFCC} + c_0 + \log E)$  and  $\Delta\Delta(\text{MFCC} + c_0 + \log E)$ . Each group is composed of 44 features; therefore, the complete vector is a 132-dimensional one.

Feature selection is applied to each feature set (FSA and FSB) using RReliefF [15, 16], with a number of neighbors varying between 1 and 50. As a result, 50 different arrangements in relevance for the total number of features considered (500 for FSA

**Table 38.1** Description of the subject sets included in the study from PARCZ

Set	No. of subjects	Mean age (y)	Std. dev. (y)
Normative females	26	61.81	9.05
Normative males	26	65.58	8.90
PD females	39	68.45	7.49
PD males	54	66.22	8.68

and 132 for FSB) are produced. Features are selected in subsets of N features (between 15 and 120) in order of relevance, from the highest to the lowest, accordingly to the classification provided by RReliefF. Each feature subset is the input to a support vector machine (SVM) with a Gaussian radial basis function (RBF) kernel [17] following the implementation given in [18]. In a preliminary work, random least squares feed-forward networks were used as a successful binary classifier [10]. In the present work, SVMs were used as an alternative to generate classification results using cross-validation of all the datasets distributed in ten groups (tenfold cross-validation). The process described is carried on for all the combinations of SVM parameters ( $C, \gamma$ ) given as  $C = [2^{-3}, 2^{-2}, \dots, 2^{12}]$  and  $\gamma = [2^{-1}, 2^{-2}, \dots, 2^{-10}]$ . The subset of N features producing the best results in terms of accuracy is selected, using the classical definition for sensitivity, specificity, and accuracy

$$\begin{aligned} \text{STV} &= \frac{\text{TP}}{\text{P}} = \frac{\text{TP}}{\text{TP} + \text{FN}} \\ \text{SPC} &= \frac{\text{TN}}{\text{N}} = \frac{\text{TN}}{\text{TN} + \text{FP}} \\ \text{ACC} &= \frac{(\text{TP} + \text{TN})}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \end{aligned} \quad (38.11)$$

with TP: true positives, TN: true negatives, FP: false positives, and FN: false negatives. The performance indices STV, SPC, and ACC for each winner subset of N features corresponding to FSA or FSB are estimated on the average of 1000 different runs over the ten groups (1000 runs of the tenfold cross-validation). Wilcoxon rank sum test (Mann–Whitney U) is used to estimate the statistical relevance of the results, from FSA and FSB, produced independently for the male and female databases.

## 38.4 Results and Discussion

Table 38.2 gives comparative results after cross-validation (1000 runs, tenfold) in terms of STV, SPC, and ACC for male and female datasets under a maximum  $p$  value  $<10^{-4}$ . The results show that AKV distributions behave better than MFCCs in all cases for both male and female datasets. The performance is better for the male than for the female dataset. This may be due to the larger number of samples in the male dataset. It is expected that these results will be extended and completed with future versions of PARCZ. A similar study is being conducted on the other vowels available ([e:, i:, o:, u:]), considering also the study of long vowels and short vowels at different utterance loudness levels.

A very relevant aspect to be commented is the use of a wide-range diadochokinetic exercise to model rapid jaw–tongue movement, when only sustained vowels were used in PD neuromotor degradation detection experiments. Wide-range diadochokinetic exercises of the kind shown in Sect. 3.1 are good models of tongue–jaw

**Table 38.2** Merit figures for the different subsets in %

Set	Features	Female STV	Male STV
FSA	Absolute kinematic Velocity	97.30	97.88
FSB	MFCCs	96.88	97.09
Set	Features	Female SPC	Male SPC
FSA	Absolute kinematic velocity	98.69	99.56
FSB	MFCCs	88.18	90.81
Set	Features	Female ACC	Male ACC
FSA	Absolute kinematic velocity	97.86	98.43
FSB	MFCCs	94.06	95.05

unstability and tremor, usually found in PD [3], considering these anomalies as range kinematic distortions of neuromotor control.

## 38.5 Conclusions

The main conclusions derived from the experimental work presented can be summarized as follows:

- A biomechanical model relating acoustic features and articulation gestures has been proposed. It could be extended to a neuromechanical one showing the relation between sEMG and articulation as an open research line.
- A method to estimate the parameters of the biomechanical model by linear regression on  $F_1$  and  $F_2$  and jaw–tongue displacements has been proposed, and the kinematics of  $F_1$  and  $F_2$  in terms of articulation has been exposed.
- The AKV distribution has been defined as an articulation feature based on formant kinematic estimates, opening the possibility of using speech-derived features to characterize anomalous articulation behavior.
- A PD detection study from phonation comparing the results of AKV distributions and MFCCs has been conducted on a database of male and female speakers. The performance of AKV distributions showed a larger detection accuracy than MFCCs on a SVM classifier.

**Acknowledgements** This work was supported by grant TEC2016-77791-C4-4-R (Plan Nacional de I+D+i, Ministry of Economic Affairs and Competitiveness of Spain), CENIE\_TECA-PARK\_55\_02 INTERREG V-A Spain-Portugal (POCTEP), and grant 16-30805A of the Czech Ministry of Health.

## References

1. Morris, M.E.: Movement disorders in people with Parkinson disease: a model for physical therapy. *Phys. Ther.* **80**(6), 578–597 (2000)
2. Sapir, S.: Multiple factors are involved in the dysarthria associated with Parkinson's disease: a review with implications for clinical practice and research. *J. Speech Lang. Hear. Res.* **57**, 1330–1343 (2014)
3. Skodda, S., Grönheit, W., Mancinelli, N., Schlegel, U.: Progression of voice and speech impairment in the course of Parkinson's disease: a longitudinal study. *Parkinson's Dis. Article ID 389195* (2013)
4. Sapir, S., Ramig, L.O., Spielman, J.L., Fox, C.: Formant centralization ratio: a proposal for a new acoustic measure of dysarthric speech. *J. Speech Lang. Hear. Res.* **53**(1), 114–125 (2010)
5. Hinton, G., Deng, L., Yu, D., Dahl, G.E., Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, V., Sainath, T.N., Kingsbury, B.: Deep neural networks for acoustic modeling in speech recognition. *IEEE Sig. Proc. Mag.* **29**, 82–97 (2012)
6. Kinnunen, T., Li, H.: An overview of text-independent speaker recognition: from features to supervectors. *Speech Comm.* **52**, 12–40 (2010)
7. Fraile, R., Sáez, N., Godino, J.I., Osma, V., Fredouille, C.: Automatic detection of laryngeal pathologies in records of sustained vowels by means of mel-frequency cepstral coefficient parameters and differentiation of patients by sex. *Folia Phoniatr. Logop.* **61**, 146–152 (2009)
8. Benba, A., Jilbab, A., Hammouch, A.: Detecting patients with Parkinson's disease using mel frequency cepstral coefficients and support vector machines. *Int. J. Electr. Eng. Inf.* **7**(2), 297–306 (2015)
9. Murphy, P.J., Akande, O.O.: Noise estimation in voice signals using short-term cepstral analysis. *J. Acoust. Soc. Am.* **121**(3), 1679–1690 (2007)
10. Gómez, P., et al.: Parkinson disease detection form speech articulation neuromechanics. *Front. Neuroinform.* **11**, 1–17 (2017)
11. Dromey, C., Jang, G.O., Hollis, K.: Assessing correlations between lingual movements and formants. *Speech Comm.* **55**, 315–328 (2013)
12. Whitfield, J.A., Goberman, A.M.: Articulatory-acoustic vowel space: application to clear speech in individuals with Parkinson's disease. *J. Comm. Disord.* **51**, 19–28 (2014)
13. Deller, J.R., Proakis, J.G., Hansen, J.H.L.: *Discrete-Time Processing of Speech Signals*. Macmillan, New York (1993)
14. Mekyska, J., Janusova, E., Gómez, P., Smekal, Z., Rektorova, I., Eliasova, I., Kostalova, M., Mrackova, M., Alonso, J.B., Faúndez, M., López de Ipiña, K.: Robust and complex approach of pathological speech signal analysis. *Neurocomput.* **167**, 94–111 (2015)
15. Kononenko, I., Šimec, E., Robnik-Šikonja, M.: Overcoming the myopia of inductive learning algorithms with RReliefF. *Applied Intelligence* **7**, 39–55 (1997)
16. Robnik-Šikonja, M., Kononenko, I.: Theoretical and empirical analysis of ReliefF and RReliefF. *Mach. Learn.* **53**(1–2), 23–69 (2003)
17. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995)
18. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**(3), 27:1–27:27 (2011)

## Chapter 39

# From “Mind and Body” to “Mind in Body”: A Research Approach for a Description of Personality as a Functional Unit of Thoughts, Behaviours and Affective States



Daniela Iennaco, Raffaele Sperandeo , Lucia Luciana Mosca, Martina Messina, Enrico Moretto , Valeria Cioffi, Silvia Dell’Orco and Mauro N. Maldonato

**Abstract** The study of personality has been developed according to different lines of research in which it is also possible to identify transversal elements that are useful for an unprecedented trans-theoretical research programme. The existence of a close relation between the contents of thought (narrative system) and the vegetative and motor functions (experiential system) of the human being is currently considered of great relevance. However, until now, the studies using classic psychological test in order to describe the contents of the narrative system customize the experiential system and produce partial descriptions and non-integrated description of the per-

---

D. Iennaco (✉) · R. Sperandeo · L. L. Mosca · M. Messina · E. Moretto · V. Cioffi · S. Dell’Orco  
SiPGI—Postgraduate School of Integrated Gestalt Psychotherapy, Torre Annunziata,  
Naples, Italy  
e-mail: [daniennaco@gmail.com](mailto:daniennaco@gmail.com)

R. Sperandeo  
e-mail: [raffaele.sperandeo@gmail.com](mailto:raffaele.sperandeo@gmail.com)

L. L. Mosca  
e-mail: [moscaluciana@libero.it](mailto:moscaluciana@libero.it)

M. Messina  
e-mail: [m.messina18@campus.unimib.it](mailto:m.messina18@campus.unimib.it)

E. Moretto  
e-mail: [enrico.more@gmail.com](mailto:enrico.more@gmail.com)

V. Cioffi  
e-mail: [dr.valeria.cioffi@gmail.com](mailto:dr.valeria.cioffi@gmail.com)

S. Dell’Orco  
e-mail: [silviadellorco@gmail.com](mailto:silviadellorco@gmail.com)

M. N. Maldonato  
Department of Neuroscience and Reproductive Sciences and Odontostomatology,  
University of Naples Federico II, Naples, Italy  
e-mail: [nelsonmauro.maldonato@unina.it](mailto:nelsonmauro.maldonato@unina.it)

sonality. Nevertheless, when the people reflect about the questions in a personality test, they present (under the action of this cognitive stimulus) a synchronous activity of the narrative and experiential system. The intention of the study proposed in this paper, with the use of biofeedback and neurofeedback methods, is to record the physiological responses of the subject while answering to the items of a personality test, in order to describe it as a functional unit of thoughts, behaviours and affective states. The personality assessment tool, Cloninger's temperament and character inventory, will be provided with a computerized system: the electrophysiological recording will be produced using the neurofeedback system "Neurobit Optima 4"; for the analysis of electrophysiological measurements, software ("BioExplorer" system) will be used to manage the brain–computer interface (BCI). It is expected that the aforementioned methodology will allow for an original description of the personality as a functional unit of thoughts, behaviours and affective states.

## 39.1 Introduction

In the field of personality studies, developed over the last twenty years according to numerous theoretical orientations, it is possible to trace transversal elements shared by the various models and to outline the conditions for a trans-theoretical research programme. Nowadays, it seems possible to structure this ambitious research programme based, on the one hand, on the many consolidated data that the various approaches to personality have produced over the years and on the other on the mathematical methods of deep learning able to produce nonlinear models to describe complex systems. Without claiming to discuss all the complex issues that arise in the creation of a personality theory that intends to integrate the different approaches, this paper will explore the nature of the relationship between thought and movement [1, 2]. The limits of the current reflection on this central element for all personality theories will be highlighted, and finally, a research method will be proposed to deal with the question from a quantitative point of view.

## 39.2 Theoretical Assumptions

An assumption substantially shared by personality researchers is the presence of two adaptive systems operating in the information processing: the experiential system, which is based on the interaction of the motor and sensory and the individual and the environment, and the narrative system, based on the use of words as a tool to organize the episodic memory [3–5]. According to the most common vision, both systems appear to have substantial differences. The experiential system is based on the overall interaction with the environment through perceptions, conscious or unconscious actions and processes of affective self-regulation [6], while the narrative system takes the form of a set of conscious verbal or imaginal mental representations and explicit

memories [7]. However, in a deeper reflection different elements emerge showing there is a close relation between both systems. The experiential system consists of a learning process based on the motor, and emotional and sensorial experience [8]. The process of learning is accomplished through the action and experimentation of situations, tasks and roles in which the subject, as an active protagonist, finds his own resources and skills for the elaboration and/or reorganization of behaviours aimed at achievement of a goal [9]. Experiential learning allows the subject to face situations of uncertainty by developing adaptive behaviours and, at the same time, to improve the ability of managing one's own emotions [10]. It also allows for the development of problem-solving skills, even through executive functions that, in a creative way, redefine possible inappropriate attitudes and enhance constructive behaviour [11]. The experience thus acquired becomes a heritage of knowledge for the subject and constitutes the new starting point for further developments [12]. This is a holistic view of learning: cognitive, emotional, volitional aspects as well as social aspects are integrated, and the person is involved in its entirety [13, 14]. Learning is achieved through direct participation and first-person discoveries. This inherent awareness leads to personal development, which stimulates an emergent and spontaneous behavioural or cognitive change [15]. Narration is a privileged tool of cultural transmission that organizes experiences, and constructs and transmits meanings [16]. Jerome Bruner defines it as a peculiar cognitive way of thought, which has always been used by the human being as a kind of “narrative creation of the self”, an essential dimension for the construction of subjective identity, and as a way of constant comparison with the other [17]. Everyone, according to Bruner, has the awareness of their learning mechanisms, of how their reasoning proceeds in order to acquire the ability to organize one's adult life [18]. Personal meaning (and personal reality) is constructed during the conceptualization and exposition of one's own narrative. Our experiences take the form of the narratives that we use to describe them, and stories are our way of organizing, interpreting and giving meaning to the experiences, ensuring them a sense of continuity [19, 20]. It is the narrative mode of thinking that allows us to reflect on experience [21]. However, when a subject organizes their experience according to a narrative modality, the self-regulating and perceptive motor processes are reactivated even if they are not visible to the observation but have a reduced intensity. Moreover, from the side of the experiential system, perception and action are frequently accompanied by verbal contents that structure and organize, during the action itself, the contents of thought [22]. Therefore, what spontaneously emerges from the conscience is something already rooted in the body and in the sensorial experience and this is the reason for which it is considered an incarnated event [23]. The two systems, therefore, appear as two different manifestations of the same process, and the fundamental distinction between them lies in the intensity of muscular involvement of the self-regulation process and in the amount of energy consumed [24]. The experiential system is characterized by an intense and visible muscular interaction with the environment, which requires a large consumption of energy and is aimed at manipulating the material elements of reality [25]; the narrative system is characterized by an imaginative interaction with the environment that does not require an evident muscular commitment and for which a smaller amount of phys-

ical energy is needed, and it allows for a sort of simulation of the manipulation of reality [26, 27]. For example, an individual who verbally and physically expresses his aggression towards another activates the experiential system that correlates with a clear muscular activity in addition to the activation of physiological and neural processes directed to the regulation of emotions. The same individual, when thinking about the anger he feels towards someone, will activate the narrative system, which presents the same muscular and neurophysiological correlation, but expressed with less intensity and less energy expenditure. These two systems, when observed in a superficial way, will appear first as behaviour and secondly as a mental act, but the biological process in both is identical and differs only in the amount of energy engaged in each process [28]. The apparent distinction between these two processes (behavioural and mental) has for years confused scholars and determined the development of different systems in order to measure psychic phenomena: some were oriented towards behaviour, while others towards thought [29–31]. In the study of personality, psychometric tools are prevalent to measure the narrative system. Over the years, in this way, we have reached measurements that, instead of describing personality as a functional unicum (constant cognitive, emotional, motivational and behavioural patterns) [32], present only the contents of the narrative system that hide rather than unveiling the processes of the experiential system [33]. The result is that the ordinary psychometric tools on the one hand are very partial descriptions of the personality that fail to identify the latter as a functional unit and on the other hand they bring together processes that are only apparently similar, since the narrative contents do not concern entirety of the experience [34]. When answering questions of a test aimed at describing one's ability to entertain intimate relationships, a subject's response will represent the set of their convictions that often underlie experiential processes of which they are not aware [35]. Similarly, it is unclear whether a person who says they experience joy when meeting a family member is really living that emotional experience and not one in the opposite direction of which they are absolutely unaware. In addition, two individuals who answer the questions of a test in the same way can express through those answers profoundly different life experiences [36]. These dynamics, substantial for the description of personality, are hidden rather than revealed by the current psychometric tools [37].

### 39.3 Aims of the Study

This research project aims to fill the gap, currently present in personality studies, between the analysis of the experiential system and the narrative one. Given the mental effort required in the process of answering the items of a personality test, the function of the experiential system will activate, but with a low energy commitment [38]; as such, we aim to measure the neurophysiological parameters of subjects involved in the compilation of personality assessment questionnaires.

This type of analysis will allow us to detect the pattern of neurophysiological activity connected to certain narrative stimuli.

- Measuring the consistency between the narrative responses to the test and the physiological activity connected to it;
- Recognizing how similar narrative responses can underlie very different experiences.

The evaluation of psychic functions, freed from the ancient body–mind dualism, will allow us to construct diagnostic tools capable of describing personality as a functional unit [39, 40].

## 39.4 Methodology

The sample group is made up of 100 persons from a private psychotherapy service. They will be subjected to the diagnostic protocol of the institute, used to exclude psychiatric and neurological disorders that have an impact on the healthy functioning of personality behaviour, with the subsequent compromise of the experiential and narrative systems. The assessment will be carried out by professionals who are experts in diagnosis. The diagnosis foresees a historical analysis, an accurate psychological examination, and will be given a battery of tests to assess the psychopathology. Specifically speaking, they will be given the SCID-5-CV and the SCID-5-PD [29].

The Structured Clinical Interview is a semi-structured interview to be used as a guide to formulating the main diagnoses of the DSM-5.

The 16PF-5 [41] test will be used to assess the functioning of the personality. The 16PF-5 is one of the world's best-known personality assessment tests, based on the measurement of 16 dimensions, functionally independent and psychologically significant, isolated during a thirty-year period through factorial studies conducted on normal and clinical groups.

The evaluation of the executive functions will be carried out through the use of the FAB. The FAB [42] consists of six subtests exploring different functions related to the frontal lobes and correlated with frontal metabolism.

The questionnaire 16PF-5 will be administered through a computerized system in which each participant will have to read the item on a tablet and select the answer using the touchscreen, synchronizing the response to each item with the following psycho- and neurophysiological findings:

- Skin conductance
- Heart rate
- Respiratory rate
- Oxygen saturation in the blood
- State of contraction of the muscular system of the main skeletal muscles
- The electrical activity of the various brain regions.

In an initial phase, participants will be evaluated in their baseline responses through the recording of the same parameters during a rest period of 5 min and during the answer to neutral questions on their socio-demographic characteristics.

The method used for electrophysiological and electroencephalographic recordings will require the positioning of the electrodes according to the “10–20” system which is a standardized method for the comparison of electrical activity of the cerebral areas. The electrodes (positioned on the scalp) can register those cortical activities of the brain regions that are close to each other. The term “10–20” refers to the placement of electrodes over 10% or 20% of electrodes of the total distance between the specific positions of the skull. Studies have shown that these placements are correlated with the corresponding cortical regions of the brain [43–45].

For the analysis of electrophysiological measurements, software for managing the brain–computer interface (BCI) will be used.

The research protocol complies with the standards of the AIP ethical code and has been approved by the ethics committee of the FISIG—Italian Federation of Schools and Institutes of Gestalt.

## 39.5 Statistical Analysis

The averages of the electrophysiological parameters recorded during the responses to the individual items will be correlated with the answers themselves, using the Spearman correlation test to evaluate the quality of the electrophysiological activity related to a certain type of response (positive or negative) related to a given item [18]. Moreover, through the hierarchical analysis of the clusters, the different electrophysiological activity patterns related to a specific item will be grouped.

Finally, using the multi-layered perceptron method [46], the most relevant electrophysiological items and patterns will be determined for the description of the specific character traits [47].

## 39.6 Tools

### 39.6.1 Sixteen Personality Factor Questionnaire-Fifth Edition (16PF-5)

The Sixteen Personality Factor Questionnaire-5 (16PF-5) is a self-report personality test, which provides a measure of the normal personality and that also may be used by psychologist and other mental health professionals, as a clinical instrument to help diagnose the psychiatric disorders. The questionnaire consists of 185 items with three possible answers, and it measures 16 primary factors, 5 global factors and 3 response style indices. The 16PF-Fifth Edition contains 185 multiple-choice items which are written at a fifth-grade reading level; the items ask simple questions about daily behaviour, interests and opinions. The 16PF provides scores regarding 16 primary personality scales and five global personality

scales, all of which are bipolar (both ends of each scale have a distinct, meaningful definition): A: detached-hot, B: crystallized intelligence, 4; D: mild-dominant, F: taciturn-enthusiast, G: unregulated-conscience, H: shy-uninhibited, I: hard-sensitive, L: confident-suspicious, M: practical-imaginative, N: naive-insecure, O: self-confident-fake senses, Q1: conservative-radical, Q2: dependent-self-sufficient, Q3: indisciplined-enterprising, Q4: quiet-tense. The 16 primary factors reveal a structure of second order with 5 factors: (1) extroversion, (2) anxiety, (3) hardness, (4) independence and (5) self-control. The instrument also includes three validity scales, which provides the measure of three indices about the subject’s response style: (1) bipolar impression management (IM) scale—scale of social desirability, (2) acquiescence (ACQ) scale—a measure of tendency to respond “true” to an item without taking into account its content and (3) infrequency (INF) scale—it makes possible to identify whether a subject responds much differently from the most of people. The test is accompanied by an online administration and scoring system that provides scores on the sixteen primary factors, the five global factors and the three response style indices, both according to the specific rules for sex and those combined.

### **39.6.2 *Electrophysiological Recording (EECG, sEMG, HRV, GSR, TEMP)***

*Tool.* For the measurement of electrophysiological activities, a “Neurobit Optima 4” system will be used, a neurofeedback system with 4 low-frequency universal channels that allow the measurement of the voltage signals, conductance, resistance and temperature (EECG, sEMG, HRV, GSR, TEMP). The instrument is able to perform an electrode impedance and circuit continuity test, has independent reference inputs for each channel, has a resolution of 16-bit measurements and is extremely accurate in voltage measurement: 1%, up to 2000 sps (scan per second), oversampling (sampling primary speed up to 8000 sps), has high resistance to electrical interference, active measurement cable shielding to reduce motion artefacts, configurable power supply interference filter (50/60 Hz or deactivated), complete galvanic isolation of the user’s body connected to the appliance [3].

*Software.* The software component is entrusted to the “BioExplorer” system. The software, used to process and analyse physiological signals and for the audio-visual presentation of the feedback, makes it possible to visualize the network for data processing using universal blocks; the construction of biofeedback protocols is simple, using a modular object-based design environment. BioExplorer is compatible with Microsoft Windows 7, 8 and 10 systems [48].

### 39.7 Expected Results and Conclusions

It is foreseeable that only some items of the 16PF-5 will be able to induce relevant electrophysiological variations and can be interpreted as specific descriptors of certain personality traits. Furthermore, it will be possible to detect different electrophysiological responses in different subjects responding to the same item and it is likely that this type of response may depend on the person's character style. Therefore, it will be possible to characterize in a more selective and specific way the set of character dimensions of a given subject.

In this first phase of the research project, the objective will be to optimize the psychometric instrument integrating it with adequate electrophysiological recordings for the description of the personality understood as a functional unit of thoughts, behaviours and affective states.

Ultimately, we intend for this research approach, through the integration of the experiential system and the narrative system into a functional unit, to become the indispensable prerequisite for the creation of a trans-theoretical personality model, which does not start from naïve Cartesian epistemology in which body and mind are considered as separate entities, but instead recognize that they belong to the same matter that expresses itself on different levels of energetic activity. The averages of the electrophysiological parameters recorded during the responses to the individual items will be correlated with the answers themselves, using the Spearman correlation test to evaluate the quality of the electrophysiological activity.

## References

1. Maldonato, N.M., Sperandeo, R., Dell'Orco, S., Cozzolino, P., Fusco, M.L., Iorio, V.S., Cipresso, P.: The relationship between personality and neurocognition among the American elderly: an epidemiologic study. *Clin. Pract. Epidemiol. Ment. Health: CP & EMH* **13**, 233 (2017)
2. Noë, A.: Perché non siamo il nostro cervello: una teoria radicale della coscienza. Raffaello Cortina (2010)
3. Inkster, M., Wellsby, M., Lloyd, E., Pexman, P.M.: Development of embodied word meanings: sensorimotor effects in children's lexical processing. *Front. Psychol.* **7**, 317 (2016)
4. Barca, L., Mazzuca, C., Borghi, A.M.: Pacifier overuse and conceptual relations of abstract and emotional concepts. *Front. Psychol.* **8** (2017)
5. Vallet, G.T., Hudon, C., Bier, N., Macoir, J., Versace, R., Simard, M.: A SEMantic and EPisodic Memory Test (SEMEP) developed within the embodied cognition framework: application to normal aging, Alzheimer's disease and semantic dementia. *Front. Psychol.* **8**, 1493 (2017)
6. Epstein, S.: Cognitive-Experiential Theory: An Integrative Theory of Personality. Oxford University Press (2014)
7. Di Nubila, R.D., Fedeli, M.: L'esperienza: quando diventa fattore di formazione e di sviluppo: dall'opera di David A. Kolb alle attuali metodologie di Experiential Learning Testimonianze e case study. Pensa multimedia (2010)
8. Johnson-Glenberg, M.C., Megowan-Romanowicz, C., Birchfield, D.A., Savio-Ramos, C.: Effects of embodied learning and digital platform on the retention of physics content: centripetal force. *Front. Psychol.* **7**, 1819 (2016)

9. Lozada, M., Carro, N.: Embodied action improves cognition in children: evidence from a study based on Piagetian conservation tasks. *Front. Psychol.* **7**, 393 (2016)
10. Kuo, C.Y., Yeh, Y.Y.: Sensorimotor-conceptual integration in free walking enhances divergent thinking for young and older adults. *Front. Psychol.* **7**, 1580 (2016)
11. Sperandeo, R., Maldonato, M., Baldo, G., Dell'Orco, S.: Executive functions, temperament and character traits: a quantitative analysis of the relationship between personality and pre-frontal functions. In: 2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), pp. 000043–000048. IEEE (2016)
12. Kandel, E.R., Schwartz, J.H., Jessell, T.M.: *Principi di neuroscienze* (1994)
13. Cantone, D., Sperandeo, R., Maldonato, M.: A dimensional approach to personality disorders in a sample of juvenile offenders. *Rev. Latinoam. Psicopatol. Fundam.* **15**(1), 42–57 (2012)
14. Sperandeo, R., Picciocchi, E., Valenzano, A., Cibelli, G., Ruberto, V., Moretto, E., et al.: Exploring the relationships between executive functions and personality dimensions in the light of "embodied cognition" theory: a study on a sample of 130 subjects. *Acta Med. Mediterr.* **34**(5), 1271–1279 (2018)
15. Rogers, C.R.: *Liberità nell'apprendimento* (1969), tr. it. a cura di R. Tettucci, Giunti-Barbera, Firenze (1973)
16. Chen, X., Liu, B., Lin, S.: Is accessing of words affected by affective valence only? A discrete emotion view on the emotional congruency effect. *Front. Psychol.* **7**, 916 (2016)
17. Bruner, J.S., Rini, R.: *La mente a più dimensioni*. Laterza (1988)
18. Setti, A., Borghi, A.M.: Embodied cognition over the lifespan: theoretical issues and implications for applied settings. *Front. Psychol.* **9**, 550 (2018)
19. Dewey, J.: *Esperienza e natura*, trad. it. Milano: Mursia (1990)
20. Adams, A.M.: How language is embodied in bilinguals and children with specific language impairment. *Front. Psychol.* **7**, 1209 (2016)
21. Jeannerod, M.: *Motor Cognition: What Actions Tell the Self*, vol. 42. Oxford University Press (2006)
22. Borghi, A.M., Setti, A.: Abstract concepts and aging: an embodied and grounded perspective. *Front. Psychol.* **8**, 430 (2017)
23. Thill, S., Twomey, K.E.: What's on the inside counts: a grounded account of concept acquisition and development. *Front. Psychol.* **7**, 402 (2016)
24. Pouw, W.T., Van Gog, T., Zwaan, R.A., Paas, F.: Augmenting instructional animations with a body analogy to help children learn about physical systems. *Front. Psychol.* **7**, 860 (2016)
25. Repetto, C., Serino, S., Macedonia, M., Riva, G.: Virtual reality as an embodied tool to enhance episodic memory in elderly. *Front. Psychol.* **7**, 1839 (2016)
26. Maldonato, M., Montuori, A., Dell'orco, S.: *The exploring Mind. Natural Logic and Intelligence of the Unconscious*. Mauro Maldonato (2013)
27. Hainselin, M., Picard, L., Manolli, P., Vankerkore-Candas, S., Bourdin, B.: Hey teacher, don't leave them kids alone: action is better for memory than reading. *Front. Psychol.* **8**, 325 (2017)
28. Maldonato, M., Sperandeo, R., Dell'Orco, S., Iennaco, D., Cerroni, F., Romano, P., Tripi, G.: Mind, brain and altered states of consciousness. *Acta Med. Mediterr.* **34**(2), 357–366 (2018)
29. American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Publishing (2013)
30. Sperandeo, R., Monda, V., Messina, G., Carotenuto, M., Maldonato, N.M., Moretto, E., Messina, A.: Brain functional integration: an epidemiologic study on stress-producing dissociative phenomena. *Neuropsychiatric Dis. Treat.* **14**, 11 (2018)
31. Sloman, S., Fernback, P.: *L'illusione della conoscenza*. Raffaello Cortina (2018)
32. Rizzolatti, G., Kalaska, J.E.: Il movimento volontario. In: Kandel et al. (eds.) *La corteccia parietale e la corteccia premotoria*, pp. 864–893 (2015)
33. Maldonato, N.M., Sperandeo, R., Dell'Orco, S., Cozzolino, P., Fusco, M.L., Iorio, V.S., Cipresso, P.: The relationship between personality and neurocognition among the american elderly: an epidemiologic study. *Clin. Pract. Epidemiol. Ment. Health CP & EMH* **13**, 233 (2017)

34. Dempster, T.: An investigation into the optimum training paradigm for alpha electroencephalographic biofeedback. Doctoral dissertation, Canterbury Christ Church University (2012)
35. Capellini, R., Sacchi, S., Ricciardelli, P., Actis-Grosso, R.: Social threat and motor resonance: when a menacing outgroup delays motor response. *Front. Psychol.* **7**, 1697 (2016)
36. Glenberg, A.M., Hayes, J.: Contribution of embodiment to solving the riddle of infantile amnesia. *Front. Psychol.* **7**, 10 (2016)
37. Evans, J.R., Abarbanel, A. (eds.): *Introduction to Quantitative EEG and Neurofeedback*. Elsevier (1999)
38. Richez, A., Olivier, G., Coello, Y.: Stimulus-response compatibility effect in the near-far dimension: a developmental study. *Front. Psychol.* **7**, 1169 (2016)
39. Cloninger, C.R., Przybeck, T.R., Svarkic, D.M., Wetzel, R.D.: *The Temperament and Character Inventory (TCI): A Guide to its Development and Use* (1994)
40. Maldonato, N.M., Sperandeo, R., Caiazzo, G., Cioffi, V., Cozzolino, P., De Santo, R.M., Nascovera, N.: Keep moving without hurting: the interaction between physical activity and pain in determining cognitive function at the population level. *PLoS ONE* **13**(6), e0197745 (2018)
41. Hofer, S.M., Horn, J.L., Eber, H.W.: A robust five-factor structure of the 16PF: strong evidence from independent rotation and confirmatory factorial invariance procedures. *Personality Individ. Differ.* **23**(2), 247–269 (1997)
42. Appollonio, I., Leone, M., Isella, V., Piamparta, F., Consoli, T., Villa, M.L., Nichelli, P.: The Frontal Assessment Battery (FAB): normative values in an Italian population sample. *Neurol. Sci.* **26**(2), 108–116 (2005)
43. Bryman, A., Cramer, D.: *Quantitative Data Analysis with IBM SPSS 17, 18 & 19: A Guide for Social Scientists*. Routledge (2012)
44. Costello, M.C., Bloesch, E.K.: Are older adults less embodied? A review of age effects through the lens of embodied cognition. *Front. Psychol.* **8**, 267 (2017)
45. Lécuyer, A., Lotte, F., Reilly, R.B., Leeb, R., Hirose, M., Slater, M.: Brain-computer interfaces, virtual reality, and videogames. *Computer* **41**(10), 66–72 (2008)
46. Maldonato, N.M., Sperandeo, R., Moretto, E., Dell'Orco, S.: A non-linear predictive model of borderline personality disorder based on multilayer perceptron. *Front. Psychol.* **9**, 447 (2018)
47. Barca, L., Mazzuca, C., Borghi, A.M.: Pacifier overuse and conceptual relations of abstract and emotional concepts. *Front. Psychol.* **8**, 2014 (2017)
48. Ninaus, M., Moeller, K., Kaufmann, L., Fischer, M.H., Nuerk, H.C., Wood, G.: Cognitive mechanisms underlying directional and non-directional spatial-numerical associations across the lifespan. *Front. Psychol.* **8**, 1421 (2017)

# Chapter 40

## Online Handwriting and Signature Normalization and Fusion in a Biometric Security Application



Carlos Alonso-Martinez and Marcos Faundez-Zanuy

**Abstract** In this paper, we analyze the combined application of signatures and capital handwriting in a biometric recognition application. We combine a signature recognition system based in a multi-section vector quantization with a handwriting text recognition system based in self-organizing maps and DTW. Due to the need to normalize the scores before the combination, we study the effect of different normalization methods and we propose the application of a logarithmic transformation for signature scores previous normalize them. Experimental results show that the identification rate raises from 86.11% using capital letter words and 96.95% using signatures up to 99.72% with a fusion of both traits. Minimum detection cost function (DCF) also improves, from 3.56 and 3.51%, respectively, up to 1.0% using the fusion of both traits.

### 40.1 Introduction

Automatic handwriting-based personal identification can be divided into two different fields: signature-based recognition and text-based recognition. Signature has a social accepted role as proof of identity, and for the reason, it has been extensively analyzed while the text-based recognition analysis has been done to a lesser extent. The paper is organized as follows. Section 40.1 is devoted to the state-of-the-art overview in online signature and handwritten text. Section 40.2 deals with experimental results using a linear combination of both traits. Section 40.3 includes conclusions and future works.

While multimodal biometric is a popular topic with a wide literature, the particular case of combining signatures and handwriting has been treated in only a few papers. The first one by Boulétreau et al. [1] is based on 48 writers that produce 20 signatures and 20 writing tasks. All the data are acquired in offline mode. The main conclusion is that in the field of authentication of documents, e.g., deeds, using signature and

---

C. Alonso-Martinez · M. Faundez-Zanuy ()  
ESUP Tecnocampus (UPF), Av. Ernest Lluch 32, 08302 Mataró, Spain  
e-mail: [faundez@tecnocampus.cat](mailto:faundez@tecnocampus.cat)

handwriting is useful to improve the recognition accuracy of a signature alone. Fractal analysis provided evidence of the independence between the behaviors of the writer when he signs and when he writes. Such independence will be a source of very enriching information within the context of signature authentication. In the paper published by Khalifa and Najoua [2], the study is also focused in offline handwriting and signatures, thus, without considering the in-air trajectories, which have been proven to contain useful information for recognition purposes. In another paper by Eshwarappa and Latte [3], they combine speech signature and handwriting in a very small database of 30 people, which is an important drawback of this paper. In fact, their speaker recognition system provides 100% identification rate and 0% verification error, making the combination with other modalities unnecessary.

In our opinion, the main obstacle to analyze and compare signatures and handwriting and the reason for the few amounts of literature on this topic is the limited number of available databases.

This paper is the continuation of our previous work based on signature [4] and handwriting [5] in two main ways: comparison between parameters of both biometric traits and the potential of a combined biometric system based on signature and handwriting.

## 40.2 Experimental Results

### 40.2.1 Biometric Database

In this paper, the BIOSECUR-ID database [6] has been used. BIOSECUR-ID main characteristics are: It is a multimodal biometric database, that includes speech, iris, face (still images and videos), handwritten signature and text, fingerprints, hand, and keystroking; with respect to handwritten text, this database defines five different tasks: a Spanish text in lower case, ten digits written separately, sixteen Spanish words in upper case, four genuine signatures, and one forgery of the three precedent subjects. The database comprises four sessions where the different tasks from 400 subjects were recorded.

### 40.2.2 Online Signature Biometric Recognition

We have used a multi-section vector quantization algorithm for signature recognition as is described in our previous work [4]. This method split signatures in N-sections of equal length and generated N codebook per user, one for each section. In the recognition step, the quantization distortion is obtained by the combination of the distortions obtained for each section. A subset of 320 subjects, the same ones that were used as train and test set, at handwritten work [5]. The four genuine signatures

from sessions 1, 2, and 3 are used for user model computation, and the ones from session 4 are used for the test.

Table 40.1 summarizes the experimental results of IDR (identification rate) from modeling each subject with a codebook. We present results for codebook sizes ranging from 1 to 8 bits (2–256 clusters) and a number of sections from 1 to 5.

We can observe how the best identification result, 96.95%, is obtained with a codebook of 4 sections and 5 bits per section (32 clusters).

In Table 40.2, we present the minimum detection cost function (DCF) for the same conditions. Here, the best value, 0.0315, is obtained with a codebook of 3 sections and 7 bits per section (128 clusters).

We choose the combination 4 sections and 5 bits per section for the subsequent combinations with handwriting text.

**Table 40.1** Identification results (IDR) for several multi-section codebook from 1 to 5 sections

Bits/section	1 section	2 sections	3 sections	4 sections	5 sections
1	54.53	79.45	66.80	92.19	66.09
2	74.61	90.94	87.11	96.17	84.06
3	88.28	94.77	92.89	96.56	92.97
4	91.25	95.49	95.31	96.80	94.53
5	91.25	94.92	95.08	96.95	94.14
6	90.78	94.77	95.39	96.41	93.67
7	90.47	93.67	95.63	95.86	92.73
8	89.30	93.59	95.70	95.39	91.80

**Table 40.2** Minimum DCF results for different multi-section codebook sections

Bits/section	1 section	2 sections	3 sections	4 sections	5 sections
1	0.2855	0.2193	0.1507	0.1288	0.1818
2	0.1997	0.1311	0.0853	0.0725	0.1082
3	0.1300	0.0818	0.0552	0.0478	0.0659
4	0.0957	0.0618	0.0429	0.0379	0.0517
5	0.0791	0.0530	0.0363	0.0351	0.0452
6	0.0715	0.0481	0.0319	0.0331	0.0433
7	0.0689	0.0468	0.0315	0.0328	0.0437
8	0.0667	0.0455	0.0323	0.0318	0.0445

### 40.2.3 *Online Handwritten Capital Letters Biometric Recognition*

As described in [5], it is based on an approach that relies on catalogs of strokes built in an unsupervised manner by means of self-organizing maps and on dynamic time warping to compare the sequences of strokes that constitute the sequences of online text. Authors took the first 50 useful subjects uppercase word tasks from 4 sessions, to obtain SOM (self-organizing maps) and the others 320 subjects in train and test tasks. Specifically, they used sessions 1 to 3 for training purposes and session 4 for the test.

In Table 40.3, we show the results of the identification rate (IDR) and detection cost function (DCF). It can be observed how identification rate (IDR) and minimum detection cost function (DCF) present a great variation in function of a word, worst IDR value is 76.25% while the best value up to 94.69%. For minimum detection cost (DCF), values vary between 0.05340 and 0.0157. Best results are obtained with the longest word, but equal length words present significant differences. The mean results show an IDR of 86.11% and a minimum DCF of 0.0356.

**Table 40.3** Identification rate and minimum detection cost function for handwriting tasks

Word number	Text	Length	IDR	DCF
1	BIODEGRADABLE	12	89.69	0.0236
2	DELEZNABLE	10	85.00	0.0292
3	DESAPROVECHAMIENTO	18	94.69	0.0157
4	DESBRIZNAR	10	84.38	0.0338
5	DESLUMBRAMIENTO	15	93.13	0.0258
6	DESPEDAZAMIENTO	15	90.63	0.0346
7	DESPRENDER	10	81.25	0.0442
8	ENGUALDRAPAR	12	85.00	0.0432
9	EXPRESIVIDAD	12	86.25	0.0371
10	IMPENETRABLE	12	85.63	0.0357
11	INEXPUGNABLE	12	94.06	0.0272
12	INFATIGABLE	11	90.00	0.0200
13	INGOBERNABLE	11	85.63	0.0365
14	MANSEDUMBRE	11	79.69	0.0555
15	ZAFARRANCHO	11	76.25	0.0540
16	ZARRAPASTROSA	13	76.56	0.0527

#### 40.2.4 Combined Approach for Biometric Recognition

In this paper, we use a linear combination of the matching scores coming from signature-based classifier and the ones obtained by the word classifier (40.1), where  $O_i$  is the score for trait  $i$  (1 per signature and 2 for handwriting words) and  $\alpha$  is the combination factor:

$$O = \alpha O_1 + (1 - \alpha) O_2 \quad (40.1)$$

This combination technique is also known as opinion fusion and requires the normalization of the distances because of the different dynamic range of each of the classifiers. We have studied several normalization techniques and their impact on the recognition system performance. A comparison of several normalization strategies can be found in [7].

#### 40.2.5 Score Normalization

In this section, we present the four normalization methods. We denote a matching score as  $o$ , from the set of all corresponding scores, being  $o'$  the corresponding normalized score.

**MinMax:** This is the simplest technique and it assures that the dynamic range of normalized scores for each of the classifiers is in the interval [0,1]. For a given set of matching scores, the normalized scores are given by (40.2):

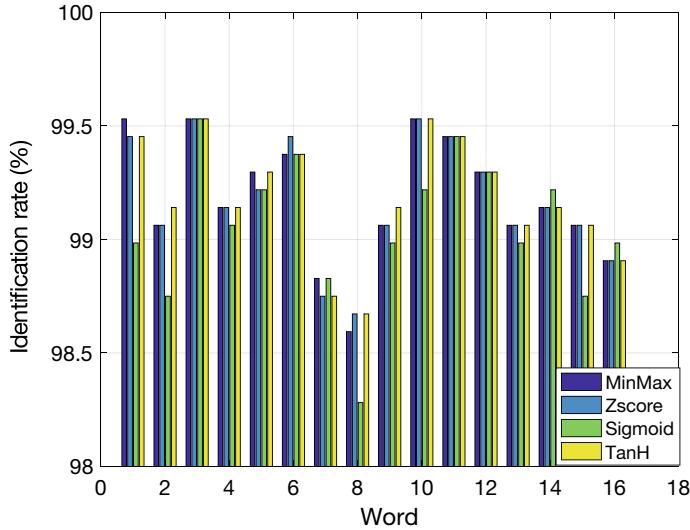
$$o'_k = (o_k - \min(O)) / (\max(O) - \min(O)) \quad (40.2)$$

**Z-score:** This method transforms the scores into distribution with mean 0 and standard deviation of 1. In (40.3),  $u$  denotes mean value and  $s$  the standard deviation of matching scores whole set.

$$o'_k = (o_k - u) / s \quad (40.3)$$

**Sigmoid:** A sigmoid function is used as a method, where the normalized scores are given by (40.4), being  $u$  the mean and  $s$  the standard deviation of matching scores set.

$$o'_k = 1 / (1 + \exp(-v_k)) \text{ where } v_k = (o_k - u) / s \quad (40.4)$$



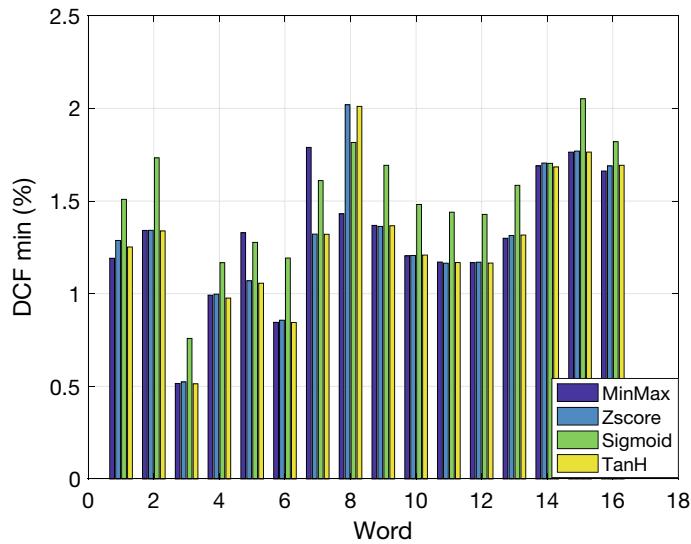
**Fig. 40.1** Identification rate for every single word using the different normalization methods

**Tanh:** This method like MinMax transforms the scores into the [0,1] range. Again,  $u$  denotes arithmetic mean and  $s$  standard deviation of the matching score set. Tanh means hyperbolic tangent operator.

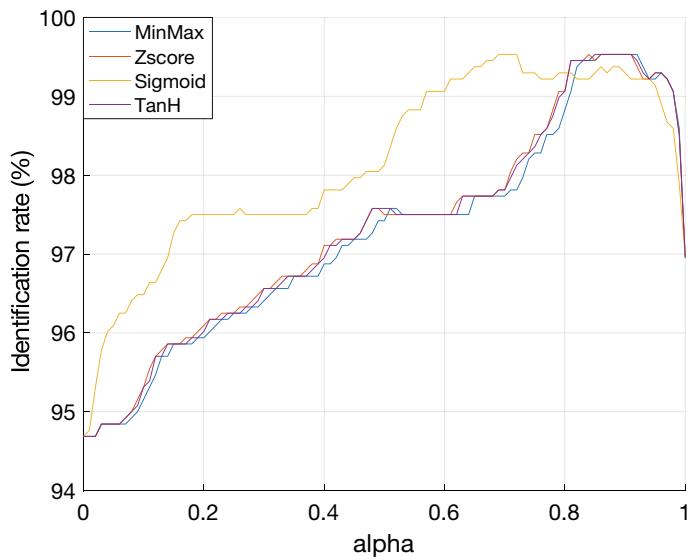
$$o'_k = 0.5[\tanh(0.01(o_k - u)/s) + 1] \quad (40.5)$$

In Figs. 40.1 and 40.2, we show the identification rate (IDR) and the minimum detection cost function (DCF) values for the different words and the four normalization methods. We can observe a significant improvement in recognition results. Average results of identification rate up from 86.11% for handwriting text and 96.95% for the signature to 99.18% using fusion with Tanh or MinMax normalization. For minimum detection cost function, the average results up from 0.0356 and 0.0351 to 0.0129 and 0.0130 using Tanh and Z-score or MinMax, respectively. The worst results are obtained with sigmoid normalization both in IDR and in DCF.

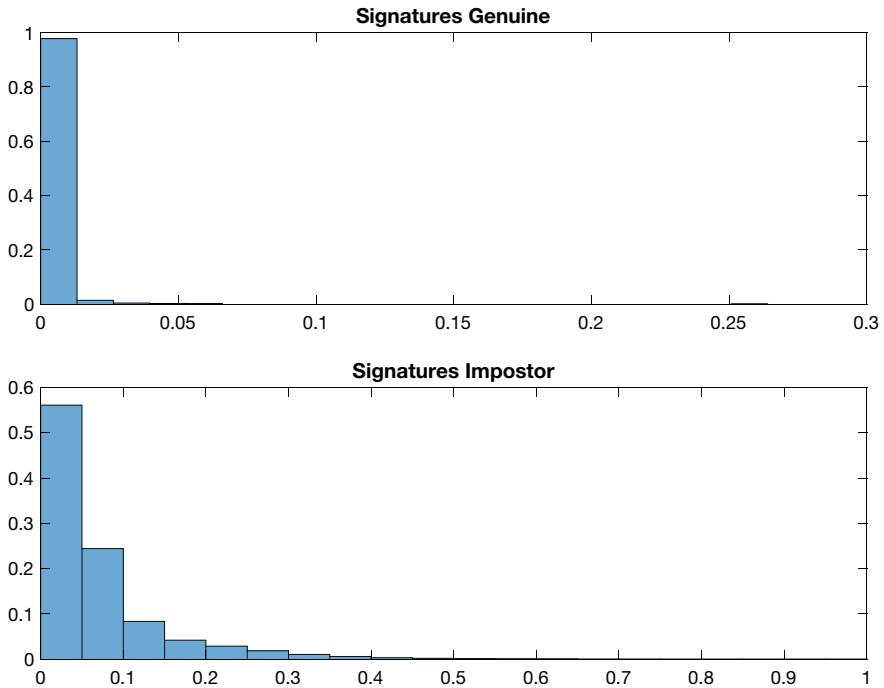
In Fig. 40.3, we represent the identification rate as a function of  $\alpha$  value for word number 3 (DESAPROVECHAMIENTO). It can be shown how the signature classifier has a strong weight in fusion because we can observe how the optimal value is achieved for values between 0.8 and 1, except for sigmoid that has the peak around an alpha value of 0.7. This means that mainly signature matching scores have more importance in the decision than handwriting ones.



**Fig. 40.2** Minimum detection cost for every single word using the different normalization methods



**Fig. 40.3** Identification rate as a function of combination factor  $\alpha$

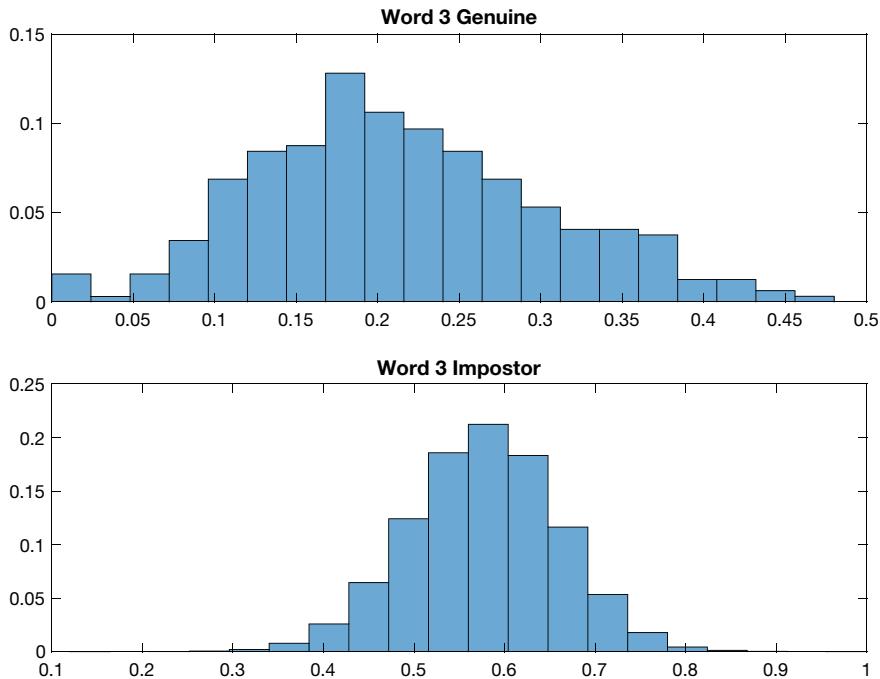


**Fig. 40.4** Histograms for signatures using MinMax normalization

When we represent the histograms of genuine and impostor normalized scores, using MinMax normalization for signatures (Fig. 40.4) and handwriting (Fig. 40.5), it can be observed that while handwriting scores follow a Gaussian distribution, the signature scores' impostor histogram shows an exponential distribution.

We propose a previous logarithmic transform for signature matching scores, previous to normalization [8]. As handwriting matching scores follow a Gaussian distribution, this previous transformation is not needed. In Fig. 40.6, it can be observed how after logarithmic transformation, the histogram for signature matching scores takes a more normal distribution.

In Figs. 40.7 and 40.8, we present the identification ratio (IDR) and minimum detection cost function (DCF) for the four normalization methods with previous logarithmic transformation of signature matching scores.



**Fig. 40.5** Histograms for word number 3 MinMax normalization

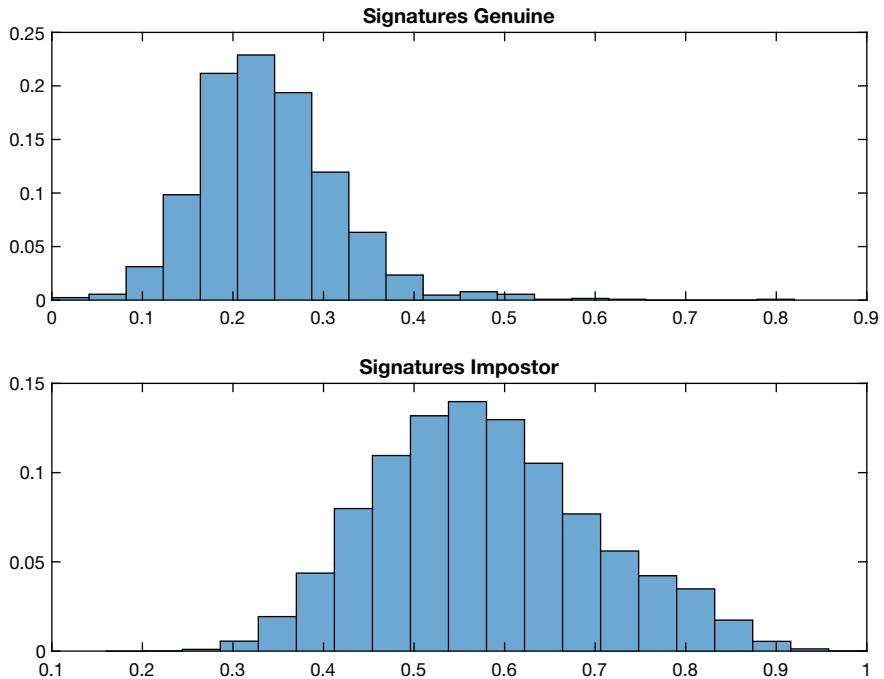
### 40.3 Conclusions

Signature and handwriting are two biometric systems widely studied, which allow good recognition rates, although, in the case of handwriting, with great dependence on the word used. The combination of both methods allows an improvement of the obtained values. However, the fact that the distances in the classifiers present different dynamic margins requires normalization prior to the combination. Normalization models typically used have been defined for distributions that fit a Gaussian model.

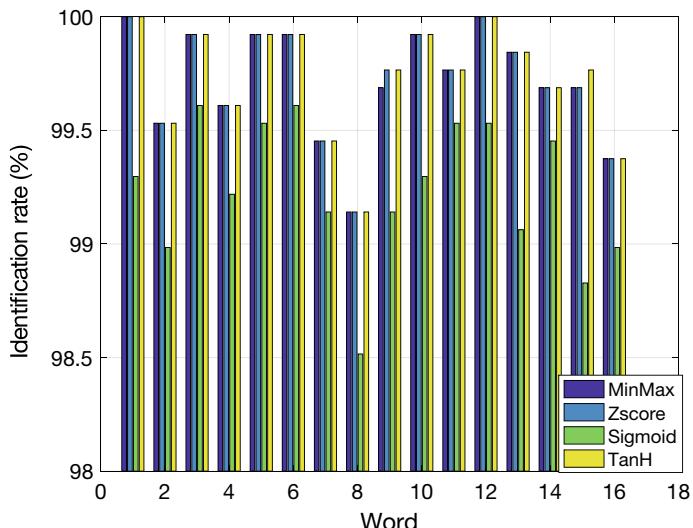
We have observed how the distribution in the case of firms adjusts more to an exponential model. For this reason, the application of a logarithmic transformation prior to normalization presents better results than combinations without such transformation.

The proposed method improves the recognition results from each individual classifier (96.95 and 94.69%) up to 100% for identification rate and from 3.51 and 1.57% up to 0.46% for minimum distance cost function (DCF) in the most favorable cases. In average, the fusion of signatures with logarithmic transformation and handwriting text, using Tanh normalization achieves 99.72% for IDR and 1% for minimum DCF.

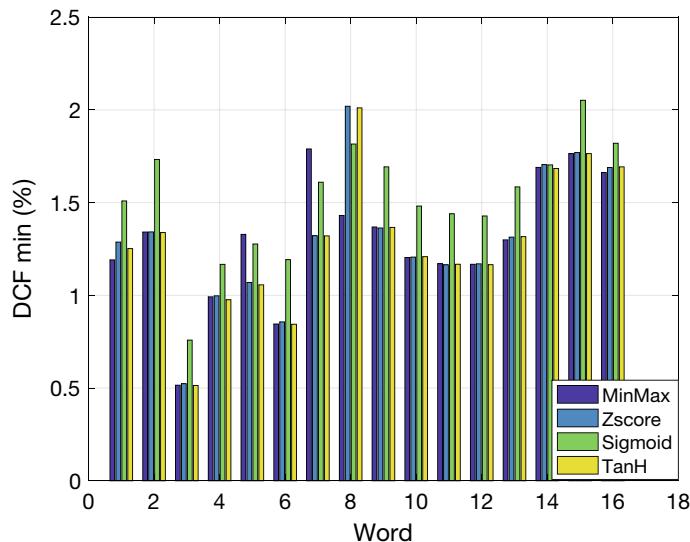
**Acknowledgements** This work has been supported by FEDER and MEC, TEC2016-77791-C4-2-R.



**Fig. 40.6** Histograms for signature MinMax normalization and log transformation



**Fig. 40.7** Identification rate for every single word using the different normalization methods and logarithmic transformation for signature scores



**Fig. 40.8** Minimum detection cost function for every single word using the different normalization methods and logarithmic transformation for signature scores

## References

1. Bouleatreu, V., et al.: Handwriting and signature: one or two personality identifiers? In: Proceedings of the 45th International Conference on Pattern Recognition, pp. 1758–1760. IEEE, Brisbane (1998)
2. Khalifa, A.B., Amara, N.E.B.: Fusion at the feature level for person verification based on offline handwriting and signature. In: Proceedings of the 2nd International Conference on Signal, Circuits and Systems, pp. 1–5. IEEE, Monastir (2008)
3. Eshwarappa, M.N., Latte, M.V.: Multimodal biometric person authentication using speech, signature and handwriting features. Int. J. Adv. Comput. Sci. Appl., 1–10 (2011) (Special Issue on Artificial Intelligence)
4. Faundez-Zanuy, M., Pascual-Gaspar, J.M.: Efficient on-line signature recognition based on multi-section vector quantization. Pattern Anal. Appl. **14**(1), 37–45 (2011)
5. Sesa-Nogueras, E., Faundez-Zanuy, M.: Biometric recognition using online uppercase handwritten text. Pattern Recogn. **45**(1), 128–144 (2012)
6. Fierrez, J., et al.: BiosecurID: a multimodal biometric database. Pattern Anal. Appl. **13**(2), 235–256 (2010)
7. Snelick, R., et al.: Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems. IEEE Trans. Pattern Anal. Mach. Intell. **27**(3), 450–455 (2005)
8. Saisani, K.L.: Dealing with non-normal data. PM&R **4**(12), 1001–1005 (2012)

# Chapter 41

## Data Insights and Classification in Multi-sensor Database for Cervical Injury



Xavi Font, Carles Paul and Eloi Rodriguez

**Abstract** The aim of this work is to produce a classification system to assess cervical injury from a multi-sensor database. This is a database that has never been gathered before, with data coming from three different sensors: inertial, EEG, and thermography. How well these sensors help to build a good model and which right features influence the response variable are key to a better understanding of the link between sensor data and the presence of a cervical injury. There is an additional data set related to the user/patient (a survey) that will provide us with a different view and a baseline to compare with the sensor data. Both data sets are characterized by few observations and many variables, such that feature selection or feature engineering is crucial to success in building a good classification system. The approach used with both data sets is penalized logistic regression. This type of models helps in the selection of the best features and also prevents overfitting by adding a penalized term over the coefficients in the objective function. To verify the ability of each data set, a k-fold cross-validation was carried out with promising results for both the sensor and survey data sets. Results demonstrate an accuracy of up to 70%, which gives a good starting point to encourage increasing the size of the database with more patients. A final discussion highlights the importance of never underestimating the information provided by the patients.

### 41.1 Introduction

Health studies always present real challenges that must be overcome in order to obtain reasonable results. The steps one should follow to conduct health studies can be found in many journals [1–3]. The basic steps involved consist of: question identification, approach analysis, experimental design, data acquisition, data analysis, and

---

X. Font (✉) · C. Paul · E. Rodriguez  
Escola Superior Politècnica (ESUPT-TCM-UPF), Mataró, Barcelona, Spain  
e-mail: [font@tecnocampus.cat](mailto:font@tecnocampus.cat)

C. Paul  
e-mail: [paul@tecnocampus.cat](mailto:paul@tecnocampus.cat)

conclusions. However, in some situations, due to a lack of budget or excessive cost many studies only pinpoint some of these steps.

The aim of this work is framed on the multi-sensor database for the cervical area. The complete data coming from three different sensors (Thermal: Infrared Thermography H2640 from Nippon Avionics, EEG: from Emotiv+ and Inertial: 3-axis accelerometer, 3-axis gyroscope and 3-axis Magnetometer) as well as a survey data have been available to assess a typical classification question: reasons behind neck injury (classify feature injury). However, given the set of sensors, we might be able to assess if there is any sensor that is more suited to the problem at hand than others. We might even compare the performance of the system between sensor and survey data.

The figures related to neck injury are very significant. For instance, they can reach more than \$30b (economic expenses in the USA) due to medical care, disability, loss of productivity, and also litigation. Another important issue comes on how to correctly diagnose patients. Diagnostic failure can bring chronic psychological symptoms (depression, anxiety, etc.). The usual approach to diagnosis is through hospital where an X-ray or MRI scan can be used based on symptoms.

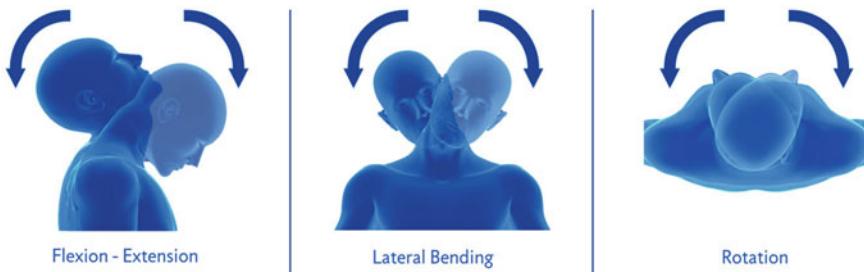
The results will apparently emerge with the exploratory data analysis or with the application of penalized regression to check the relationship between the set of features contained in the design matrix  $X$  and the target variable  $Y$  that describes neck injury. This preliminary study needs to address a huge amount of data which has emerged from the three different sensors. Thus, before deploying any advanced techniques, it is mandatory to reduce the number of features. These highly unwanted dimensions need a suitable process of reduction to increase data insights and help in classification methods. Many advanced methods cover this field of data reduction techniques [4–6].

## 41.2 Experiment Setup and the Data Set

The data set used in this paper comes from an experiment designed to obtain certain measures related to the range of motion involving the neck area. Each user goes through a process that involves: extension/flexion (or yes movement), inclination (or I do not know movement) and rotation (or no movement). These three steps are shown in Fig. 41.1.

Before the acquisition process started, users fill in a brief survey with some questions related to general statements like gender, age, and so on. Some of these questions were placed with no further intention than to check the physical condition of the user/patient. Some of the data gathered from these survey are described in the following list:

- *riskProf*: whether your work puts the column and cervical area at risk, because of efforts
- *regularSport*: do you do sport regularly? Yes/No

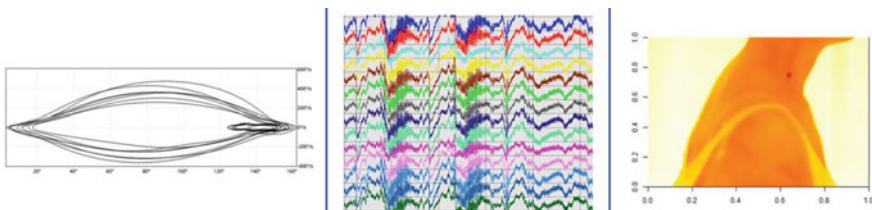


**Fig. 41.1** Movements (from left to right): Yes; I don't know; No

- *alcoholDrink*: how often do you drink alcohol? Never/Low–Moderate/Moderate–High
- *sleep*: how many hours do you sleep? 3–6/6–9/9–12
- *injuryBef*: have you ever had an accident (degree of it)? Never/Low/Moderate/High
- *injuryNow*: have you had an accident quite recently (degree of it)? None/Moderate/High
- *stressLev*: what is your stress level? 0–10
- *painLinkStr*: what type of pain do you have? None/Headache/Cervical
- *levelPain*: degree of pain: None/Slight/Moderate/High.

Basically, we have data coming from two different sources: first, the survey information given by the user (some obviously subjective) which represents user's perception. Second, data coming from three different sensors. We gather data from the user's physiology through three different sensors: an inertial sensor, a thermographic camera, and a EEG headset. The goal is to classify a binary variable *cervicalInjury* (saying whether the user has a cervical injury) according to these covariates which are analyzed first with the sensor data and then with the survey data.

The amount of data available for each sensor is different. In the following subsections, we will describe the data that was used to feed the model. In order to reduce and simplify the complexity of the original data set (see Fig. 41.2), two straightforward approaches are mainly used—dimensionality reduction and/or feature engineering. Without wishing to confuse the former one is a process of using domain knowledge



**Fig. 41.2** Data from sensors (from left to right): inertial data (position vs. velocity); EEG data; thermographic image (lateral right picture)

to extract new features from the original data to help with interpretation and mostly support the machine learning algorithms work. This is the approach used in this work.

### **41.2.1 Inertial Data**

Instead of working with the complete data coming from the inertial sensor, it represents: extension/flexion, inclination left/right and rotation left/right (plus the same with velocity and acceleration) only percentage and deviation is computed to reduce the data to the following variables: *yesInerPer*, *dontknowInerPer*, *noInerPer*, *maxDevExtFlex*, *maxDevIncl*, *maxDevRot*, *maxYesDev*, *maxDontKnowDev* and *maxNoDev*. This set of data was carefully selected with the help of a physician using the position velocity plot (see first picture from Fig. 41.2).

### **41.2.2 Thermographic Data**

The data was basically a thermographic image with temperature information [7]. To reduce this set of images drastically, we just maintain the different temperature of the highest temperature point on each of the sides studied (right, back and left). At the end of the process, three variables were created: *rightDif*, *backDif* and *leftDif*. Again, the reduction of data is remarkable. From a set of two groups of ten thermographic images with a resolution of  $640 \times 480$ , we just retain a number that represents the difference (in  $^{\circ}\text{C}$ ) between before and after the experiment. The right picture from Fig. 41.2 shows a red spot linked to the highest temperature.

### **41.2.3 EEG Data**

Data was acquired with the EMOTIV + EEG headset that gives 14 channels, plus additional gyroscope data. Some of this data was highly correlated with the inertial sensors. Before removing all this data, we decided to average some of the sensors and keep variance information (deviations and IQR). The following variables were computed: *corticDevEEG*, *corticIqrEEG*, *temporalDevEEG*, *temporalIqrEEG*, *parietalDevEEG* and *parietalIqrEEG*.

### **41.2.4 Data Summary**

The number of cases ready for the study ranged between 20 and 60 years old. There were 14 subjects with no cervical injury and 13 with some level of injury. The complete set of users are shown in Table 41.1.

**Table 41.1** Subjects enrolled in the multi-sensor database for cervical area

	Subjects with injury	Healthy subjects	Total
20–40 years	4 women + 3 men	4 women + 3 men	8 women + 6 men
41–60 years	3 women + 3 men	3 women + 4 men	6 women + 7 men
Total	7 women + 6 men	7 women + 7 men	14 women + 13 men

The problem is reduced to a data set that can be split into two mutually exclusive data sets: sensor and survey.

$$\mathcal{D} = \mathcal{D}_{\text{sensors}} + \mathcal{D}_{\text{survey}}$$

where  $|\mathcal{D}_{\text{sensors}}| = 27 \times 18$  and  $|\mathcal{D}_{\text{survey}}| = 27 \times 31$ .

### 41.3 Data Insights

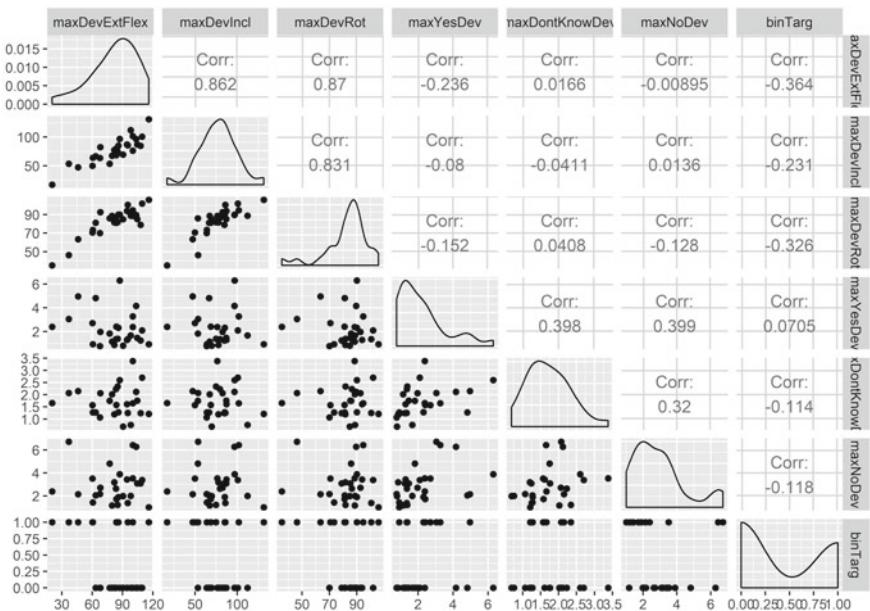
With the new features already computed, it is now possible to see some graphical representations related to inertial, EEG, and thermographic data. Because the target variable is binary, the relationship with the target is not easily seen. For instance, Fig. 41.3 shows the output between the  $Y$  variable and the remaining  $X$  variables (all from inertial sensors) at the bottom of the representation. There is no clear feature that stands out as the best to select.

With a detailed look of the data coming from the EEG sensor, the message given is even clearer. There are no signs of linear relationship between EEG features and the target variable (see Fig. 41.4). Nevertheless, the correlation matrix shows a strong relationship between the EEG variables. From a statistical point of view, this type of variable will generate multicollinearity problems. From the perspective of signal processing, the data seems to show a lot of interference. It is true that the device used is more suited for brain–computer interface (BCI) than medical studies.

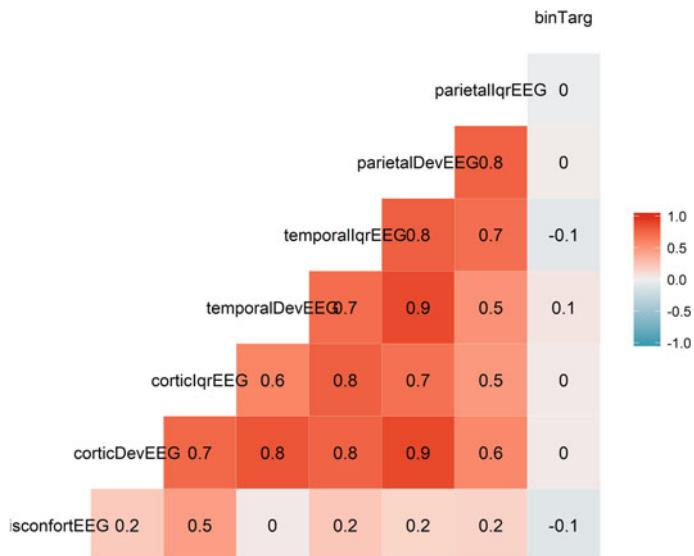
From the perspective of thermal data, Fig. 41.5 shows the target variable and the factor regular sport (*regularSport*). No evident conclusion related to the link between injured versus healthy and the remaining variables is shown in the plot.

### 41.4 Classification Through Penalized Regression

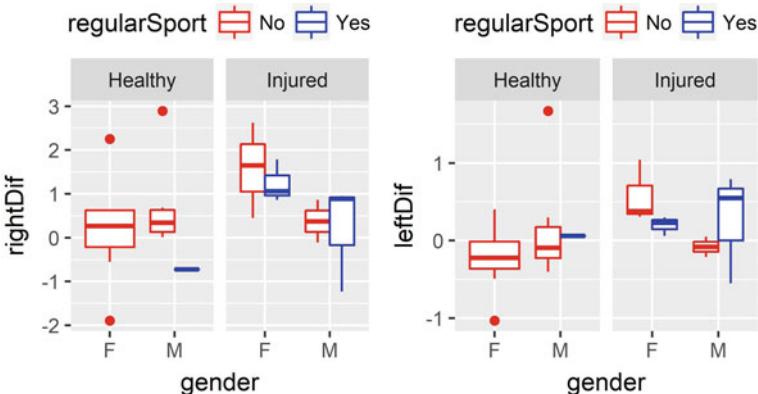
The complexity of the data is evident from the exploratory data analysis previously performed. This complexity can be interpreted in different ways: the number of users is pretty low, the number of sessions conducted by users varies from 1 to 2, and the data obtained from the feature engineering is not good enough.



**Fig. 41.3** Inertial data matrix with the target variable



**Fig. 41.4** Correlation matrix from EEG variables and the target variable



**Fig. 41.5** Distributions from rightDif and leftDif across some categorical variables

To overcome some of the problems that are present in general regression models, a penalized regression model will be used. The idea is to overcome two problems at once, that is: selecting the variables that have an effect on the target variable and avoiding overfitting.

The logistic regression is a model which does not work when applied in a straight way. However, through a penalized model, the problem at hand can be solved. When the number of variables is greater than the number of observations ( $p > n$ ) or just close enough, the algorithm does not offer convergence.

The function for the penalized logistic regression uses the form of a negative binomial log-likelihood:

$$\min_{(\beta_0, \beta) \in \mathbb{R}^{p+1}} - \left[ \frac{1}{N} \sum_{i=1}^N y_i \cdot (\beta_0 + x_i^T \beta) - \log(1 + e^{(\beta_0 + x_i^T \beta)}) \right] + \lambda [(1-\alpha) \|\beta\|_2^2 / 2 + \alpha \|\beta\|_1] \quad (41.1)$$

We deployed penalized regression through three possible alternatives: Lasso, Ridge, and Elastic Net (see [8, 9]). Lasso shrinks the regression coefficients toward zero by penalizing through the sum of the absolute coefficients (L1 norm). Ridge produces a shrinkage of the coefficients by penalizing the sum of the squared coefficients (L2 norm). And, the last approach Elastic Net tries to find a trade-off between Lasso and the Ridge penalties.

## 41.5 Results

The two data sets were used to provide a reasonable approach because the number of observations is not big enough. Thus, it is highly recommended to use a k-fold cross-validation to ensure a better estimation of the performance of the classification algorithms tested.

### 41.5.1 Sensor Data

Sensor data with ridge regression gives an average accuracy close to 60% and picks as best predictors three variables related to inertial movement and one related to thermal data. Interestingly, inertial data shows variables linked to the three exercises used in the experimentation. The Lasso approach gives a solution that only selects two variables: one from the inertial sensor and one from the thermal data. Accuracy has improved to 66%. The last option goes through the Elastic Net, which shows a mixed selection of inertial variables (*dontknowInerPer* and *noInerPer*) and thermal variables (*leftDif* and *RightDif*). The accuracy in this last method reaches 70%. Such cases are shown in Tables 41.2, 41.3 and 41.4. As illustrated in these tables, they provide us with the best selected coefficients (feature selection) and their estimated values.

### 41.5.2 Survey Data

The analysis of survey data at the beginning was a step which was not intended to perform. The reason to avoid the data survey was that the sensor data was more

**Table 41.2** Regression coefficients for sensor data with ridge regression

don'tknowInerPer	maxDevExtFlex	maxDevRot	leftDif
-0.018825	-0.0152218	-0.0140756	0.0149467

**Table 41.3** Regression coefficients for sensor data with Lasso regression

maxDevExtFlex	leftDif
-0.0018906	0.0949152

**Table 41.4** Regression coefficients for sensor data with Elastic Net regression

leftDif	don'tknowInerPer	noInerPer	rightDif
0.4672984	-0.3652328	-0.1806678	0.1171872

**Table 41.5** Regression coefficients for survey data with Ridge regression

regularSportYes	riskProfYes	injuryNowHigh	injuryNow Moderate	painLinkStr Cervical
0.1394922	0.1394923	0.145211	0.1556112	0.1778166

**Table 41.6** Regression coefficients for survey data with Lasso regression

injuryNowHigh	injuryBefLow	injuryBefHigh	painLinkStr Cervical	riskProfYes
0.2742874	0.2981609	0.387151	0.5445093	0.7842023

objective and more reliable. Thus, this section is intended to provide a reasonable baseline to compare with the sensor data.

It is not strange to see that many of the variables picked by the Ridge regression method look reasonable. The variable that tells us whether the user has pain in the cervical area obtained the highest estimated value. The next two variables selected describe information about a recent trauma on the cervical area, which certainly looks related to the injury he or she is suffering from. The Ridge process obtains an accuracy close to 67%.

With Lasso approach, it turns out that the variable related to pain is again selected. The first variable selected is related to the risk of the job. This is understood as a risk for cervical injuries. The remaining variables describe injuries back in the past or more recently. Again, the accuracy is closed to 70%. Tables 41.5 and 41.6 can be seen as the best features selected from the whole set of survey variables and their corresponding estimated value.

## 41.6 Conclusions

The main conclusion drawn from this preliminary work is to never underestimate the information provided by users through a survey or in any other way, basically, because it may be very helpful in obtaining new insights or add further data to the whole data set under scrutiny.

Sensors which are robust to movements like inertial sensors and thermographic camera have shown its ability to explain the target variable. On the other side of the sensors used, EEG does not provide any reasonable result whatsoever. Since the results show how two completely different sensors help in characterizing the target variable (as a way to encourage diversification in data acquisition), the analysis will be extended to a larger set of data.

**Acknowledgements** This work has been supported by FEDER and MEC, TEC2016-77791-C4-2-R.

## References

1. Squires, H., Chilcott, J., Akehurst, R., Burr, J., Kelly, M.P.: *Value Health* **19**(5), 588–601 (2016)
2. Jones, C.M., Clavier, C., Potvin, L.: *Soc. Sci. Med.* **177**, 69–77 (2017)
3. de Zoysa, I., Habicht, J.P., Peltó, G., Martínez, J.: *Bull. World Health Organ.* **76**(2), 127–133 (1998)
4. Wang, S., Ding, C., Hsu, C.-H., Yang, F.: *Future Gener. Comput. Syst.* (2018)
5. Tao, C., Feng, J.: *Comput. Stat. Data Anal.* **107**, 131–148 (2017)
6. Forzani, L., García Arancibia, R., Llop, P., Tomassi, D.: *Comput. Stat. Data Anal.* **125**, 136–155 (2018)
7. Fernández-Cuevas, I., Bouzas Marins, J.C., Arnáiz Lastras, J., Gómez Carmona, P.M., Piñonosa Cano, S., García-Concepción, M.Á., Sillero-Quintana, M.: *Infrared Phys. Technol.* **71**, 28–55 (2015)
8. Zou, H., Hastie, T.: *J. Royal Stat. Soc. Series B* **62**(2), 301–320 (2005)
9. Tibshirani, R.: *J. Royal Stat. Soc. Series B* **58**(1), 267–288 (1996)

## Chapter 42

# Estimating the Asymmetry of Brain Network Organization in Stroke Patients from High-Density EEG Signals



**Nadia Mammone, Simona De Salvo, Silvia Marino, Lilla Bonanno, Cosimo Ieracitano, Serena Dattola, Fabio La Foresta and Francesco Carlo Morabito**

**Abstract** Following a stroke, the functional brain connections are impaired, and there is some evidence that the brain tries to reorganize them to compensate the disruption, establishing novel neural connections. Electroencephalography (EEG) can be used to study the effects of stroke on the brain network organization, indirectly, through the study of brain-electrical connectivity. The main objective of this work is to study the asymmetry in the brain network organization of the two hemispheres in case of a lesion due to stroke (ischemic or hemorrhagic), starting from high-density EEG (HD-EEG) signals. The secondary objective is to show how HD-EEG can detect such asymmetry better than standard low-density EEG. A group of seven stroke patients was recruited and underwent HD-EEG recording in an eye-closed resting state condition. The permutation disalignment index (PDI) was used to quantify the coupling strength between pairs of EEG channels, and a complex network model was constructed for both the right and left hemispheres. The complex network analysis allowed to compare the small-worldness (SW) of the two hemispheres. The impaired hemisphere exhibited a larger SW ( $p < 0.05$ ). The analysis conducted using traditional EEG did not allow to observe such differences. In the future, SW could be used as a biomarker to quantify longitudinal patient improvement.

### 42.1 Introduction

There is an increasing evidence that cerebral stroke involves the disruption and subsequent reorganization of functional brain networks both close to the lesion and far away from it [1]. Therapy and rehabilitation aim at helping the brain to restore neural connections, compensating for disrupted circuits. Nowadays, choosing the best

---

N. Mammone (✉) · S. De Salvo · S. Marino · L. Bonanno  
IRCCS Centro Neurolesi Bonino-Pulejo, Via Palermo c/da Casazza, SS. 113, Messina, Italy  
e-mail: [nadia.mammone@ircsme.it](mailto:nadia.mammone@ircsme.it)

C. Ieracitano · S. Dattola · F. La Foresta · F. C. Morabito  
Mediterranea University of Reggio Calabria,  
Via Graziella, Feo di Vito, 89060 Reggio Calabria, Italy

treatment is a challenging process that can take weeks and is still not objectively measurable. Unfortunately, the ability of the brain to recover seems to decline over time; therefore, it is of paramount importance to find an objective biomarker of functional recovery. To this purpose, electroencephalography might be of great help as it is a non-invasive and well-tolerated neurophysiological examination which the patient can repeatedly and safely undergo in the short term.

Graph theory plays a key role in the study of cerebral complex network behavior. Cortical areas are represented as nodes, while edges connecting nodes represent the functional or structural connectivity between the areas [2, 3]. Functional connectivity, derived from EEG, investigates statistically significant correlations, in the frequency and/or time domain, between the electrical activity of different brain areas [4]. Brain functional connectivity is characterized by both local specialization and global integration which can be measured by two graph indices: the clustering coefficient ( $CC$ ), an index of functional segregation, and the characteristic path length ( $\lambda$ ), an index of functional integration [5]. There is some evidence that stroke can disrupt the balance between local processing and global functioning. Up to the present, most studies on functional reorganization after stroke are mainly derived from functional magnetic resonance imaging (MRI) [2].

The study of the functional connectivity in stroke patients derived from EEG signals is a barely explored field. Zeng et al. [6] evaluated the effect of stroke rehabilitation applying a novel nonlinear dynamic complexity measure to EEG signals, the mean nonlinearly separable complexity degree (MeanNLSD). They observed that MeanNLSD is smaller at the lesion regions. Caliandro et al. [2] tested whether ischemic stroke in the acute stage may determine changes in the small-worldness of cortical networks. They estimated the network characteristics in 30 consecutive stroke patients in acute stage. Network rearrangements were mainly detected in delta, theta, and alpha bands when patients were compared with healthy subjects. Liu et al. [7] investigated the nonlinear features of functional connectivity, derived from EEG signals, in patients with acute thalamic ischemic stroke and healthy subjects. Lempel-Ziv complexity (LZC), sample entropy (SampEn), and partial directed coherence (PDC) were calculated, and the stroke group exhibited weaker cortical connectivity. Zappasodi et al. [8] analyzed the resting EEGs of 36 patients and 19 healthy controls. The inter-hemispheric asymmetry (FDasy) and spectral band powers were calculated. FD was smaller in patients than in controls thus revealing the loss of complexity probably linked to the global system dysfunction.

The present paper aims at investigating the possible asymmetry in the brain network organization of the two hemispheres in stroke patients and finding a correlation between the presence of the lesion and the brain-electrical network behavior, through the analysis of high-density EEG (HD-EEG) signals. The secondary objective is to assess if such differences are detectable when the analysis is based on low-density EEG. The brain network organization is studied by means of a complex network model based on the estimation of the coupling strength between pairs of EEG signals. Seven stroke patients enrolled at IRCCS Centro Neurolesi Bonino-Pulejo (Messina, Italy) were involved in the study.

The paper is organized as follows: Sect. 42.2 illustrates how patients were enrolled in the study, how their EEGs were recorded and preprocessed, and EEG signals were used to construct the complex network model. Section 42.3 reports the results and Sect. 42.4 draws some conclusions.

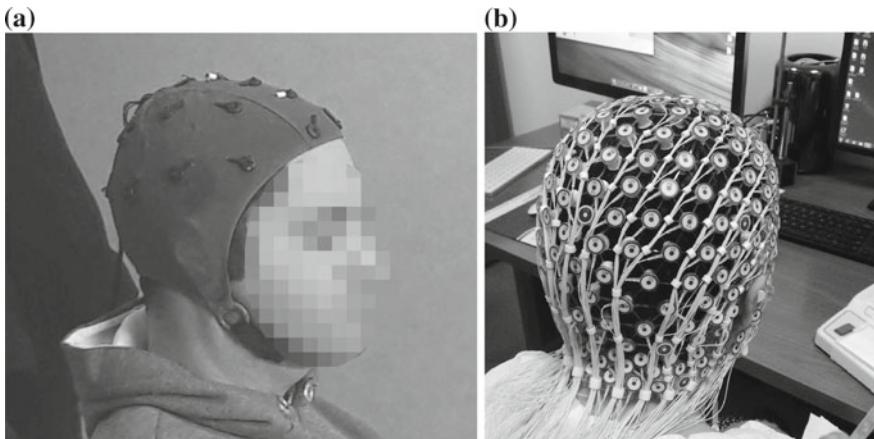
## 42.2 Estimating the Brain Network Organization

### 42.2.1 Patients' Recruitment

Seven patients (mean age 65 years, SD 12 years; six males and one female) were enrolled at the rehabilitative post-stroke unit of IRCCS Centro Neurolesi Bonino-Pulejo (Messina, Italy). Patients with previous ischemic or hemorrhagic stroke were excluded. The patients were clinically evaluated before EEG recording. Patients with a history of other neurological disease, history of traumatic brain injury, defects in sight, previous depression, or other psychiatric illnesses were excluded. The patients showed ischemic (3) and hemorrhagic (4) stroke to magnetic resonance imaging (MRI).

### 42.2.2 HD-EEG Recording and Preprocessing

EEG signals were recorded by means of a high-density 256-channels *EGI Sensor Net* that belongs to the *Electrical Geodesics EEG system* (Fig. 42.1b). The central channel ( $Cz$ ) was the reference electrode, and 250Hz was the sampling rate. Ev-



**Fig. 42.1** EEG recording system: **a** standard 19-channels EEG recording cap. **b** high-density 256-channels *Electrical Geodesics EEG system*

ery channel records the differential potential between its own electrode location  $x$  and Cz. Following the EGI guidelines, electrode impedances were kept lower than  $50\text{ k}\Omega$ . The scalp is uniformly covered by the electrodes of the sensor net shown in Fig. 42.1b. The HD-EEG was recorded in an eye-closed resting state. The physician reviewed the EEG traces in real time in order to detect a possible pattern of drowsiness and keep the subject awake. The EEG signals were bandpass-filtered in the range 1–40 Hz by means of *Net Station EEG software* of the *Electrical Geodesics EEG system* in order to capture all the EEG rhythms: delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), gamma ( $>30$  Hz). The filtered recordings were cleaned by the EEG experts who rejected the artifactual epochs. For every patient, four minutes of artifact-free HD-EEG were selected. The electrodes located on the cheeks and neck were excluded by the analysis because they are not expected to carry valuable information about the cortical activity and they are usually heavily corrupted by artifacts. In particular, 162 channels were considered, symmetrically distributed over the scalp, 81 on the right hemisphere, 81 on the left one. The HD-EEG recordings were finally exported as a MATLAB .mat file. The algorithms for the subsequent complex network analysis (Sect. 42.2.3) were implemented in MATLAB 2016a (The MathWorks, Inc., Natick, MA, USA).

### 42.2.3 EEG-based Complex Network Analysis

The aim of the present research was to explore the ability of HD-EEG-based complex network analysis to describe the asymmetry in the brain-electrical network organization in stroke patients. A graph representation of the coupling strength between EEG channels was adopted. The  $i$ th channel represents the  $i$ th node of the graph. The weight of an edge connecting two nodes (electrodes)  $i$  and  $j$  is the *coupling strength* between the two recorded EEG signals  $x_i$  and  $x_j$ . The properties of the graph describe the behavior of the brain-electrical network. Every HD-EEG recording was segmented into 1-s non-overlapping epochs, and it was processed epoch by epoch. Thus, for every patient, 240 epochs were analyzed. PDI was recently introduced as a measure of dissimilarity between time series, and it was shown to be inversely proportional to the coupling strength between them [9, 10]. For every epoch “ $e$ ”, the  $\text{PDI}^e(x, y)$  between electrodes  $x$  and  $y$  was calculated.  $\text{PDI}^e(x, y)$  is the  $(x, y)$ -th element of the dissimilarity matrix  $\mathbf{PDI}^e$  of epoch  $e$ . Since  $\text{PDI}^e(x, y)=\text{PDI}^e(y, x)$ ,  $\mathbf{PDI}^e$  matrix is symmetrical (undirected graph). Summarizing, the adopted model is an undirected weighted graph. Once the whole HD-EEG recording of a subject is processed, the sequence of matrices  $\mathbf{PDI}^e$  is determined (with epoch  $e = 1, \dots, 240$ ). The entire sequence is then normalized so that the elements of the matrices range between 0 and 1. For every epoch  $e$ , the *connection matrix*  $\mathbf{CM}^e$  is then constructed as the complementary of the dissimilarity matrix:  $\mathbf{CM}^e = 1 - \mathbf{PDI}^e$ . Given a weighted graph, the characteristic path length ( $\lambda$ ) quantifies the integration of a network [11]. It is calculated as the mean of the shortest path length ( $d_{i,j}^w$ ) between all the possible pairs of  $n$  nodes:

$$\lambda = \frac{1}{n(n-1)} \sum_{i \neq j} d_{i,j}^w$$

The average clustering coefficient ( $CC$ ) quantifies the segregation of a network as it measures in what extent the nodes tend to group together with the nearest neighbor nodes:

$$CC^w = \frac{1}{n} \sum_{i=1}^n CC_i^w = \frac{1}{n} \sum_{i=1}^n \frac{2t_i^w}{k_i(k_i - 1)}$$

where  $CC_i$  is the clustering coefficient of node  $i$ . It depends on the number of triangles around node  $i$  ( $t_i$ ) and the maximum possible number of triangles around that node,  $k_i(k_i - 1)/2$ , where  $k_i$  is the degree of node  $i$  [12].  $CC$  and  $\lambda$  are involved in the “small-worldness” phenomenon [13]. Small-worldness is quantified by the small-world coefficient (SW) which depends on  $\lambda$  and  $CC$  as well as on  $\lambda_r$  and  $CC_r$  which are the characteristic path length and the clustering coefficient of a random graph with the same number of nodes and edges. SW is defined as:

$$SW = \frac{CC/CC_r}{\lambda/\lambda_r}$$

where  $\lambda_r = \frac{\ln(n)}{\ln(\frac{E}{n}-1)}$  and  $CC_r = \left(\frac{E}{n}\right)$ . SW is an index of balance between the local and global integration of the network, and network with a small-world organization is characterized by short path length and high clustering. For every patient and every epoch, the connection matrix  $\mathbf{CM}^e$  was analyzed by means of the *Brain Connectivity Toolbox* [12] to extract the small-worldness (SW) of the graph model (i.e., the SW of the brain-electrical network in that epoch).

#### 42.2.4 Permutation Disalignment Index

PDI was first introduced in [9]. It is based on time series projection into symbols (the so-called *motifs* [14]). PDI is a multivariate descriptor of the dissimilarity between two or more time series. It projects equally spaced samples of two or more signals into the  $m$ -dimensional embedding space (where  $m$  is the *embedding dimension* [14]), and then it estimates the randomness of the dissimilarity between the motifs. The distance between two consecutive selected samples is the *time lag*  $L$ . Given two time series  $x_i$  and  $x_j$ , the probability  $p_{x_i,x_j}(\pi_k)$  of the simultaneous occurrence of every possible motif  $\pi_k$  (with  $k = 1, \dots, m!$ ) in both  $x_i$  and  $x_j$  is estimated. PDI between  $x_i$  and  $x_j$  is defined as:

$$PDI(x_i, x_j) = \frac{1}{1-\alpha} \log \left[ \sum_{k=1}^{m!} p_{x_i,x_j}(\pi_k)^{\alpha} \right] \quad (42.1)$$

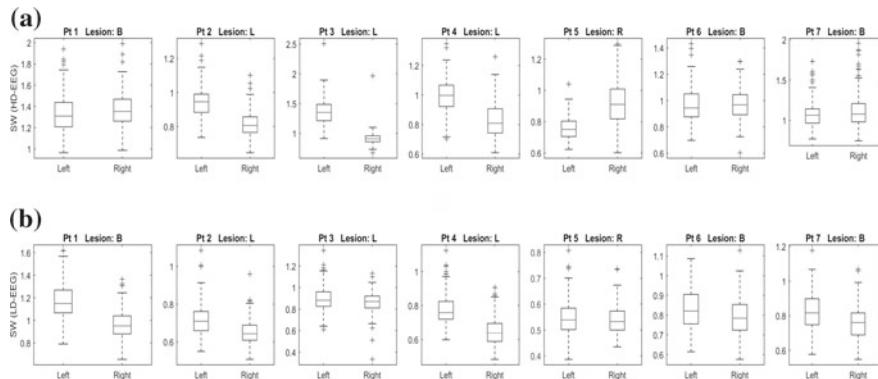
where  $p_{x_i, x_j}(\pi_k)$  is the probability that a given motif  $\pi_k$  simultaneously occurs both in  $x_i$  and  $x_j$ , and  $\alpha$  is the order of entropy according to Renyi [15]. PDI( $x_i, x_j$ ) is inversely proportional to the coupling strength between  $x_i$  and  $x_j$ , and coupled signals are indeed expected to exhibit the same motifs with increased probability [9]. In the present work,  $m = 3$  and  $L = 1$ , and  $\alpha = 2$  were selected, as detailed in [9].

## 42.3 Results

The complex network analysis described in Sect. 42.2.3 was applied to the HD-EEG recordings of the seven stroke patients. The main objective of the present work is to study the possible asymmetry in the network organization of the two hemispheres to evaluate the possible correlation between the presence of the lesion due to stroke and the brain-electrical network behavior. The secondary objective is to assess if such differences are more pronounced when the analysis is based on HD-EEG rather than on standard low-density EEG. When the analysis was based on HD-EEG, 162 electrodes (nodes) were equally divided into two subsets, the right sub-network and the left sub-network, then the graph representation of each sub-network was constructed and the SW of the two sub-networks was calculated, as detailed in Sect. 42.2.3. The same procedure was followed with low-density EEG. The 16 nodes (Fp1, Fp2, F7, F3, F4, F8, T3, C3, C4, T4, T5, P3, P4, T6, O1, O2) were divided into two subsets (right and left hemispheres), and the complex network analysis was carried out over the two sub-networks. In both cases (HD-EEG and LD-EEG), given a patient under consideration, the SW of the right and left hemisphere were estimated for every epoch. In this way, a vector **SWr** is estimated for the right hemisphere and a vector **SWl** for the left hemisphere, both with 240 elements, as many as the epochs. Figure 42.2a shows the boxplots of **SWr** and **SWl** vectors (populations) of every patient, obtained through the HD-EEG-based analysis; whereas, Fig. 42.2b refers to the results achieved through LD-EEG-based analysis. It is worth to note that the impaired hemispheres exhibited larger SW. Since a network with a small-world organization is characterized by short path length and high clustering coefficient, impaired hemispheres seem to be associated with lower randomness of the connectivity and a higher clustering, probably due to the disrupted connections.

Table 42.1 reports the statistical comparison between **SWr** and **SWl**, for each patient. In particular, since the Shapiro–Wilk [16] test showed that the two vectors had a not normal distribution, the Mann–Whitney test [17] was adopted to compare the two populations. The results show that, when estimated from the HD-EEG-based complex network model, the SW reflects the brain asymmetry. In particular, when the lesion is not bilateral, the impaired hemisphere exhibits a larger SW ( $p < 0.05$ ) than the healthy hemisphere; when the lesion is bilateral, **SWr** and **SWl** are not significantly different ( $p > 0.05$ ).

The analysis based on standard LD-EEG did not allow to find significant differences between the two hemispheres. This result endorses that HD-EEG, despite the re-



**Fig. 42.2** Boxplot of the small-worldness of the right hemisphere (**SW<sub>r</sub>**) and of the left one **SW<sub>l</sub>**, resulting from the PDI-based complex network analysis of the EEGs of the seven stroke subjects. The horizontal mark in the boxes represents the median, the upper and lower edges of the boxes represent the first and third quartile, the whiskers depict to the extreme data points that are not considered outliers. The site of the lesion is indicated as L (left hemisphere), R(right hemisphere), or B (bilateral). **a** SW resulting from the complex network analysis of HD-EEG; **b** SW resulting from the complex network analysis of standard LD-EEG;

**Table 42.1** Statistical comparison (*p* values) of the small-worldness of the two hemispheres achieved with high-density (HD) and low-density (LD) EEG

	Pt 1	Pt 2	Pt 3	Pt 4	Pt 5	Pt 6	Pt 7
SW (HD)	0.059	5.99e-45	8.58e-78	5.81e-42	1.89e-41	0.273	0.065
SW (LD)	5.81e-42	3.88e-22	0.024	9.34e-42	0.371	2.84e-05	4.04e-12

dundancy associated with the high number of EEG channels and with the volume conduction effect, allows to better describe the properties of the brain-electrical network. In the future, the study will be detailed in the various sub-bands of the EEG signals (delta, theta, alpha, beta, gamma) and will be extended to a larger number of patients. Patients will be studied longitudinally to evaluate the effects of the treatment during the patient's follow-up. In case of not bilateral impairment, the effectiveness of the treatment is expected to be linked to a balancing of the small-worldness between the two hemispheres; in case of bilateral impairment, the SW is expected to decrease globally. SW, and other network parameters, could also be used as neurofeedback in active patient rehabilitation. The EEG is indeed a totally non-invasive and well-tolerated diagnostic tool that is well suited for use in the brain-computer-based rehabilitation protocols.

## 42.4 Conclusions

Stroke is a phenomenon that disrupts the functional brain connectivity. Apparently, the brain later tries to compensate for the damage by establishing novel neural connections. In the present work, high-density EEG was used to quantify the effects of stroke on the brain network organization, indirectly, through the study of the brain-electrical connectivity. The primary objective was to study the asymmetry in the brain network organization of the two hemispheres, in the presence of a lesion due to stroke, by means of a complex network model based on HD-EEG signals. The secondary objective was to show how the HD-EEG is able to describe the asymmetry better than low-density EEG. To this purpose, a group of seven stroke patients was recruited and monitored through HD-EEG. Each HD-EEG was partitioned into 1-s non-overlapping epochs. Epoch by epoch, a complex network model was extrapolated where the nodes of the network correspond to the EEG electrodes, while the edges of the network correspond to the coupling strength between the EEG signals. The permutation disalignment index (PDI) was adopted to estimate the complementary of the coupling strength between pairs of EEG signals. To study the asymmetry in the network organization, two sub-networks were constructed, one including the electrodes of the left hemisphere and one including the electrodes of the right one. The small-worldness of the two sub-networks, left and right, was estimated epoch by epoch, for every patient. The hemisphere affected by the lesion showed a SW ( $p < 0.05$ ) larger than the healthy one. The analysis conducted using traditional EEG did not allow to detect such differences. In the future, SW, and other parameters extracted from the HD-EEG-based complex network analysis, could be used as biomarkers of the longitudinal improvement of stroke patients.

**Acknowledgements** The present work was funded by the Italian Ministry of Health, Project Code GR-2011-02351397.

## References

1. Grefkes, C., Fink, G.R.: Connectivity-based approaches in stroke and recovery of function. *Lancet Neurol.* **13**(2), 206–216 (2014)
2. Caliandro, P., Vecchio, F., Miraglia, F., Reale, G., Della Marca, G., La Torre, G., Lacidogna, G., Iacovelli, C., Padua, L., Bramanti, P., et al.: Small-world characteristics of cortical connectivity changes in acute stroke. *Neurorehabil. Neural Repair* **31**(1), 81–94 (2017)
3. Ieracitano, C., Duun-Henriksen, J., Mammone, N., La Foresta, F., Morabito, F.C.: Wavelet coherence-based clustering of EEG signals to estimate the brain connectivity in absence epileptic patients. In: *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 1297–1304. IEEE (2017)
4. Miraglia, F., Vecchio, F., Bramanti, P., Rossini, P.M.: EEG characteristics in “eyes-open” versus “eyes-closed” conditions: small-world network architecture in healthy aging and age-related brain degeneration. *Clin Neurophysiol.* **127**(2), 1261–1268 (2016)
5. Frantzidis, C.A., Vivas, A.B., Tsolaki, A., Klados, M.A., Tsolaki, M., Bamidis, P.D.: Functional disorganization of small-world brain networks in mild Alzheimer’s Disease and amnestic Mild

- Cognitive Impairment: an EEG study using Relative Wavelet Entropy (RWE). *Front Aging Neurosci.* **6**, 224 (2014)
- 6. Zeng, H., Dai, G., Kong, W., Chen, F., Wang, L.: A novel nonlinear dynamic method for stroke rehabilitation effect evaluation using EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**(12), 2488–2497 (2017)
  - 7. Liu, S., Guo, J., Meng, J., Wang, Z., Yao, Y., Yang, J., Qi, H., Ming, D.: Abnormal EEG complexity and functional connectivity of brain in patients with acute thalamic ischemic stroke. *Comput. Math. Methods Med.* **2016**, 9 (2016)
  - 8. Zappasodi, F., Olejarczyk, E., Marzetti, L., Assenza, G., Pizzella, V., Tecchio, F.: Fractal dimension of EEG activity senses neuronal impairment in acute stroke. *PLoS ONE* **9**(6), e100199 (2014)
  - 9. Mammone, N., Bonanno, L., De Salvo, S., Marino, S., Bramanti, P., Bramanti, A., Morabito, F.C.: Permutation disalignment index as an indirect, EEG-based, measure of brain connectivity in MCI and ad patients. *Int. J. Neural Syst.* <https://doi.org/10.1142/S0129065717500204>
  - 10. Mammone, N., De Salvo, S., Ieracitano, C., Marino, S., Marra, A., Corallo, F., Morabito, F.C.: A permutation disalignment index-based complex network approach to evaluate longitudinal changes in brain-electrical connectivity. *Entropy* **19**(10), 548 (2017)
  - 11. Fornito, A., Zalesky, A., Bullmore, E.: *Fundamentals of Brain Network Analysis*. Academic Press (2016)
  - 12. Rubinov, M., Sporns, O.: Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**(3), 1059–1069 (2010)
  - 13. Watts, D.J., Strogatz, S.H.: Collective dynamics of small-world networks. *Nature* **393**(6684), 440–442 (1998)
  - 14. Bandt, C., Pompe, B.: Permutation entropy: a natural complexity measure for time series. *Phys. Rev. Lett.* **88**(17), 174102 (2002)
  - 15. Renyi, A.: On measures of information and entropy. In: *Proceedings of the Fourth Berkeley Symposium on Mathematics, Statistics and Probability*, pp. 547–561 (1961)
  - 16. Shapiro, S.S., Wilk, M.B.: An analysis of variance test for normality (complete samples). *Biometrika* **52**(3, 4), 591–611 (1965)
  - 17. Mann, H.B., Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.* **18**(1), 50–60 (1947)

# Chapter 43

## Preliminary Study on Biometric Recognition Based on Drawing Tasks



Josep Lopez-Xarbau, Marcos Faundez-Zanuy  
and Manuel Garnacho-Castaño

**Abstract** In this paper, we analyze the possibility to identify people using online handwritten tasks different from the classic ones (signature and text). Our preliminary results reveal that some drawings offer a reasonably good identification rate, which is 78.5% in the case of multi-sectional vector quantization applied to classify circles and house drawing tasks. To the best of our knowledge, this is the first paper devoted to biometric recognition based on drawing tasks.

### 43.1 Introduction

Handwritten tasks can be used for biometric recognition purposes. Signature recognition is the most popular one due to its original role as authenticator. Identification accuracy is good enough too in the case of handwritten text. Even in the case of capital letters, as we checked in our previous work [8]. However, to the best of our knowledge, the possibility to identify people using handwritten drawings has not been analyzed. In this paper, we present some preliminary results about the recognition accuracies of several online acquired drawings.

It is well known that drawing test can reveal health condition of the performer, such as mild cognitive impairment/Alzheimer [6], Parkinson's disease [1], emotional states [7], and drug abuse [5]. On the other hand, is it also possible to identify people using different freehand drawings? If the answer is yes, then privacy of the user can be compromised [4]. If the answer is no, then privacy of the user is maintained. Or asked in a different way: Once a drawing test is done for health analysis, can it be used for different purposes?

---

J. Lopez-Xarbau · M. Faundez-Zanuy (✉) · M. Garnacho-Castaño  
Fundació Tecnocampus, Universitat Pompeu Fabra, Avda. Ernest Lluch 32, 08302 Mataró,  
Barcelona, Spain  
e-mail: [faundez@tecnocampus.cat](mailto:faundez@tecnocampus.cat)

## 43.2 Database

We have acquired a database of 14 users. These users performed eight tasks on the ‘training’ day, and they did them again, the following week, on the ‘test’ day. The training set consists of the following tasks:

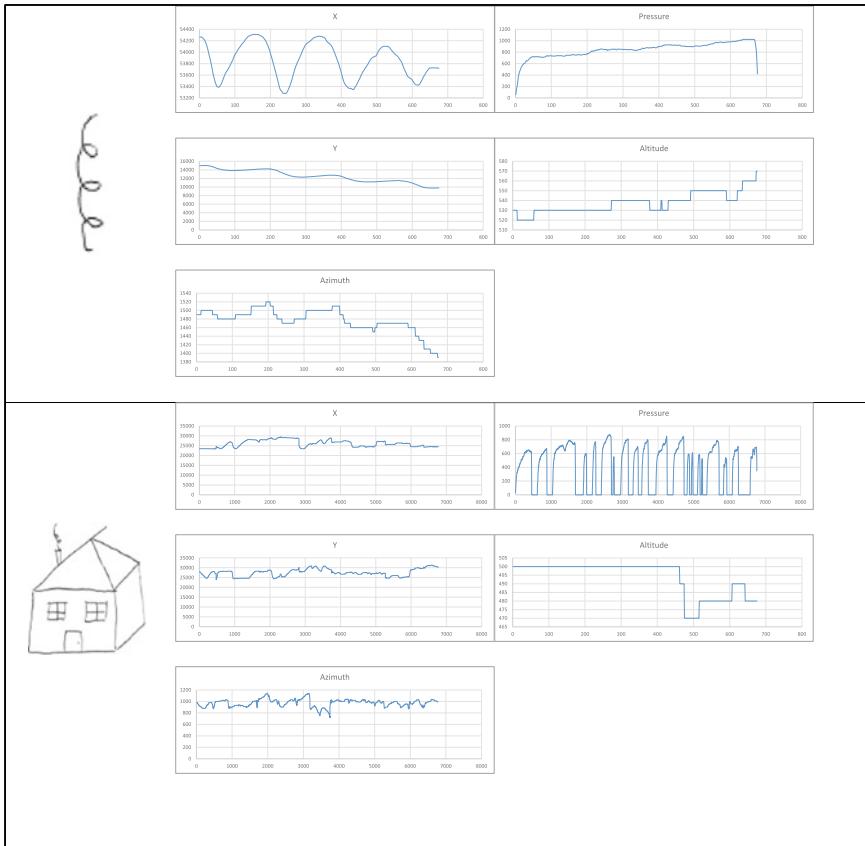
1. Pentagon drawing test
2. House drawing copied from an example
3. Archimedes spiral
4. Signature
5. Concentric circles
6. Handwritten text in capital letters
7. Spring drawing
8. Handwritten text in cursive letters.

Figure 43.1 shows an example of the acquisitions of a specific user.

Figure 43.2 shows the dynamic information acquired by the digitizing tablet Intuos Wacom when drawing the spring and the house copying tasks.

Número de usuario: U1		S10		F4		Fecha:	
<input type="checkbox"/> 1		<input type="checkbox"/> 2		<input type="checkbox"/> 3		<input type="checkbox"/> 5	(10 círculos mano dominante)
<input type="checkbox"/> 6	Escribir en mayúsculas: BIODEGRADABLE <input type="text" value="BIODEGRADABLE"/>	<input type="checkbox"/> 7		<input type="checkbox"/> 8	Firma 	<input type="checkbox"/> 9	(Mano dominante) 
Escribir en letra cursiva: A qué kilómetros de estos hermanos Wenceslao aportó su							
<input type="checkbox"/> 7							

Fig. 43.1 Example of acquired sheet from a specific user



**Fig. 43.2** Dynamic information acquired during realization of the spring drawing task

### 43.3 Recognition Algorithms

In this preliminary result paper, we have decided to apply some basic pattern recognition techniques that provided successful results in online signature recognition: dynamic time warping (DTW) [2] and multi-sectional vector quantization (MSVQ) [3].

DTW measures the similarity between two temporal sequences, which may vary in speed. DTW requires different realizations of the same task that must be done always in the same temporal order. We have applied direct comparison between testing input signals and all the user models in the database. The model that provides minimum distance reveals the identity of the user that performed the handwriting task.

MSVQ consists of splitting the signal into several parts, and one codebook is fitted to each of these parts. Then an input signal is split into the same amount of

parts, and each part is quantized with its corresponding codebook. The input vector is replaced by the nearest centroid of the codebook. Using only one section ignores the temporal evolution, while multi-section takes into consideration the order of the strokes. This process is repeated for all the models inside the database, which contains one MSVQ model for each enrolled user. The model that provides the lowest quantization distortion reveals the identity of the user that performed the input signal.

## 43.4 Experimental Results

Table 43.1 shows the identification rates obtained with DTW algorithm for each of the tasks.

Table 43.1 reveals very poor results except in the case of signature, where it is a very accurate method. This is probably due to the fact that the strokes of the drawings are not always performed in the same temporal order, and thus, the DTW fails to align the input drawing with a drawing of the same shape performed by the same user.

Figures 43.3, 43.4, 43.5, and 43.6 show the experimental results for each of the drawing tasks as function of the number of codebook sections and bits per section.

Table 43.2 summarizes the optimal values for each drawing task. In those cases, where the same optimal value is obtained with different parameter setup, we have chosen the model that implies a minimum number of total parameters.

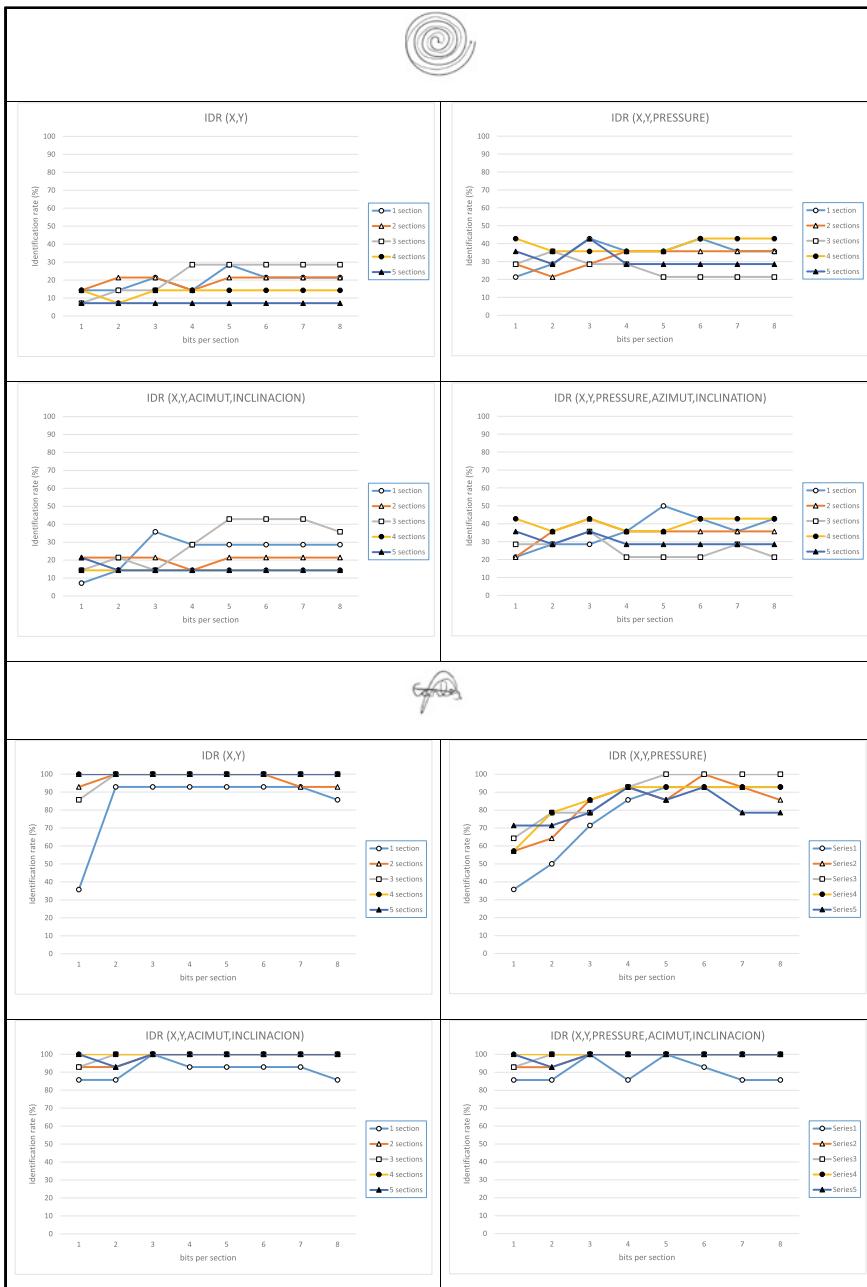
Using MSVQ, we observe some success identifying people in some drawing tasks, which are the spring drawing and house drawing. However, these results are still far from the good results obtained with the signature. However, with a more sophisticated recognition algorithm, better results could be obtained.

**Table 43.1** Identification rates with DTW for different tasks

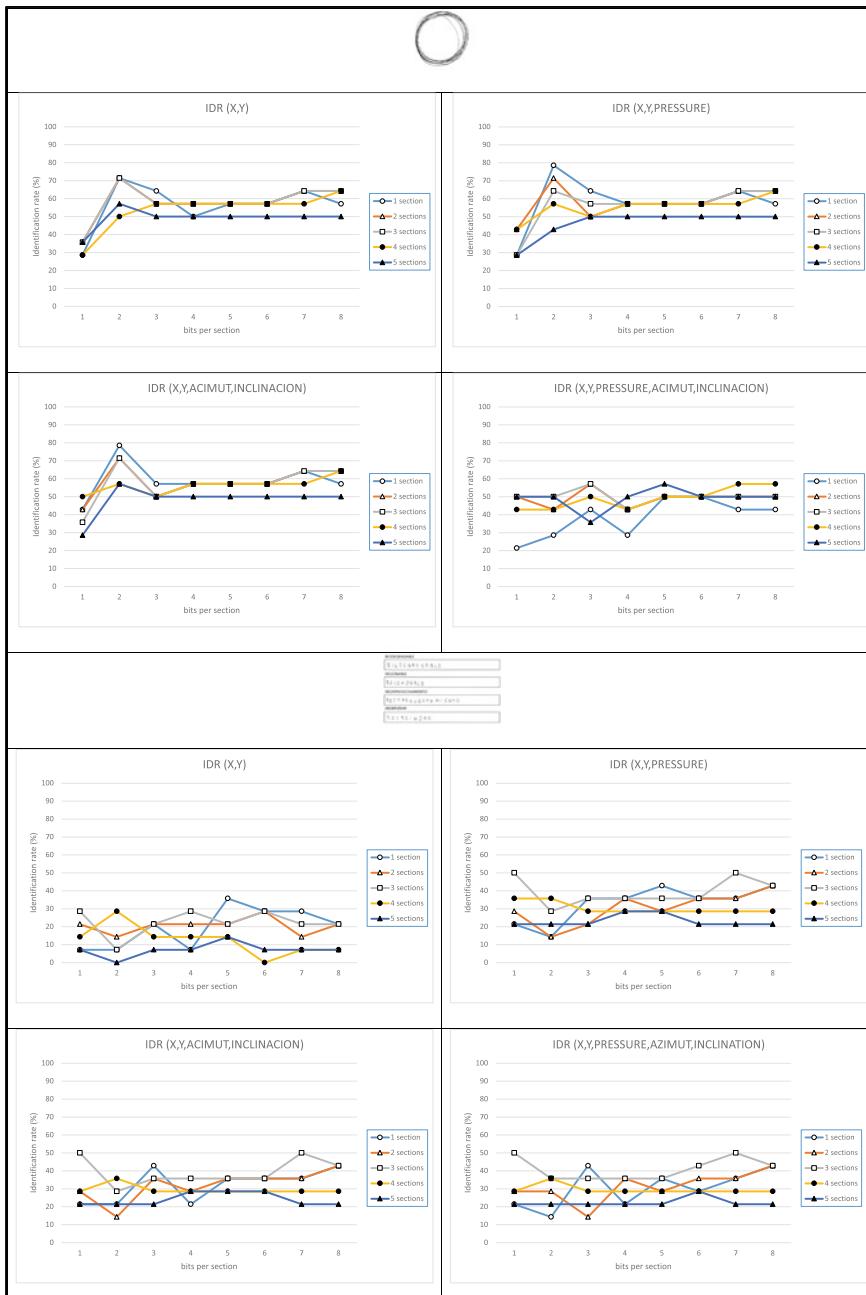
Task	DTW IDR
Signature	100
Pentagons	21.4
House	21.4
Spiral	7.1
Words	35.7
Spring	35.7
Phrase	35.7



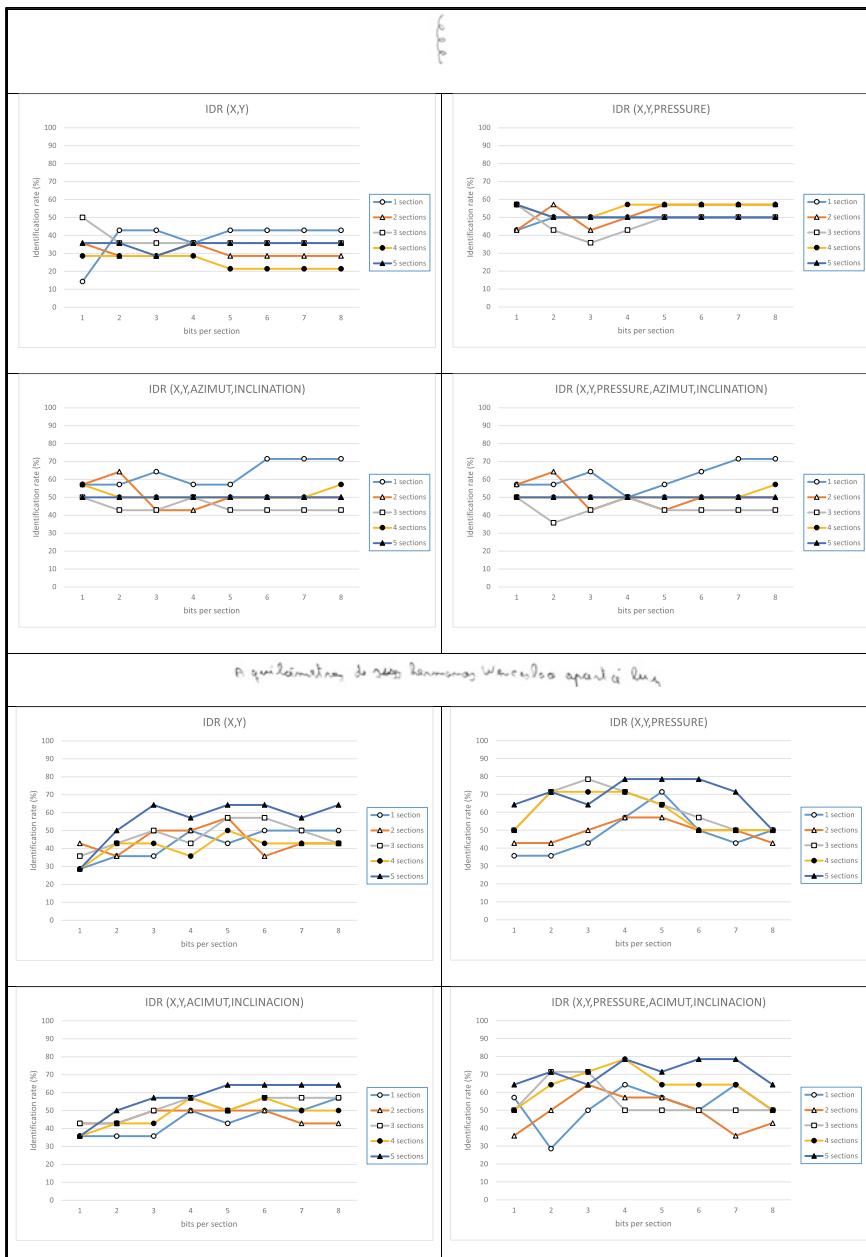
**Fig. 43.3** MSVQ results versus number of sections and bits per section from Pentagon and house drawing tasks



**Fig. 43.4** MSVQ results versus number of sections and bits per section from spiral and signature drawing tasks



**Fig. 43.5** MSVQ results versus number of sections and bits per section from circles and handwritten text in capital letters drawing tasks



**Fig. 43.6** MSVQ results versus number of sections and bits per section from spring and handwritten text in cursive letters

**Table 43.2** Identification rate (IDR) and optimal setup values for MSVQ classification

Task	MSVQ IDR	Parameter setup	Optimum value
Pentagons	50	X, Y, PRESSURE	2 sections, 5 bits per section
House	78.5	X, Y, PRESSURE, AZIMUT, INCLINATION	5 sections, 2 bits per section
Spiral	50	X, Y, PRESSURE, AZIMUT, INCLINATION	1 sections, 5 bits per section
Signature	100	X, Y	2 sections, 2 bits per section
Circles	78.5	X, Y, PRESSURE	1 sections, 2 bits per section
Capital letters	50	X, Y, PRESSURE	3 sections, 1 bits per section
Spring	71.4	X, Y, AZIMUT, INCLINATION	1 sections, 6 bits per section
Cursive letters	78.5	X, Y, PRESSURE	5 sections, 4 bits per section

## 43.5 Conclusions

Based on the experimental results of this paper, we have found that DTW and MSVQ work very fine for biometric identity recognition using online signature. On the other hand, only limited accuracy is obtained with other tasks. However, houses, circles, and cursive letters offer a 78.5% accuracy, which implies a good potential to identify people using handwritten tasks. However, the database used in this paper is quite limited, and a larger database and a statistical analysis should be done in a future work. In addition, a different recognition method could provide better results. In fact, this is what happens with the capital letters task. DTW and MSVQ offer poor results, but self-organizing maps (SOMs) showed much better results in a previous work [8].

**Acknowledgements** This work has been supported by FEDER and MEC, TEC2016-77791-C4-2-R.

## References

1. Drotar, P., Mekyska, J., Rektorova, I., Masarova, L., Smekal, Z., Faundez-Zanuy, M.: Evaluation of handwriting pressure for differential diagnosis of Parkinson's disease. *Artif. Intell. Med.* **67**, 39–46 (2016)
2. Faundez-Zanuy, M.: On-line signature recognition based on VQ-DTW. *Patt. Recogn.* **40**, 981–992 (2007). ISSN: 0031-3203
3. Faundez-Zanuy, M., Pascual-Gaspar, J.M.: Efficient on-line signature recognition based on multi-section VQ. *Patt. Anal. Appl.* **14**(1), 37–45 (2011)
4. Faundez-Zanuy, M., Mekyska, J.: Privacy of online handwriting biometrics related to biomedical analysis. In: Vielhauer, C. (Ed.) *IET in User-Centric Privacy and Security in Biometrics*, Nov 2017. Book: <https://doi.org/10.1049/pbse004e> Chapter: [https://doi.org/10.1049/pbse004e\\_ch2](https://doi.org/10.1049/pbse004e_ch2). e-ISBN: 9781785612084 Book chapter

5. Foley, R.G., Miller, L.: The effects of marijuana and alcohol usage on handwriting. *Forensic Sci. Int.* **14**(3), 159–164 (1979)
6. Garré-Olmo, J., Faundez-Zanuy, M., Lopez-de-Ipiña, K.: Kinematic and pressure features of handwriting and drawing: preliminary results between patients with mild cognitive impairment, Alzheimer disease and healthy controls. *Curr. Alzheimer Dis.* **14**(9), 960–968. <https://doi.org/10.2174/1567205014666170309120708>
7. Likforman, L., Esposito, A., Faundez-Zanuy, M.: EMOTHAW: a novel database for emotion recognition from handwriting. *IEEE Trans. Human-Machine Syst.* **47**(2), 273–284 (2017)
8. Nogueras, E.S., Faundez-Zanuy, M.: Biometric recognition using online uppercase handwritten text. *Patt. Recogn.* **45**, 128–144 (2012)

## Chapter 44

# Exploring the Relationship Between Attention and Awareness.



## Neurophenomenology of the Centroencephalic Space of Functional Integration

**Mauro N. Maldonato, Raffaele Sperandeo, Anna Esposito, Ciro Punzo  
and Silvia Dell'Orco**

**Abstract** Although there is no established theory, there is no longer any doubt about the multiplicity of the structures involved in the attentional processes. Attention is involved, in fact, in several fundamental functions: consciousness, perception, motor action, memory and so on. For several decades, the hypothesis that attention is highly variable (for extension and clarity) in terms of consciousness has been quite influential, which would range within itself in relation to its changes of state: from sleep to wakefulness, from drowsiness to twilight state of consciousness, from confusion to hyperlucidity, from dreamlike to oneiric states. More recently, other fields of considerable theoretical importance have linked attention to emotion, to affectivity or primary autonomous psychic energy or to social determinants. In this paper, we shall demonstrate how paying attention to something does not mean becoming aware of it. A series of experiments has shown that these are two distinct mental states.

---

M. N. Maldonato (✉)

Department of Neuroscience and Reproductive and Odontostomatological Sciences,  
University of Naples Federico II, Naples, Italy  
e-mail: [nelsonmauro.maldonato@unina.it](mailto:nelsonmauro.maldonato@unina.it)

R. Sperandeo

Department of Human Sciences, University of Basilicata, Potenza, Italy  
e-mail: [raffaele.sperandeo@gmail.com](mailto:raffaele.sperandeo@gmail.com)

A. Esposito

Department of Psychology, University of Campania Luigi Vanvitelli, Naples, Italy  
e-mail: [iiass.annaesp@tin.it](mailto:iiass.annaesp@tin.it)

C. Punzo

Pontifical Lateran University, Rome, Italy  
e-mail: [punzo.ciro@virgilio.it](mailto:punzo.ciro@virgilio.it)

S. Dell'Orco

Department of Humanistic Studies, University of Naples Federico II, Naples, Italy  
e-mail: [silvia.dellorco@unina.it](mailto:silvia.dellorco@unina.it)

This decoupling could represent a useful mechanism for the ability to survive that has developed during the course of evolution.

## 44.1 Modal Levels of Attention

Despite formidable advances in the neurosciences, many things on the phenomenon of attention continue to be overlooked. We know little about the way in which attention directs the mind towards objects, actions and aims while maintaining a certain level of tension for a variable time period. We know little of the way in which it has to do with awareness or why, apart from the evidence produced by Broadbent [1] only a minimal percentage of input reaches the brain, given that this comes into contact with our sensory systems with the same intensity. Finally, we know little about why most of these things, while entering into the perceptual field, do not emerge in our awareness. However, we do know that, in the selection of stimuli, attention is strongly influenced by individual expectations. It is individual expectations that often ‘decide’ which objects and experiences must reach our awareness and which must remain below minimum thresholds. Playing a decisive role is the law of interests [2], which regulates the selection of objects or topics on which attention focuses.

If it is true that we pay attention to what motivates feelings and emotions, it is equally true that attention is selective, spontaneous, intentional and directed towards an aim [3]. In Principles of Psychology [4], in the field of attention James brilliantly grasped different moments that constitute waiting, observation and reflection. Starting from its analysis, it is possible to identify the following characteristics of attention:

- A. waiting attention: this prepares for action and is always conditioned by expected events. It includes both reflex behaviour and activities of animal life, in the short or long term, and intentional conduct sustained by generic instinctual drives. For example, there are animals that seem to display great patience in highly specialized contexts: for example, the way that predators stalk their prey or the spider that awaits the signals of its prey. There are, furthermore, those situations of human awareness such as the crouching hunter, the angler waiting for a fish to bite or the sniper ready to hit a target [5]. The time frames for this kind of attention cannot be precisely determined and do not last very long. They may be limited to a precise moment or expand according to the situation, but are always characterized by a strong and alert vigilance and tension.
- B. observant attention: this indicates the distance between the subject and the surrounding scene, which the former observes down to the smallest detail. Here, the decisive factors are interest and motivation and attention is manifested in all its modal characteristics (hesitation, fatigue, drowsiness) and in terms of maintenance (driving a car on a quiet street).
- C. reflective attention: this is characterized by intense focusing on inner object. Here, the attention is expressed with an act of awareness on an inner space.

This internal observation corresponds to a psychological inquiry supported by an intended attentional process such as in yoga, finding the solution to a logical-mathematical problem, in meditation, in ascetic exercise and in many contemplative activities.

Do physiological indicators of attention exist? At a corporeal level, there is the response of orientation observable on being presented with a new stimulus. There are also signs such as pupillary dilation, peripheral vasoconstriction, cerebral vasodilation, the decrease in muscular activity and the arrest of the EEG alpha rhythm with its replacement by an irregular beta rhythm [6] Arousal describes a psycho-physiological activation level that varies along a continuum (from sleep to agitation) and is decisive in determining the efficiency of a subject's performance. If at reduced levels of activation, there is distractibility [7], and at very high levels, there is a reduction in performance efficiency due to the reduction in the degree of attention [8].

The theory of levels of activation proves to be important especially for relations between the ARAS and cortical activity. Attention and levels of activation are related aspects, but do not overlap. In fact, if the activation is a state of the organism structured along a continuum, attention is a selective function correlated with the activation levels [9]. The levels of attention depend on the activation levels of the organism, which in turn depend on the peripheral inputs and on its internal conditions. Intense inputs urge the attention to select information on the basis of its biological or psychological relevance. The basic attentional mechanisms are supported by innate or acquired elements, in resonance with the object of attention, while the ARAS represents the root of its primary neural levels that project to the cerebral cortex [10]. The attentional energy can be modulated, even neutralized, in order to direct the attention elsewhere, such as in the case of an unexpected object that has entered the sense-perception field of action. In this sense, only efficient supplementary functions can allow the analysis and processing of the subjective meaning of the object of attention. For this reason, the sensory inputs must be able to reach the cortex, even if they appear to be excluded from the field of attention. Apart from the meaning and value of each individual object, attention is strongly stimulated by the perceptual freshness and by variations in the perceptual field. In the absence of changes, attention declines, giving space to the imagination [11]. As we know, when something new captures one's attention, the reiteration of a stimulus provokes echoes of habit and even inhibition. Attention can be activated by suggestions or thoughts that have emotional or affective salience for an individual, confirmation of the role of the cortex in attentional mechanisms.

## 44.2 Procedures in Continuous Evolution

Neurobiological Research has more recently revealed the extreme complexity of the attentional phenomenon. During experiments, it has been studied from different perspectives: (a) individual variability in anticipation and its influence on reaction

speed [12]; (b) reaction times [13]; (c) the difference between reaction times on being presented with two synchronous stimuli originating from different sensory modalities [14]; (d) duration and sustainability [15]; (e) intensity and degree [16]; (f) concentration and focus [17]; and (g) the activation and the so-called arousal of attention [8]. Beyond what has been confirmed by experiment, it is known that attention renders mental states clearer, livelier and more aware. In General Psychopathology, Japers demonstrates the coexistence of related elements in attention, although they are not superimposable: (a) magnetism exercised by an object or topic; (b) clarity and vivacity of the contents, with resonances that are at times predictable but otherwise unexpected and mysterious; and (c) cognitive, emotional and motivational effects [18] of the first elements on the psychic processes. The first element is derivable only through introspection; the second calls into question critical awareness; and the third is open to objective study.

If it is true that attention increases the efficiency of the psychic processes, it is not uncommon for their acceleration (and intensification) to cause a weakening of the activities that are not focused. There needs to be a more accurate analysis of the relationship between spontaneous (or reflex) attention [19] and immediate interest, as well as of voluntary attention, directed at deferred interests in time, which can cause the feeling of an onerous effort [9]. The different levels and reaction times and both natural and cultural acts and structural constraints are relevant. It is, therefore, precisely thanks to the attention that the memory can be regarded not as a repository of static images, but as a set of procedures in continuous evolution [20]. The ability to remember is strongly connected to motor action. Every new experience represents the ability to preserve the traces of one's past life, saving from oblivion events that contribute the most to the construction of personal identity and one's personal and cultural narrative [21]. This dynamic appears all the more natural especially if one considers that attention is always attention directed towards something [22], i.e. pure intentionality. Attention, namely, fixes intent on the object on which it rests, becoming one with it in an immediate unity. It is through this intentionality that awareness gives meaning to the experience. It is precisely the dynamic nature of attention that demonstrates that psychic life is not a monolithic assembly of isolatable phenomena, but a set of relationships in continuous development. Attention focuses on an object and relates to it in its singularity. In this way, it is possible to perceive. Attention introduces things to the world. Its enhancement is a vehicle for the widening of the field of consciousness and for the acquisition of other fields of consciousness, as occurs also to contemplatives of all religions [23].

#### 44.3 Functional Asymmetries Between Attention and Awareness

When the attentional selection has reached the limits, if attention is not focused, or if in the end a part of the visual field is insufficient, perception is compromised.

But what becomes of the information that does not reach one's awareness? Is it possible for a stimulus which is ignored to be processed in any case by the cognitive system and then excluded from consciousness? Furthermore, what is the relationship between attention and awareness [24] and what is the function of the latter?

The theme of the non-awareness is extremely complex. There can be no subjective record for the obvious impossibility of accessing the processing of information [25]. It is therefore necessary to use indirect evidence, based on the measurement of independent elements from the voluntary subjective response. Below are four strategies of analysis: visual masking, priming, dichotic listening and neglect.

Visual masking [26] is a method that allows one to establish indirectly and with greater safety if the mental process that is being studied is really unaware. It consists of presenting a stimulus target followed by another stimulus that hides it, thus making it difficult, if not impossible, to identify it. This method makes it possible to directly verify whether the masking has been effective and if the target has been perceived consciously, by asking the subject to identify it. If the responses are not random, one may reasonably conclude that an analysis of the target can be attributed to non-aware processes [24].

Priming is a phenomenon that allows one to obtain indirect evidence of an unconscious analysis of information [27]. It is a phenomenon of facilitation produced by a stimulus on a further subsequent stimulus. The interesting finding that emerged in these experiments is that if the first is masked so that its identity is not recognized consciously, the effect of priming is obtained equally. Therefore, even if a word is not consciously perceived, it is still able to influence a conscious response.

Dichotic listening [28] is an experiment in which right-handed subjects are made to listen to verbal stimuli and then asked to repeat the stimulus that was the clearest. The experiment has shown that subjects tend to report more frequently what they heard in their right ear, while for non-verbal stimuli such as musical tones the opposite occurs. This behaviour seems to reflect both the specialization of the cerebral hemispheres and the structure of the auditory pathways. The two hemispheres have the tasks of processing different information and among the major differences found are those relating to linguistic-verbal skills, prevailing in the left hemisphere, and musical skills, prevailing in the right hemisphere.

Neglect [29] is a phenomenon in which a stimulus is presented in the visual hemifield of patients who have not reported the presence of any stimulus. In this case, the drawing of an object belonging to a determined category produced facilitation on a target word presented in a central position. This phenomenon shows non-aware processing both in experimental conditions on healthy subjects and in neglect patients.

The literature reports increasing evidence showing how there are psychic processes that also operate in the absence of attention with sophisticated processing of stimuli. But in what way is attention conversely linked to conscious processes [24]? According to some authors, it is attention that allows some information to reach consciousness. Attention is the access door to the awareness that can only contain a limited amount of information. In this sense, the space of functional integration [30] is a sort of selection processor and information filter. Its role is to strategically

plan and deliberately give rise to motor actions and intuitive decisions [31] and to operations of veto and prohibition for actions that are triggered automatically.

#### 44.4 Future Research Lines

Although intimately linked, attention and awareness are phenomena which do not overlap. The experimental evidence reported above shows that particular stimuli can reach one's awareness even in the absence of attention. In the last fifty years, it has been seen that the information (objects or events) selected by the attention are processed and perceived more efficiently. Attention moves in space, from one object to another, and can be directed both voluntarily and automatically, especially in the case of unexpected events. In certain cases, when there is inadequate attention, perception comes to the rescue: for example, potentially dangerous events may pass completely unnoticed if the attention is not ready and properly directed. The phenomenon defined as change blindness [32] shows precisely that it is not enough to look to see: in order to consciously perceive something the intervention of attention is always required.

In the coming years, research will need to be focussed on attention as a channel of privileged access to awareness [33], albeit not the only one. Although many cognitive processes are conducted unconsciously, awareness seems to be important in order to have strategic control of our actions and to prevent automatic processes such as reflexes from entirely controlling our behaviour.

### References

1. Broadbent, D.E.: A mechanical model for human attention and immediate memory. *Psychol. Rev.* **64**(3), 205 (1957)
2. Burnham, W.H.: Attention and interest. *Am. J. Psychol.* **19**(1), 14–18 (1908)
3. Fenske, M.J., Raymond, J.E.: Affective influences of selective attention. *Curr. Dir. Psychol. Sci.* **15**(6), 312–316 (2006)
4. James, W.: *The Principles of Psychology*, Vol. 1, p. 474. Holt, New York (1890)
5. Callieri, B., Maldonato, N.M., Di Petta, G.: *Lineamenti di psicopatologia fenomenologica*. Guida Editori, Napoli (1999)
6. Oken, B.S., Salinsky, M.C., Elsas, S.M.: Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clin. Neurophysiol.* **117**(9), 1885–1901 (2006)
7. Escera, C., Alho, K., Schröger, E., Winkler, I.W.: Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiol. Neurotol.* **5**(3–4), 151–166 (2000)
8. Eysenck, M.: *Attention and Arousal: Cognition and Performance*. Springer Science & Business Media, Berlin (2012)
9. Zomeren, A.H., Brouwer, W.H.: *Clinical neuropsychology of attention*. Oxford University Press, New York (1994)
10. Maldonato, N.M.: The ascending reticular activating system. In: *Recent Advances of Neural Network Models and Applications*, pp. 333–344. Springer, Cham (2014)
11. Oliverio, A., Maldonato, N.M.: The creative brain. In: *2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom)*, pp. 527–532. IEEE (2014)

12. Mahadevan, M.S., Bedell, H.E., Stevenson, S.B.: The influence of endogenous attention on contrast perception, contrast discrimination, and saccadic reaction time. *Vis. Res.* (2017)
13. Jipp, M.: Reaction times to consecutive automation failures: a function of working memory and sustained attention. *Hum. Factors* **58**(8), 1248–1261 (2016)
14. Shiffrin, R.M., Grantham, D.W.: Can attention be allocated to sensory modalities? *Percept. Psychophys.* **15**(3), 460–474 (1974)
15. Jones, R.G.: An applied approach to psychology of sustainability. *Ind. Organ. Psychol.* (2017)
16. Booth, J.N., Tomporowski, P.D., Boyle, J.M., Ness, A.R., Joinson, C., Leary, S.D., Reilly, J.J.: Associations between executive attention and objectively measured physical activity in adolescence: findings from ALSPAC, a UK cohort. *Mental Health Phys. Act.* **6**(3), 212–219 (2013)
17. Makovski, T., Jiang, Y.V.: Distributing versus focusing attention in visual short-term memory. *Psychon. Bull. Rev.* **14**(6), 1072–1078 (2007)
18. McGuinness, D., Pribram, K.: The neuropsychology of attention: Emotional and motivational controls. In: *The Brain and Psychology*, pp. 95–139 (1980)
19. Baroody, A.J., Li, X.: The construct and measurement of spontaneous attention to a number. *Eur. J. Develop. Psychol.* **13**(2), 170–178 (2016)
20. Rosenfield, I.: A invenção da memória: uma nova visão do cérebro. Nova Fronteira (1994)
21. Klein, S.B., Nichols, S.: Memory and the sense of personal identity. *Mind* **121**(483), 677–702 (2012)
22. Eilan, N.: Perceptual intentionality. *Attention and consciousness*. Royal Inst. Philos. Suppl. **43**, 181–202 (1998)
23. Thompson, E.: Neurophenomenology and contemplative experience. *The Oxford Handbook of Science and Religion*, pp. 226–235 (2006)
24. Robinson, P.: Attention and awareness. *Lang. Awareness Multilingualism* 1–10 (2016)
25. Yiend, J.: The effects of emotion on attention: a review of attentional processing of emotional information. *Cogn. Emot.* **24**(1), 3–47 (2010)
26. Enns, J.T., Di Lollo, V.: What's new in visual masking? *Trends Cogn. Sci.* **4**(9), 345–352 (2000)
27. Klinger, M.R., Burton, P.C., Pitts, G.S.: Mechanisms of unconscious priming: I. Response competition, not spreading activation. *J. Experim. Psychol. Learn. Memory Cogn.* **26**(2), 441 (2000)
28. Moray, N.: Attention in dichotic listening: affective cues and the influence of instructions. *Quart. J. Experim. Psychol.* **11**(1), 56–60 (1959)
29. Corbetta, M., Shulman, G.L.: Spatial neglect and attention networks. *Annu. Rev. Neurosci.* **34**, 569–599 (2011)
30. Maldonato, N. M., Oliverio, A., Esposito, A.: Neuronal symphonies: Musical improvisation and the centrencephalic space of functional integration. *World Futures* 1–20 (2017)
31. Maldonato, N.M., Dell'Orco, S., Sperandeo, R.: When intuitive decisions making, based on expertise, may deliver better results than a rational, deliberate approach. In: Esposito, A., Faundez-Zanuy, M., Morabito, F.C., Pasero, E. (eds.) *Multidisciplinary Approaches to Neural Computing*. Springer, Cham (2018)
32. Simons, D.J., Rensink, R.A.: Change blindness: past, present, and future. *Trends Cogn. Sci.* **9**(1), 16–20 (2005)
33. Maldonato, N.M., Sperandeo, R., Dell'Orco, S., Iennaco, D., Cerroni, F., Romano, P., Tripi, G.: Mind, brain and altered states of consciousness. *Acta Medica Mediterranea* **34**(2), 357–366 (2018)

# Chapter 45

## Decision-Making Styles in an Evolutionary Perspective



**Silvia Dell'Orco, Raffaele Sperandeo, Ciro Punzo, Mario Bottone,  
Anna Esposito, Antonietta M. Esposito, Vincenzo Bochicchio  
and Mauro N. Maldonato**

**Abstract** Naturalistic decision-making (NDM) investigates the cognitive strategies used by experts in making decisions in real-world contexts. Unlike studies conducted in the laboratory, the NDM paradigm is applied to real human interactions, often characterized by uncertainty, risk, complexity, time pressures and so on. In this approach, the role of experience is crucial in making possible a quick classification of decision-making situations and therefore in making an effective, rapid and prudent choice. Models of behaviour resulting from these studies represent an extraordinary resource for research and for the application of decision-making strategies in high-

---

S. Dell'Orco (✉)

Department of Humanistic Studies, University of Naples Federico II, Naples, Italy

e-mail: [silvia.dellorco@unina.it](mailto:silvia.dellorco@unina.it)

R. Sperandeo

SiPGI—Postgraduate School of Integrated Gestalt Psychotherapy, Torre Annunziata, Italy

e-mail: [raffaele.sperandeo@gmail.com](mailto:raffaele.sperandeo@gmail.com)

C. Punzo

Pontifical Lateran University, Rome, Italy

e-mail: [punzo.ciro@virgilio.it](mailto:punzo.ciro@virgilio.it)

M. Bottone · M. N. Maldonato

Department of Neuroscience and Reproductive and Odontostomatological Sciences, University of Naples Federico II, Naples, Italy

e-mail: [mario.bottone@unina.it](mailto:mario.bottone@unina.it)

M. N. Maldonato

e-mail: [nelsonmauro.maldonato@unina.it](mailto:nelsonmauro.maldonato@unina.it)

A. Esposito

Department of Psychology, University of Campania Luigi Vanvitelli, Naples, Italy

e-mail: [iiass.annaesp@tin.it](mailto:iiass.annaesp@tin.it)

A. M. Esposito

National Institute of Geophysics and Volcanology, Naples, Italy

e-mail: [antonietta.esposito@ingv.it](mailto:antonietta.esposito@ingv.it)

V. Bochicchio

Department of Humanities, University of Calabria, Rende, Italy

e-mail: [vincenzo.bochicchio@unical.it](mailto:vincenzo.bochicchio@unical.it)

risk environments. They particularly underline not only that most of the critical decisions that we take are based on our intuition, but that the ability to recognize patterns and other signals that allow us to act effectively is a natural extension of experience.

## 45.1 Introduction

The research paradigm, called naturalistic decision-making, arose from the acknowledgement of the fallacy of the patterns that prevailed during the twentieth century, which represented human decision-making out of context and is based on purely conjectural and abstract dynamics. In reality, in real-world contexts individuals do not make decisions by relying on the laws of the calculus of probability and forms of marginal utility or statistical calculations [1]. Instead of starting from models of normative reasoning, individuals use certain cognitive strategies to cope with the pressure of time, uncertainty, missing information and unclear objectives and finally select a possible course of action.

Naturalistic decision-making begins with the study of actual conduct and then probes the systematic deviations [2, 3]. One of the first sources of funding for the application of NDM to field research was provided by the Army Research Institute for the behavioural and social sciences. As a result of the accident of 1988—in which an Iran Air passenger jet, for reasons not entirely clear, was shot down by a US surface-to-air missile causing the death of all 290 passengers including 66 children—the US Navy became interested in this paradigm in real contexts, with the objective of assisting experts in making decisions under conditions of extreme uncertainty, time pressure and risk [4]. In a paper that emerged from the first conference on NDM (1989), Lipshitz [5] identified several naturalistic models developed at the same time.

The first concerns Hammond's so-called cognitive continuum theory [6] according to which various forms of cognitive processing (intuitive, analytical or simply common sense) would be situated along a continuum which puts intuitive and analytical processing at opposite ends of the spectrum. The characteristics of reasoning (e.g. control, awareness, speed) vary according to the degree and the structural characteristics of the tasks.

A second model is the skill-rule-knowledge (SRK) paradigm postulated by Rasmussen [7]. Skills, rules and knowledge represent three different ways of processing information, which are used in different ways depending on the experience of the decision-maker and familiarity they have with the situation. It is divided into three different types of behaviour that have the following characteristics:

1. Skill-based behaviour: this is behaviour related to routine and based essentially on skills learned, is characterized by a kind of unconscious reasoning and is therefore very limited in terms of cognitive commitment.
2. Rule-based behaviour: this type of behaviour is guided by rules according to which the decision-maker has to perform known tasks. In other words, the

decision-maker recognizes the situation and applies it to the most appropriate procedure. The cognitive commitment is slightly higher [7].

3. Knowledge-based behaviour: this type of behaviour is aimed at the resolution of new situations involving decision-making. Due to the lack of familiarity with the task, the decision-maker has no rules or specific procedures to which to refer and must therefore make a high cognitive commitment in order to search for a creative and effective solution.

The third decision-making model, which is based on recognition [2], will be discussed in more detail below.

## 45.2 Recognition-Primed Decision Model

Beginning from several studies conducted on the decision-making processes of experts in disparate fields (doctors, firefighters, military commanders, pilots, etc.), Klein and colleagues [2] have shown that in critical conditions—for example, those featuring tight time constraints or, in the case of indecision, serious consequences—their choices do not correspond to regulatory models and maximizing [8], but are based on a repertoire of action plans arising from experience. Thanks to Klein et al., the experts were able to quickly locate the most relevant signals, identifying objectives and plausible action plans in that particular decision-making context [9]. For example, when in an emergency situation the commander of a patrol of firemen does not decide what to do quickly and effectively [10, 11] it is highly likely that they may endanger the lives of many people. In such situations, objectives are not always clear (save people or rapidly extinguish the fire?). The information is incomplete (it is possible that one does not have a clear idea of the plan of the building on fire or the material contained therein), and the intervention procedures are not coded (sometimes, it is necessary to act creatively to discover a way to free, for example, a wounded person in a car accident) Zsambok [12, 13]. In situations like these, informative, cognitive and time constraints direct the subject towards a single objective: to make a quick assessment of the situation and choose the most feasible, timely and inexpensive alternative.

The recognition-primed model [2] works under the fundamental presupposition (in many respects counter-intuitive) that decisions by experts are not preceded by an analytical assessment of the advantages and disadvantages of each alternative decision [14]. In fact, experts begin to make decisions on the basis of an analytical assessment of the different options available and of the potential consequences. An expert in decision-making acts on the basis of four elements: achievable objectives, relevant signals, expectations and plausible courses of action. Should these elements define the familiarity and therefore the plausibility of an intuitive action, or does the situation require an inference and a conscious deliberation [1]? An individual with a high degree of expertise in a given field executes their own choices through three basic steps:

1. Experiencing the situation: they experience the situation, observing the manner in which events play out.
2. Analysing the situation: they consider the situation, quickly and intuitively, taking into account previous experience, identifying the objectives to be achieved, the most important clues to monitor, the possible developments of the situation and, finally, the possible courses of action to be undertaken.
3. Implementing the decision: they choose an action plan. The assessment of the latter step does not take place by making a comparison with other action plans, but by identifying a plausible solution and therefore by satisficing [15] (Fig. 45.1).

The RPD model is a perspicuous combination of intuition and analytical reasoning. In fact, if on the one hand the quick comparison between the actual situation and the situations already lived in the past implies a quick and intuitive processing of information, on the other hand, predicting and mentally simulating the consequences of potential decisions imply an analytically aware form of reasoning [16].

### 45.3 Cognition and Individual Differences in Decision-Making Styles

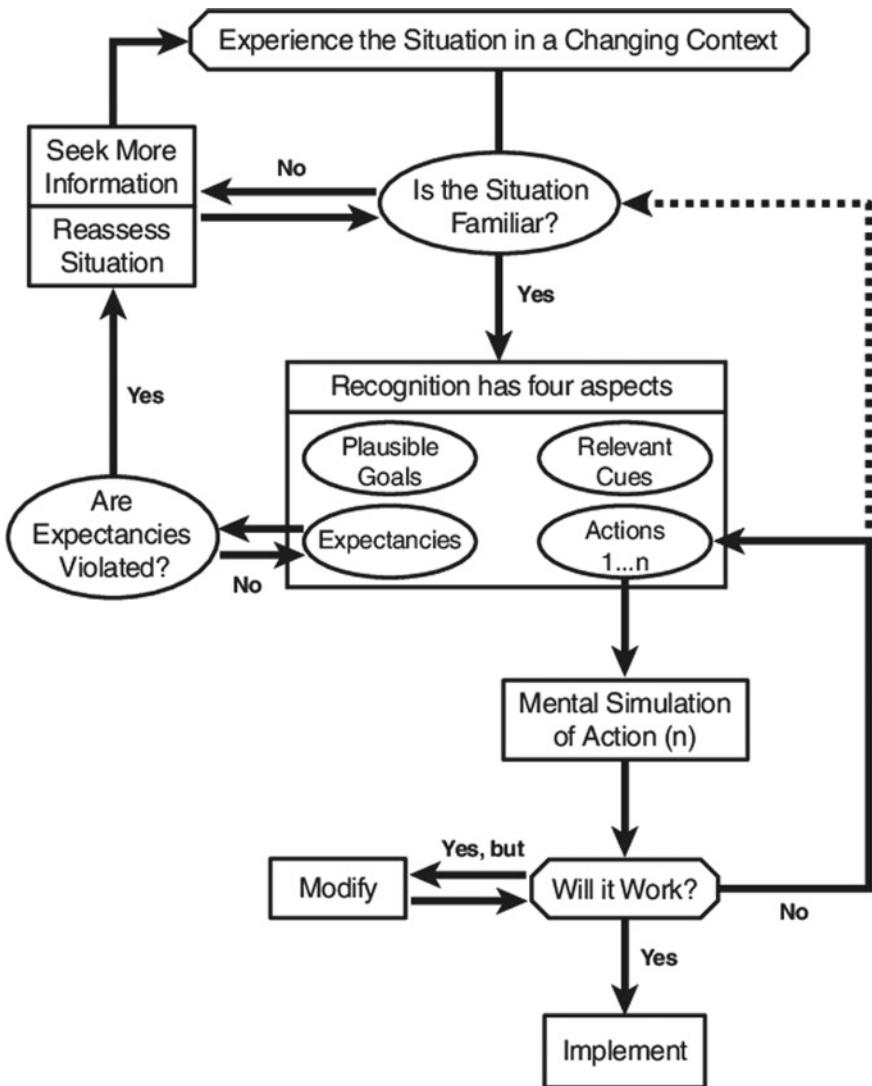
If on the one hand one's degree of expertise is an important indicator (and predictor) of decision-making behaviour, on the other hand the latter seems to be affected by other variables such as cognitive and motivational aspects, personality [17], age [18], one's general socio-economic status and more [19].

Beginning from a systematic analysis of these and other variables, some authors have hypothesized the existence of real decision-making styles that would affect the manner in which each individual acts [20, 21]. This is the propensity of a subject to adopt one particular cognitive strategy more frequently than another, depending on the context [22].

Over the years, multiple cognitive styles and decision-making processes have been studied that are often attributable to different polarities: the tendency to be independent/context dependent; the preference for a type of serial rather than holistic processing [23]; the preference for a certain structural property used in processing information; the tendency to adopt an impulsive, reflective, deliberative or intuitive style and so on [24].

One of the most prominent classifications in the literature on decision-making styles has been made by Scott and Bruce [21] and identifies five decision-making styles:

1. Rational style: the subject tends to systematically look for the information and to concentrate on the details. This search style ascertains plenty of options, the consequences of which should be processed logically.
2. Intuitive style: the subject directs their attention towards the global aspects of the situation, and the decisions are often guided by their feelings and intuitions. The



**Fig. 45.1** Recognition-primed decision-making model [3]

information is based on important details and tacit knowledge which becomes the basis for adaptive decisions.

3. Employee style: the subject seeks out tips and advice before taking a decision.
4. Avoidant style: the subject tends to postpone or to avoid making each decision.
5. Spontaneous style: the subject tends to decide as quickly as possible.

More recently, the focus of research in the context of decision-making styles has moved from description to prediction [25], particularly of the mechanisms that are

the forerunners of cognitive tasks relating complexes to decision-making [26]. It has been suggested, in fact, that if there are decision-making styles that individuals employ more frequently than others, it is equally true that these are not rigid and immutable, but tend to be flexible depending on the decision-making context or even simply depending on the emotional state of the decision-maker at the moment of choosing [27]. In fact, one decision-making style amounts to a sequence of cognitive operations that are chosen with respect to a series of factors such as the manner in which the information is proposed, the complexity of the problem in the decision-making context and the characteristics of the decision-maker [28].

As suggested by empirical and theoretical research carried out in the field of psychology of decision-making, cognitive strategies follow paths that are often different from those postulated by a rational economic choice. According to the model developed by Payne et al. [29], the decision-making process is a highly contingent and adaptive form of information processing with which individuals use heuristic decision-making strategies as a response to limited information processing capacity and the complexity of decision-making duties [30, 31]. It is a fundamental characteristic of our cognitive system, and more generally of our nervous system, and is precisely responsible for the extraordinary flexibility of decision-making strategies at our disposal. In fact, in making a choice, individuals first consider accuracy and cognitive effort not as absolute attributes linked to a strategy, but as properties contingent on a single situation [32]. This assessment—detectable both a priori (top-down) and during the processing of the decision itself (bottom-up)—can influence the choice of different decision-making strategies available [33]. The strategy selected will be the one that will lead to the making of a good decision with less effort.

In general, the most frequent strategies of simplification are commonly classified as compensatory and non-compensatory [29]. The former is based on a judgment quantity and is implemented when the attributes that describe the various decision-making alternatives are mutually commensurate according to the values of attractiveness/utility [34]. In other words, an individual chooses an alternative that has an attribute that compensates for the sacrifice that they are willing to make while not considering other appreciable attributes. Non-compensatory strategies on the other hand are used for those decision-making problems in which options and criteria cannot be measured and the attractiveness of an option with respect to a certain criterion cannot be compensated by the greater attractiveness of the same option with respect to another criterion [35]. It often happens that individuals have to mediate between accuracy and cognitive effort in selecting the strategy more suited to the demands of the task: in this case, it is necessary to have a certain flexibility in the adoption of strategies.

The decision-making process, considered as a cognitive activity with limited capacity [36], aims at satisfying multiple goals, such as minimizing the emotional burden of conflicting values between alternatives, reaching socially acceptable and justifiable decisions and making accurate decisions that maximize the benefits and minimize the cognitive effort of acquiring and processing information [37]. The latter is defined on the basis of the amount of time and type of mental operations required to implement a certain decision-making strategy. Zipf [38] proposed the principle

of least cognitive effort, according to which the selected strategy ensures minimal effort in reaching a specific desired result. The strategies involving more accurate choices are often those that involve more effort, and this indicates how the selection of strategies is the result of a compromise between the desire to take the most correct decision and that which requires least effort.

## 45.4 Conclusions and Future Perspectives of Research

The development of NDM has clearly outlined how the vast majority of decision-making strategies are not based on a logical structure [16]. It is not a coincidence that rational rules are systematically rejected on a daily basis and that the world is swarming with inconsistencies, behavioural incoherencies and incorrect calculations. Do we ascribe everything to a natural tendency of the mind, or is the description of our decision-making processes inadequate and misleading? We need a new representation of our relationships that starts from the recognition of the fundamental role played by those natural stratagems that allow us to deal effectively with situations of uncertainty. In this direction goes the idea of ecological rationality [39, 40], which emphasizes how difficulties in coping with uncertainty do not depend on inadequate neuro-physiological architecture, but from our representation of uncertainty. It is, in fact, a crucial element of a new theory of knowledge, based not only on evidence derived from natural logic, but especially by new and ongoing neuroscientific evidence. The latter shows how the subcortical structures, mediating all sensory inputs arising from the periphery, have as their highest level precisely which centrencephalic space of functional integration transfers information to the prefrontal cortex, only after subtle and complex processing. This represents a reversal of the hierarchical and pyramidal decision-making model, and as evidenced implicitly by NDM, it outlines structural-functional dynamics rooted in what is called embodied cognition. According to Lipshitz and Strauss [41], who have conducted numerous studies on heuristic strategies with which individuals face uncertainty in ‘natural’ contexts [42, 43], Kahneman and Tversky’s approach is insufficient. In fact, the standard methodology to cope with uncertainties—synthesizable with the RQP heuristic formula (reduce: reduce uncertainty by making a systematic search for information; quantify: quantify uncertainty that cannot be reduced; plug: enter the result into a formula that incorporates uncertainty as a factor in the selection of the preferred alternative)—does not provide individual environmental feedback present in the vast majority of decision-making situations [16]. Otherwise, NDM describes the way in which policy-makers face uncertainty without resorting to calculations or the RQP heuristic.

Today thanks to increasingly more accurate contributions to methods of brain imaging, it is increasingly clear that most of the ‘carriers’ that guide our behaviour are unaware and that, by reason of their actions, will advance ex post rational justification based on pure intuition [42, 44]. This framework, which interweaves the

act of choosing with biologically based research on the unconscious spheres of our relationships, represents a fascinating and complex horizon for future research.

Increasing evidence regarding adaptation [30] to unforeseen situations and, therefore, improvisation, shows that the ability to modulate our actions is extraordinarily broader than traditional theories of executive functions have so far claimed [45, 46]. It is our belief that a drastic revision of the definition of executive functions will bring new models of interpretation-making and motor behaviour that up to the present have been reduced to rigid patterns and functions.

## References

1. Zsambok, C.E., Klein, G. (eds.): *Naturalistic Decision Making*. Psychology Press, New York (2014)
2. Klein, G.A., Orasanu, J.E., Calderwood, R.E., Zsambok, C.E.: Decision making in action: models and methods. In: *This Book is an Outcome of a Workshop Held in Dayton, OH, 25–27 Sep 1989*. Ablex Publishing, New York (1993)
3. Klein, G.L.: A recognition-primed decision (RPD) model of rapid decision making. *Decision Making in Action: Models and Methods*. Ablex Publishing, New York (1993)
4. Beach, L.R., Lipshitz, R.: Why classical decision theory is an inappropriate standard for evaluating and aiding most human decision making. In: *Decision Making in Aviation*, 85 (2017)
5. Lipshitz, R.: Converging themes in the study of decision making in realistic settings. In: Klein, G.A., Orasanu, J., Calderwood, R., Zsambok, C.E. (eds.) *Decision Making in Action: Models and Methods*, pp. 103–137. Ablex, Norwood, NJ (1993)
6. Hammond, K.R., Hamm, R.M., Grassia, J., Pearson, T.: Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Trans. Syst. Man Cybern.* **17**(5), 753–770 (1987)
7. Rasmussen, J.: Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Trans. Syst. Man Cybern.* **3**, 257–266 (1983)
8. Maldonato, N.M., Dell'Orco, S.: The natural logic of action. *World Futures* **69**(3), 174–183 (2013)
9. Morton, A.: *Disasters and Dilemmas: Strategies for Real-Life Decision Making*. Wiley, Hoboken (2017)
10. Klein, G.A., Calderwood, R., Clinton-Cirocco, A.: Rapid decision making on the fire ground. In: *Proceedings of the Human Factors Society Annual Meeting* (vol. 30, no. 6, pp. 576–580). Sage, Los Angeles, CA (1986)
11. Lipshitz, R., Omodei, M., McLennan, J., Wearing, A.: What's burning? The RAWFS heuristic on the fire ground. In: *Expertise Out of Context*, pp. 97–112 (2007)
12. Zsambok, C.E.: Implications of a recognitional decision model for consumer behavior. In: *ACR North American Advances* (1993)
13. Schiebener, J., Brand, M.: Decision making under objective risk conditions—a review of cognitive and emotional correlates, strategies, feedback processing, and external influences. *Neuropsychol. Rev.* **25**(2), 171–198 (2015)
14. Maldonato, N.M., Dell'Orco, S., Sperandeo, R.: When intuitive decisions making, based on expertise, may deliver better results than a rational, deliberate approach. In: Esposito, A., Faundez-Zanuy, M., Morabito, F.C., Pasero, E. (eds.) *Multidisciplinary Approaches to Neural Computing*. Springer, Cham (2018)
15. Simon, H.A.: Theories of bounded rationality. *Decis. Organ.* **1**(1), 161–176 (1972)
16. Klein, G.: A naturalistic decision making perspective on studying intuitive decision making. *J. Appl. Res. Mem. Cogn.* **4**(3), 164–168 (2015)

17. Juanchich, M., Dewberry, C., Sirota, M., Narendran, S.: Cognitive reflection predicts real-life decision outcomes, but not over and above personality and decision-making styles. *J. Behav. Decis. Mak.* **29**(1), 52–59 (2016)
18. Mata, R., Josef, A.K., Lemaire, P.: Adaptive decision making and aging. In: *Aging and Decision Making*, pp. 105–126 (2015)
19. Lin, X., Featherman, M., Brooks, S.L., Hajli, N.: Exploring gender differences in online consumer purchase decision making: an online product presentation perspective. In: *Information Systems Frontiers*, pp. 1–15 (2018)
20. Stanovich, K.E., West, R.F.: Individual differences in rational thought. *J. Exp. Psychol. Gen.* **127**(2), 161 (1998)
21. Scott, S.G., Bruce, R.A.: Decision-making style: the development and assessment of a new measure. *Educ. Psychol. Measur.* **55**(5), 818–831 (1995)
22. Riding, R.J., Glass, A., Douglas, G.: Individual differences in thinking: cognitive and neurophysiological perspectives. *Educ. Psychol.* **13**(3–4), 267–279 (1993)
23. Pask, G.: Learning strategies, teaching strategies, and conceptual or learning style. *Learning Strategies and Learning Styles*, pp. 83–100. Springer, Boston, MA (1988)
24. Epstein, S., Pacini, R., Denes-Raj, V., Heier, H.: Individual differences in intuitive—experiential and analytical—rational thinking styles. *J. Pers. Soc. Psychol.* **71**(2), 390 (1996)
25. Maldonato, N.M., Dell'Orco, S.: The predictive brain. *World Futures* **68**(6), 381–389 (2012)
26. Bruine de Bruin, W., Parker, A.M., Fischhoff, B.: Individual differences in adult decision-making competence. *J. Pers. Soc. Psychol.* **92**(5), 938 (2007)
27. Scheibejenne, B., Von Helversen, B.: Selecting decision strategies: the differential role of affect. *Cogn. Emot.* **29**(1), 158–167 (2015)
28. Hollnagel, E.: Decisions about “what” and decisions about “how”. In: *Decision Making in Complex Environments*, pp. 37–46. CRC Press, Boca Rotan (2017)
29. Payne, J.W., Bettman, J.R., Johnson, E.J.: *The Adaptive Decision Maker*. Cambridge University Press, Cambridge (1993)
30. Maldonato, N.M., Dell'Orco, S.: How to make decisions in an uncertain world: heuristics, biases, and risk perception. *World Futures* **67**(8), 569–577 (2011)
31. Gigerenzer, G.: Towards a rational theory of heuristics. *Minds, Models and Milieux*, pp. 34–59. Palgrave Macmillan, London (2016)
32. Glöckner, A., Hilbig, B.E., Jekel, M.: What is adaptive about adaptive decision making? A parallel constraint satisfaction account. *Cognition* **133**(3), 641–666 (2014)
33. Laureiro-Martínez, D., Brusoni, S.: Cognitive flexibility and adaptive decision-making: evidence from a laboratory study of expert decision-makers. *Strateg. Manag. J.* **39**, 1031–1058 (2018)
34. Jekel, M., Glöckner, A.: How to identify strategy use and adaptive strategy selection: the crucial role of chance correction in weighted compensatory strategies. *J. Behav. Decis. Mak.* **31**, 265–279 (2016)
35. Shevchenko, Y., Bröder, A.: The effect of mood on integration of information in a multi-attribute decision task. *Acta Physiol. (Oxf.)* **185**, 136–145 (2018)
36. Maldonato, N.M.: Undecidable decisions: rationality limits and decision-making heuristics. *World Futures* **63**(1), 28–37 (2007)
37. Hogarth, R.M., Reder, M.W.: *Rational Choice: The Contrast Between Economics and Psychology*. University of Chicago Press, Chicago (1987)
38. Zipf, G.K.: *Human Behaviour and the Principle of Least-Effort*. Addison-Wesley, Reading, Cambridge MA (1949)
39. Gigerenzer, G.: The adaptive toolbox: toward a Darwinian rationality. *Nebr. Symp. Motiv.* **47**, 113–144 (2001)
40. Mousavi, S.: Ecological rationality of heuristics in psychology and economics. In: *Routledge Handbook of Behavioral Economics*, 88 (2016)
41. Lipshitz, R., Strauss, O.: Coping with uncertainty: a naturalistic decision-making analysis. *Organ. Behav. Hum. Decis. Process.* **69**(2), 149–163 (1997)

42. Maldonato, N.M., Dell'Orco, S.: Making decisions under uncertainty emotions, risk and biases. *Advances in Neural Networks: Computational and Theoretical Issues*, pp. 293–302. Springer, Cham (2015)
43. Gore, J., Ward, P.: Naturalistic decision making under uncertainty: theoretical and methodological developments—an introduction to the special section. *J. Appl. Res. Mem. Cogn.* (2018)
44. Klein, G.: The power of Intuition. Currency-Doubleday, New York, NY (2003)
45. Sperandeo, R., Picciocchi, E., Valenzano, A., Cibelli, G., Ruberto, V., Moretto, E., Monda, V., Messina, A., Dell'Orco, S., Di Sarno, A.D., Marsala, G., Polito, A.N., Longobardi, T., Maldonato, N.M.: Exploring the relationships between executive functions and personality dimensions in the light of “embodied cognition” theory: a study on a sample of 130 subjects. *Acta Medica Mediterranea* **34**(5), 1271–1279 (2018)
46. Sperandeo, R., Moretto, E., Baldo, G., Dell'Orco, S., Maldonato, N.M.: Executive functions and personality features: a circular interpretative paradigm. In: 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), pp. 000063–000066. IEEE (2017)

## Chapter 46

# The Unaware Brain: The Role of the Interconnected Modal Matrices in the Centrencephalic Space of Functional Integration



**Mauro N. Maldonato, Paolo Valerio, Raffaele Sperandeo,  
Antonietta M. Esposito, Roberto Vitelli, Cristiano Scandurra,  
Benedetta Muzii and Silvia Dell'Orco**

**Abstract** For over a century, it has been accepted that there exists a remote psychic space that influences our way of thinking, perception, decision-making and so on. This space, defined by Freud as the ‘unconscious’, embodies the psychic element that we are unaware of. It is a space that is an extension and a wider representation of the complex and sophisticated metapsychological apparatus he conceived. With respect to the conscious sphere (whose related anatomical function concerns the encephalic trunk, diencephalon and associative cortical areas), this unconscious

---

M. N. Maldonato (✉) · P. Valerio · R. Vitelli · C. Scandurra

Department of Neuroscience and Reproductive and Odontostomatological Sciences,  
University of Naples Federico II, Naples, Italy  
e-mail: [nelsonmauro.maldonato@unina.it](mailto:nelsonmauro.maldonato@unina.it)

P. Valerio

e-mail: [paolo.valerio@unina.it](mailto:paolo.valerio@unina.it)

R. Vitelli

e-mail: [rvitelli@unina.it](mailto:rvitelli@unina.it)

C. Scandurra

e-mail: [cristiano.scandurra@unina.it](mailto:cristiano.scandurra@unina.it)

R. Sperandeo

Department of Human Sciences, University of Basilicata, Potenza, Italy  
e-mail: [raffaele.sperandeo@gmail.com](mailto:raffaele.sperandeo@gmail.com)

A. M. Esposito

National Institute of Geophysics and Volcanology, Naples, Italy  
e-mail: [antonietta.esposito@ingv.it](mailto:antonietta.esposito@ingv.it)

B. Muzii

Intradepartmental Program of Clinical Psychology, AOU, University of Naples Federico II,  
Naples, Italy  
e-mail: [benedetta.muzii@gmail.com](mailto:benedetta.muzii@gmail.com)

S. Dell'Orco

Department of Humanistic Studies, University of Naples Federico II, Naples, Italy  
e-mail: [silvia.dellorco@unina.it](mailto:silvia.dellorco@unina.it)

dimension relates to the limbic lobe and specific areas of the frontal, parietal and temporal lobes. This complex neurophysiological system connects and coordinates the sensory, emotional, cognitive and behavioural systems. Its sophisticated adaptive functions, implied and unaware, allow the prefrontal cortex to transform a huge amount of information into explicit behaviour, thus affecting our behaviour in terms of executive functions, decision-making, moral judgments and so on. In this paper, we advance the hypothesis that these subcortical components constitute interconnected modal matrices that intervene under certain circumstances to respond to environmental requirements. In this space of functional integration, they act as an intermediary between the frontal cortex, the limbic system and the basal ganglia and are a key player in the planning, selection and decision to carry out appropriate actions. Due to their generativity and intramodal and extramodal connections, it is plausible to assume that they also play a role in mediation between unconscious and conscious thought.

## 46.1 Unconscious Cognitive Processing

The intricate relationship between the conscious and unconscious continues to fascinate and to confuse our understanding of mental life and relationships. For over a century, psychoanalysis attempted to explain clinical phenomena through metapsychological constructs, the cognitive sciences through computational models and neuroscience through the application of psychoanalytic categories to neural processes. Despite the enormous efforts made, many questions still remain unanswered. For example, how and to what extent is our conscious experience influenced by the dynamic unconscious? Is it possible to distinguish between the unconscious processes examined by psychoanalysis and those examined by neuroscience?

During the late 1970s and early 1980s, through the electrical stimulation of the premotor cortex, Libet [1] showed that a motor action is preceded and triggered by an electric potential generated in the brain [(“readiness potential” (RP)] lasting less than half a second. For example, if you are tapping your finger on the table, you are certain to perceive the contact in ‘real time’ [2]. In reality, this is an illusion: we become aware only about half a second later. The brain knows that we intend to act well before we become aware. Awareness comes with a fait accompli, and what remains to us is only the impression of having decided. Resuming the same experimental paradigm, Haggard [3] has shown through transcranial magnetic stimulation (TMS) that the beginning of a movement occurs with a delay of 200 ms, while motor awareness does not incur any delay. If motor awareness was an ‘upstream’ phenomenon, TMS would not lead to delays. Conversely, if it were a ‘downstream’ phenomenon, it should be delayed as much as the actual start of the movement [3]. The research by Libet and Haggard suggests that awareness is supported by a synchronisation between different brain regions and is an essential process in higher integration level [4].

Today, it is accepted that ‘lower-level’ information processing (e.g. sensorimotor reflexes) operates outside of awareness and therefore outside of the higher level processes: logical, semantic and so on. In recent decades, various studies have shown the existence of an unconscious mental activity [5]. It has been shown that stimuli can be activated unconsciously, processed at the cortical level and achieve the highest levels of representation, as in the case of patients who are blind [6, 7], who have prosopagnosia [8] or with implicit awareness in hemineglect [9].

## 46.2 The Thresholds of Subliminal Perception

Studies of brain imaging on neural correlates of subliminal perception (sensory inputs that remain under the threshold of awareness), in which subjects were exposed to conscious and unconscious visual stimuli, showed an intense activation of the posterior occipitotemporal areas that extends to fronto-parietal areas [10]. In an experiment defined as ‘stealth’, it has been seen that a word—projected on a screen for a few tenths of a centisecond followed simultaneously by an image which prevented conscious perception [11]—comes to awareness when the interval between the word and its ‘mask’ is approximately 50 ms. That time could also be lower if the word arouses emotions or greater attention [12]. Without warning, in fact, the access of information to a second fronto-parietal processing step could be inhibited. However, research has shown that it is possible to recognise an object corresponding to the masked word, even without the awareness of having seen it [13]. The masked stimuli, unlike those unmasked ones, elicit a greater occipital–temporal activity and activate distant areas of the brain in the competition for access to knowledge. Therefore, to become aware of a stimulus, it must be sufficiently intense and receive focused attention—an operation that can be thwarted by other tasks or stimuli [14].

Studies on subliminal priming [15] have shown the effects on the evaluation and interpretation of information. Even when it is not perceived as a stimulus, it influences not only actions, thoughts, feelings, learning or memory but also motor responses as a masked word or a digit, influencing the perceptual sphere, or lexical semantics. Ultimately, subliminal words activate cognitive processes associated with the meanings of words, even without the awareness of this effect [16]. These studies seem to suggest, on the one hand, that certain stimuli, albeit processed by our sensory devices, do not reach the threshold of awareness. On the other hand, that conscious processing is influenced by complex neural subcortical networks inside one centrencephalic space of functional integration [17]. In fact, if it is true that compared to stimuli overspill, high enabling is not a sufficient condition for access to awareness [18], it is equally true that the activation of the cortex is often weaker for subliminal stimuli. Recording with intracranial electrodes has shown that subliminal words perceived unconsciously can consolidate in lasting cerebral processes [19]. Furthermore, transcranial magnetic stimulation (TMS) in distinct areas can selectively inhibit subliminal activation affecting lexis decision-making and pronunciation [20]. Some studies suggest that inhibition is only possible with a conscious inhibitory control, while

others argue that stimuli perceived unconsciously can activate responses inhibited in a second moment [21].

Dehaene and his group have shown that during a primary masked activity, repetitive transcranial magnetic stimulation (rTMS) on the left premotor cortex does not influence the reaction times triggered by the subliminal presentation of prime numbers [22] and this would suggest that subliminal priming does not depend on the activation of the premotor or locomotory cortex. In the first masked task, motor control would be influenced by automatic processes mediated by basal ganglia circuits [23]. Therefore, motor inhibition may be carried out at two different levels: The first is mediated by the prefrontal cortex (PFC), through supraliminal and voluntary stimuli, and depends on the conscious detection of important signals [24]; the second is mediated by cortico-striatal structures (thalamus, the caudate nucleus and perhaps the rear parietal cortex) which may exclude the PFC [25].

### 46.3 Unconscious Affective Logic

Cognitive science research helps us to understand sophisticated unconscious processes but do not seem to be sufficient to describe the dynamics (emotional, affective, motivational and so on) that modulate our conscious life [26, 27]. The recognition of the influence that unconscious emotions and motivations have on our psychic life has contributed to a more significant representation of the functioning of our relational life [28, 29]. Functional imaging is gradually clarifying the nature of instincts, inclinations and basic emotions and their role in the life of the mind [30]. We have now finally understood that our actions are influenced by our emotional and archaic motivational systems.

Neurobiological studies on unwitting emotional-affective nodes seem to confirm the intuition that has guided psychoanalysis [31]. Individuals not only act on the basis of unconscious feelings, but feel things without knowing they feel them and decide without the awareness of deciding [32, 33]. On the other hand, subjective emotions are based on a meta-abstract representation of the physiological state of the body which involves the right frontal lobe of the insular cortex that modulates feelings, emotions and affective expressions related to the limbic system [34, 35].

A large amount of research has highlighted how unwitting emotional states can guide the decision-making process [36]. Some scholars maintain that even most ‘carriers’ that guide our behaviour are unaware and that an individual may not be able to clarify goals and motivations of their own behaviour [37]. In certain circumstances, some actions begin without awareness of the objectives to be achieved, of the reasons to act or the dynamics of decision-making with regard to what is said and done [38]. They reveal little access to processes underlying their mental states (and the connections between them) advancing ex-post rational justification evaluations of their actions based on pure intuition [39].

Since they are unprovable empirically, unconscious dynamics are deduced ex post. However, it is impossible to doubt their existence. Unconsciously acquired

schemes, formidable tools for the adaptive success of the species, should be more deeply examined in the analysis of rational behaviour. The same visceral sensations are often the best guides for the action and more effective decision-making and satisfactory of logical inferences. Several years ago, Damasio [40] analysed the behaviour of patients with medial prefrontal cortex lesions (mPFC), who with their reasoning ability intact showed severe limitations in the use of emotional affective associations—their actions, i.e. so were largely inflexible and often made the wrong decisions [41].

## 46.4 Could Desire Perhaps Be an Illusion?

From the evolutionary point of view, there are two processing modes of thought: one conscious and another unconscious [42]. These modes are highly adaptive, come into play depending on the circumstances and have different characteristics. For example, simple issues can be better addressed through rational decisions, while complex issues can be addressed through intuitive decisions [43], which have a high adaptivity due to their ability to process large amounts of information, which, however, may often be imprecise and false [44]. To confirm this, there is not only laboratory evidence, but also examples from real life. In fact, while on the one hand, the purchase of problematic products is more effective when the consumer relies on intuitive decisions; on the other hand, it is more efficient when the purchase of simple products is entrusted to conscious choices. In reality, there is no evidence of the greater effectiveness of unconscious decisions [45]. Even if we know that reflection also has negative effects on the quality of the decisions, it is nevertheless true that, in making choices, risking ‘informed deliberation’ is equal to or exceeds the efficacy of unconscious thought [46]. In general, deliberations based on logical reasoning more often than not improve the effectiveness of decisions and performance.

Several studies suggest that reasoning is a construction based on prior unconscious processes [47]. If this is true, i.e. if simple actions can be carried out by the activation of the brain before an individual is aware [2], then conscious desire (and consequently freewill) would be illusory. For the rest, the intention that precedes an action is not necessarily conscious. The unconscious processes can generate illusions both about the voluntary nature of its intentions and about its own actions. Wegner [37] has argued that a voluntary experience, an expression of the simultaneity of thought and action, is inevitably the expression of an illusory inference. A thought in itself is not the real cause of the action [48]. In support of this thesis, there are studies carried out with methods of Bayesian decision-making [49] that show how our ‘unconscious’ structures allow us to make better decisions. In general, individuals analyse features better in context by relying on their own intuitions [43].

But in what way, do intuitions, mental images and creative expressions emerge and take part in our relationships? Think of what happens, for example, in the elaboration of a speech, in musical improvisation [17] or in any artistic creation, awareness comes after [23]. In that sense, then, do we continue to speak of awareness? How do we know

why we do the things we do, why we choose the things that we choose, why we prefer things that we prefer? Is awareness really only an illusion? It is clear that the answer to this question has important consequences for our ethical and legal systems [50]. The conscious and voluntary nature of our acts has to do with matters which examine science and the law on sensitive issues regarding the imputability of responsibility and the causal relationships between the brain and moral decisions [51]. Research has not yet clarified if we are able to exert full moral responsibility and if this has universal characteristics [52]. We know that our life is marked by continuous tensions with forces that attempt to assert themselves against vetoes and that ban will [53], but as yet there is no evidence of the primacy of unconscious structures over the will. Even if at certain psychological costs, the will prevails, for the most part, over these unconscious forces and, therefore, even if we perceive constraints on our actions, our will prevails. [54]. Certainly, we do perceive constraints on our actions, but within certain limits, we also feel able to control them.

## 46.5 Conclusions and Future Explorations

The question arises again: if awareness comes after an action, what is its origin? At the turn of the twentieth century, psychoanalysis stated that the ego is not master in his own home and is governed by a hidden sovereignty [55]. This assumption has represented the cornerstone of psychoanalysis and has opened new avenues for the understanding of human nature [56]. The influence of the unconscious, its functions and its processes is implemented in every mental expression and behaviour, as well as in every relationship [57]. Most of the perception of the inner processing stimuli, like external ones, takes place outside of awareness [58]. Unconscious processes constitute the inextricable background of mental, social and individual life, therefore of conscious life. If this is true, what is awareness? Is it perhaps the effect of the area of the brain that, for evolutionary reasons, developed for other purposes? It is likely that evolution has left the element that makes flexible decisions uncertain [59]. Awareness, indeed, allows for accurate and detailed actions [60]. It allows us to act on the basis of the facts, to contribute to ongoing actions for an effective adaptation to environmental contexts [61]. It helps one recognise what is important in the ocean of available data and even establishes hierarchies for new needs, desires and purposes. In short, it is the space in which the answers to certain problems are discovered and processed by correlating different evidence and thus shaping the consequences. Nevertheless, it is a field of action that has been left undetermined by evolution [62]. It comes into play only under certain conditions. To determine what should or should not take place is an unconscious and profound processing of information. Not by chance does a stimulus emerge to the awareness only after a difficult process of selection. It is along this path of the development that man has achieved its greater adaptivity, which has transformed simple forms of cognition into higher forms of individuality.

## References

1. Libet, B.: Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behav. Brain Sci.* **8**(4), 529–539 (1985)
2. Libet, B.: *Mind Time: The Temporal Factor in Consciousness*. Harvard University Press, Cambridge, London (2004)
3. Haggard, P.: Conscious intention and motor cognition. *Trends Cogn. Sci.* **9**(6), 290–295 (2005)
4. Haggard, P., Clark, S.: Intentional action: conscious experience and neural prediction. *Conscious. Cogn.* **12**(4), 695–707 (2003)
5. Morsella, E.: The function of phenomenal states: supramodular interaction theory. *Psychol. Rev.* **112**(4), 1000–1021 (2005)
6. Zigmond, M.J., Coyle, J.T., Rowland, L.P.: *Neurobiology of brain disorders: biological basis of neurological and psychiatric disorders*. Elsevier, London (2014)
7. Faingold, C., Blumenfeld, H.: *Neuronal Networks in Brain Function, CNS Disorders, and Therapeutics*. Academic Press, London (2013)
8. Bate, S., Tree, J.J.: The definition and diagnosis of developmental prosopagnosia. *Q. J. Exp Psychol.* **70**(2), 193–200 (2017)
9. Cicerone, K.D., Langenbahn, D.M., Braden, C., Malec, J.F., Kalmar, K., Fraas, M., Azulay, J.: Evidence-based cognitive rehabilitation: updated review of the literature from 2003 through 2008. *Arch. Phys. Med. Rehabil.* **92**(4), 519–530 (2011)
10. Dehaene, S.: The error-related negativity, self-monitoring, and consciousness. *Perspect. Psychol. Sci.* **13**(2), 161–165 (2018)
11. Dehaene, S.: *Consciousness and the Brain: Deciphering How the Brain Codes our Thoughts*. Penguin, New York (2014)
12. Dehaene, S.: *Les neurones de la lecture: La nouvelle science de la lecture et de son apprentissage*. Odile Jacob, Paris (2007)
13. Dehaene, S., Jobert, A., Naccache, L., Ciuci, P., Poline, J.B., Le Bihan, D., Cohen, L.: Letter binding and invariant recognition of masked words: behavioral and neuroimaging evidence. *Psychol. Sci.* **15**(5), 307–313 (2004)
14. Cohen, M.A., Cavanagh, P., Chun, M.M., Nakayama, K.: The attentional requirements of consciousness. *Trends Cogn. Sci.* **16**(8), 411–417 (2012)
15. Jaskowski, P., Skalska, B., Verleger, R.: How the self controls its “automatic pilot” when processing subliminal information. *J. Cogn. Neurosci.* **15**(6), 911–920 (2003)
16. Verleger, R., Jaskowski, P., Aydemir, A., van der Lubbe, R.H., Groen, M.: Qualitative differences between conscious and nonconscious processing? On inverse priming induced by masked arrows. *J. Exp. Psychol.* **133**, 494–515 (2004)
17. Maldonato, M., Oliverio, A., Esposito, A.: Neuronal symphonies: musical improvisation and the centrencephalic space of functional integration. *World Futures*, 1–20 (2017)
18. Kunde, W., Kiesel, A., Hoffmann, J.: Conscious control over the content of unconscious cognition. *Cognition* **88**(2), 223–242 (2003)
19. Lingnau, A., Vorberg, D.: The time course of response inhibition in masked priming. *Percept. Psychophys.* **67**(3), 545–557 (2005)
20. Nakamura, K., Hara, N., Kouider, S., Takayama, Y., Hanajima, R., Sakai, K., Ugawa, Y.: Task-guided selection of the dual neural pathways for reading. *Neuron* **52**, 557–564 (2006)
21. Breitmeyer, B., Ogmen, H., Ramon, J., Chen, J.: Unconscious and conscious priming by forms and their parts. *Visual Cogn.* **12**(5), 720–736 (2005)
22. Dehaene, S., Brannon, E.: *Space, Time and Number in the Brain: Searching for the Foundations of Mathematical Thought*. Academic Press, London (2011)
23. Oliverio, A., Maldonato, M.: The creative brain. In: 2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom), pp. 527–532 (2014)
24. Chenery, H.J., Angwin, A.J., Copland, D.A.: The basal ganglia circuits, dopamine, and ambiguous word processing: a neurobiological account of priming studies in Parkinson’s disease. *J. Int. Neuropsychol. Soc.* **14**(3), 351–364 (2008)

25. Seiss, E., Praamstra, P.: The basal ganglia and inhibitory mechanisms in response selection: evidence from subliminal priming of motor responses in Parkinson's disease. *Brain* **127**(2), 330–339 (2004)
26. Berlin, H.A.: The neural basis of the dynamic unconscious. *Neuropsychoanalysis* **13**(1), 5–31 (2011)
27. Bargh, J.A., Chartrand, T.L.: A practical guide to priming and automaticity research. In: Reis, H., Judd, C. (eds.) *Handbook of Research Methods in Social Psychology*, pp. 253–285. Cambridge University Press, New York (2000)
28. Zellner, M.R., Watt, D.F., Solms, M., Panksepp, J.: Affective neuroscientific and neuropsychoanalytic approaches to two intractable psychiatric problems: why depression feels so bad and what addicts really want. *Neurosci. Biobehav. Rev.* **35**(9), 2000–2008 (2011)
29. Damasio, A.: *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. Vintage Book, London (2000)
30. Panksepp, J., Solms, M.: What is neuropsychoanalysis? Clinically relevant studies of the minded brain. *Trends Cogn. Sci.* **16**(1), 6–8 (2012)
31. Solms, M.: *The Feeling Brain: Selected Papers on Neuropsychoanalysis*. Karnac Books, London (2015)
32. Maldonato, N.M., Dell'Orco, S.: How to make decisions in an uncertain world: heuristics, biases, and risk perception. *World Futures* **67**(8), 569–577 (2011)
33. Sperandeo, R., Monda, V., Messina, G., Carotenuto, M., Maldonato, N.M., Moretto, E., Messina, A.: Brain functional integration: an epidemiologic study on stress-producing dissociative phenomena. *Neuropsychiatric Dis. Treat.* **14**(11), 11–19 (2018)
34. Sachs, M.E., Habibi, A., Damasio, A., Kaplan, J.T.: Decoding the neural signatures of emotions expressed through sound. *NeuroImage* **174**, 1–10 (2018)
35. LeDoux, J.E., Brown, R.: A higher-order theory of emotional consciousness. In: *Proceedings of the National Academy of Sciences*, 201619316 (2017)
36. Maldonato, N.M., Dell'Orco, S., Sperandeo, R.: When intuitive decisions making, based on expertise, may deliver better results than a rational, deliberate approach. In: Esposito, A., Faundez-Zanuy, M., Morabito, F.C., Pasero, E. (eds.) *Multidisciplinary Approaches to Neural Computing*. Series: Smart Innovation, Systems and Technologies, vol. 69, pp. 369–377. Springer, Gewerbestrasse (2017)
37. Wegner, D.M.: *The Illusion of Conscious Will*. MIT Press, Cambridge, MA (2002)
38. Maldonato, N.M., Dell'Orco, S.: Making decisions under uncertainty emotions, risk and biases. *Advances in Neural Networks: Computational and Theoretical Issues*, pp. 293–302. Springer, Cham (2015)
39. Westen, D.: The scientific status of unconscious processes: Is Freud really dead? *J. Am. Psychoanal. Assoc.* **47**(4), 1061–1106 (1999)
40. Damasio, A.R.: The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Phil. Trans. R. Soc. Lond. B* **351**(1346), 1413–1420 (1996)
41. Maldonato, N.M., Dell'Orco, S.: The natural logic of action. *World Futures* **69**(3), 174–183 (2013)
42. Maldonato, N.M.: The ascending reticular activating system. *Recent Advances of Neural Network Models and Applications*, pp. 333–344. Springer, Cham (2014)
43. Gigerenzer, G.: *Gut Feelings: The Intelligence of the Unconscious*. Penguin, London (2007)
44. Kahneman, D.: A perspective on judgment and choice: mapping bounded rationality. *Am. Psychol.* **58**(9), 697 (2003)
45. Maldonato, N.M., Dell'Orco, S.: The predictive brain. *World Futures* **68**(6), 381–389 (2012)
46. Maldonato, N.M.: From neuron to consciousness: for an experience-based neuroscience. *World Futures* **65**(2), 80–93 (2009)
47. Maldonato, N.M., Dell'Orco, S.: Toward an evolutionary theory of rationality. *World Futures* **66**(2), 103–123 (2010)
48. Maldonato, N.M.: Undecidable decisions: rationality limits and decision-making heuristics. *World Futures* **63**(1), 28–37 (2007)

49. Ellison, A.M.: An introduction to Bayesian inference for ecological research and environmental decision-making. *Ecol. Appl.* **6**(4), 1036–1046 (1996)
50. Illes, J.: *Neuroethics: Anticipating the Future*. Oxford University Press, Oxford (2017)
51. Gazzaniga, M.S.: *The Consciousness Instinct: Unraveling the Mystery of How the Brain Makes the Mind*. Farrar Straus and Giroux, New York (2018)
52. Shepherd, J.: Free will and consciousness: experimental studies. *Conscious. Cogn.* **21**(2), 915–927 (2012)
53. Evers, K.: The contribution of neuroethics to international brain research initiatives. *Nat. Rev. Neurosci.* **18**(1), 1 (2016)
54. Levy, N.: *Consciousness and Moral Responsibility*. Oxford University Press, Oxford (2014)
55. Solms, M.: The conscious id. *Neuropsychoanalysis* **15**(1), 5–19 (2013)
56. Maldonato, N.M.: The wonder of reason at the psychological roots of violence. *Advances in Culturally-Aware Intelligent Systems and in Cross-Cultural Psychological Studies*, pp. 449–459. Springer, Cham (2018)
57. Freud, S.: Appendix C to the unconscious. Stand. Edn. **14**, 159–215 (1915)
58. Merker, B.: Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. *Behav. Brain Sci.* **30**(1), 63–81 (2007)
59. Maldonato, N.M., Dell'Orco, S.: How to make decisions in an uncertain world: heuristics, biases, and risk perception. *World Futures* **67**(8), 569–577 (2011)
60. Dijksterhuis, A., Nordgren, L.F.: A theory of unconscious thought. *Perspect. Psychol. Sci.* **1**(2), 95–109 (2006)
61. Mangold, S., Hagmayer, Y.: Unconscious vs. conscious thought in causal decision making. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 33, no. 33 (2011)
62. Payne, J.W., Samper, A., Bettman, J.R., Luce, M.F.: Boundary conditions on unconscious thought in complex decision making. *Psychol. Sci.* **19**(11), 1118–1123 (2008)

# Author Index

## A

- Alahi, Alexandre, 21  
Alonso-Martinez, Carlos, 453  
Álvarez-Marquina, Agustín, 431  
Amorese, Terry, 331  
Araki, Ryosuke, 291  
Azhari, Atiqah, 395

## B

- Baldi, Mario, 201  
Ballan, Lamberto, 21  
Barbiero, Pietro, 213, 223, 281, 305  
Barzanti, Luca, 73  
Bernardo De, Andrea, 55  
Bertotti, Andrea, 281, 305  
Betti, Alessandro, 177  
Bevilacqua, Vitoantonio, 83, 257  
Bochicchio, Vincenzo, 503  
Bonanno, Lilla, 403, 475  
Bortone, Ilaria, 257  
Bottone, Mario, 503  
Bourbakis, Nikolaos, 331  
Brunetti, Antonio, 83, 257  
Buonanno, Amedeo, 201  
Buongiorno, Domenico, 257

## C

- Cabri, Alberto, 95  
Camastra, Francesco, 47  
Cascarano, Giacomo Donato, 257  
Cermenati, Luca, 167  
Ciaramella, Angelo, 119  
Cioffi, Federico, 347  
Cioffi, Valeria, 443

Ciravegna, Gabriele, 223, 281, 305

Cirrincione, Giansalvo, 213, 223, 235, 247, 281, 305

Cirrincione, Maurizio, 213, 235

Comminiello, Danilo, 11

Corazza, Marco, 145

Cordasco, Gennaro, 331

Coscia, Pasquale, 21

Costello, Rachel, 383

Cozza, Federico, 107

Cuciniello, Marialucia, 331

## D

Dagnes, Nicole, 223

Dattola, Serena, 403, 475

De Salvo, Simona, 403, 475

Dell'Orco, Silvia, 415, 443, 495, 503, 513

## E

Esposito, Anna, 3, 331, 495, 503

Esposito, Antonietta M., 55, 331, 503, 513

Esposito, Gianluca, 395

## F

Fabbricino, Irene, 415

Faundez-Zanuy, Marcos, 3, 453, 485

Ferone, Alessio, 189

Ferrara, Salvatore, 55

Ferretti, Claudio, 269

Font, Xavi, 465

Franchini, S., 313

Frontera, Patrizia, 61

Furukawa, Yujiro, 291

**G**

- Gabrieli, Giulio, 395  
 Galdi, Paola, 107  
 Garnacho-Castaño, Manuel, 485  
 Gelardi, Giuseppe, 83  
 Gennaro Di, Angelo, 371  
 Gennaro, Giovanni Di, 201  
 Gilardi, Maria Carla, 269  
 Giove, Silvio, 73  
 Giudicepietro, Flora, 55  
 Gnisci, Augusto, 347, 371  
 Gómez-Rodellar, Andrés, 431  
 Gómez-Vilda, Pedro, 431  
 Gori, Marco, 177

**H**

- Han, Changhee, 269, 291  
 Hataya, Ryuichiro, 269  
 Hayashi, Hideaki, 291

**I**

- Iennaco, Daniela, 443  
 Ieracitano, Cosimo, 61, 475

**K**

- Koutsombogera, Maria, 359, 383  
 Kumar, Rahul R., 235

**L**

- La Foresta, Fabio, 403, 475  
 Leva Di, Giuseppina, 415  
 Lo Re, G., 313  
 Loconsole, Claudio, 257  
 Lopez-Xarbau, Josep, 485

**M**

- Maffei, Luigi, 347  
 Malchiodi, Dario, 167  
 Maldonato, Mauro N., 331, 415, 443, 495, 503, 513  
 Mamnone, Nadia, 61, 403, 475  
 Maratea, Antonio, 189  
 Marcolin, Federica, 223  
 Marino, Nicola, 83  
 Marino, Silvia, 403, 475  
 Masulli, Francesco, 95  
 Masullo, Massimiliano, 347  
 Mauri, Giancarlo, 269, 291  
 Mazzola, Sergio, 83  
 Meghraoui, Djamilia, 431  
 Mekyska, Jiri, 431  
 Messina, Martina, 443  
 Midiri, M., 313  
 Militello, Carmelo, 269

**Monteriù, Andrea, 35**

- Morabito, Francesco Carlo, 3, 61, 403, 475  
 Moretto, Enrico, 415, 443  
 Morfino, Valerio, 133  
 Mosca, Lucia Luciana, 443  
 Muzii, Benedetta, 513

**N**

- Nadali, Leonardo, 145  
 Nagano, Yudai, 269  
 Nakayama, Hideki, 269, 291  
 Nardone, Davide, 119  
 Nobile, Marco S., 269

**O**

- Ospedale, Francesco, 201

**P**

- Palacios-Alonso, Daniel, 431  
 Palmieri, Francesco A.N., 21, 201  
 Panella, Massimo, 157  
 Pantó, Fabiola, 61  
 Pappalardo, Lucia, 55  
 Parpinel, Francesca, 145  
 Pascale, Aniello, 347  
 Pasero, Eros, 3, 235, 247  
 Pasqua, Gabriele, 107  
 Paul, Carles, 465  
 Paviglianiti, Annunziata, 61  
 Pavone, Luigi, 107  
 Pezzi, Alessandro, 73  
 Piccolo, Elio, 213, 223, 281  
 Piccolo, Piccolo, 305  
 Pizzi, Claudio, 145  
 Principi, Emanuele, 35  
 Punzo, Ciro, 495, 503

**R**

- Rampone, Salvatore, 133  
 Randazzo, Vincenzo, 235, 247  
 Razi, Gennaro, 47  
 Reverdy, Justine, 359  
 Rodriguez, Eloi, 465  
 Romeo, Luca, 35  
 Rovetta, Stefano, 95  
 Rundo, Leonardo, 269, 291  
 Russo, Pietro, 157

**S**

- Salerno, S., 313  
 Sarno Di, Alfonso Davide, 415  
 Savarese, Silvio, 21  
 Scandurra, Cristiano, 513  
 Scardapane, Simone, 11

Scarpiniti, Michele, 11  
Senese, Vincenzo Paolo, 347, 371  
Sergi, Ida, 347, 371  
Serra, Angela, 107  
Sperandeo, Raffaele, 415, 443, 495, 503, 513  
Squartini, Stefano, 35  
Staiano, Antonino, 119

**T**

Tagliaferri, Roberto, 107  
Tangherloni, Andrea, 269  
Terranova, M.C., 313  
Tonda, Alberto, 281  
Troncone, Alda, 331  
Trotta, Gianpaolo Francesco, 83, 257

**U**

Uncini, Aurelio, 11

**V**

Vaccarino, Francesco, 213  
Valerio, Paolo, 513  
Vesperini, Fabio, 35  
Vitabile, Salvatore, 269, 313  
Vitelli, Roberto, 513  
Vogel, Carl, 331, 359, 383

**W**

Weitschek, Emanuel, 133

**Z**

Zanaboni, Anna Maria, 167  
Zhang, Jin, 269