Quantitative Methods in the Humanities and Social Sciences

Quantitative Methods in the Humanities and Social Sciences is a book series designed to foster research-based conversation with all parts of the university campus – from buildings of ivy-covered stone to technologically savvy walls of glass. Scholarship from international researchers and the esteemed editorial board represents the far-reaching applications of computational analysis, statistical models, computer-based programs, and other quantitative methods. Methods are integrated in a dialogue that is sensitive to the broader context of humanistic study and social science research. Scholars, including among others historians, archaeologists, new media specialists, classicists and linguists, promote this interdisciplinary approach. These texts teach new methodological approaches for contemporary research. Each volume exposes readers to a particular research method. Researchers and students then benefit from exposure to subtleties of the larger project or corpus of work in which the quantitative methods come to fruition.

More information about this series at http://www.springer.com/series/11748

Arjuna Tuzzi
Editor

Tracing the Life Cycle of Ideas in the Humanities and Social Sciences

Springer

Editor
Arjuna Tuzzi
Department of Philosophy, Sociology, Education
and Applied Psychology
University of Padova
Padova, Italy

ISSN 2199-0956 ISSN 2199-0964 (electronic)
Quantitative Methods in the Humanities and Social Sciences
ISBN 978-3-319-97063-9 ISBN 978-3-319-97064-6 (eBook)
https://doi.org/10.1007/978-3-319-97064-6

Library of Congress Control Number: 2018956144

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

Some years ago, I made, together with my students, some experiments aimed to test the Piotrowski-Altmann law on textual data from newspapers. The Piotrowski-Altmann law explains and describes the dynamics of the spread of new elements in a language and the dynamics of how elements of a language disappear. The formula which represents this law is

$$p(t) = \frac{1}{1 + ae^{-bt}}$$

It can be obtained as the solution to a differential equation which describes the dynamics of language change as a function of time. Apparently, the parameter b represents the velocity of change and can be interpreted as a bunch of linguistic and extralinguistic factors. The results of these tests gave perfect support to the hypothesis on language change and showed various forms of temporal behaviour of the function. Some words were on the increase; others could be observed while they were losing momentum. A special group reached a peak within 1 day and started decreasing the next day. Of course, there was no hope to single out the individual factors which contributed to the empirical values of the parameters and thus to a detailed interpretation of our results. We were happy enough with the empirical support to the law and a catalogue of several progression forms we found and could interpret in individual cases.

When Arjuna Tuzzi told me that she was planning a project based on distant reading using a quantitative approach aimed at data on the "history of ideas" in several scientific disciplines, I was not very optimistic at a first thought. It was clear that the search of such a history of concepts was methodologically very similar to the dynamics of linguistic elements because the concepts, or ideas, as taken from texts, are found in the form of terms in texts. I remembered my impressions from the experiments with my students. The results were excellent from a pure scientific point of view but did not look useful with respect to a chance to apply them. But then I thought: "What about if someone smarter than I am turned the process the

other way round? Starting from one or two extra-linguistic factors and analysing the frequency dynamics of words or chunks found in the texts?". This was exactly the idea behind Arjuna Tuzzi's plan. And now I became enthusiastic.

A member of the scientific community has always some knowledge about his/her discipline: there are concepts, research questions, pioneers and important personalities, significant publications, debates and controversies, leading paradigms, failures and many more, which an informed colleague will be familiar with. On the other hand, no one is able to cover a discipline totally. The older a discipline, the harder a good picture on the basis of individual descriptions will be. After some decades, even a relatively young science becomes not even remotely comprehensible by a single person. Young colleagues are not yet able to gain an overview; older ones are less open to new developments. Thus, personal knowledge of a discipline is always incomplete and biased. A more complete picture can be obtained, of course, by reading as many relevant original books and articles as possible. This would become a project for decades, while the corresponding discipline keeps changing. Such a situation calls for statistics—the only method to collect reliable information in spite of fragmentary data. The project Arjuna Tuzzi was talking about suddenly seemed to provide the only possible way to achieve a "history of ideas" in several disciplines from texts and other data sources.

Now, I am tracking the project with rapt attention.

University of Trier
Trier, Germany

Reinhard Köhler

Contents

Contributors

Michele A. Cortelazzo University of Padova, Padova, Italy

Pierdaniele Giaretta University of Padova, Padova, Italy

Giuseppe Giordan University of Padova, Padova, Italy

Stefano Ondelli University of Trieste, Trieste, Italy

Pasquale Pavone Università degli Studi di Modena e Reggio Emilia, Modena, Italy

Valentina Rizzoli University of Padova, Padova, Italy

Chantal Saint-Blancat University of Padova, Padova, Italy

Stefano Sbalchiero University of Padova, Padova, Italy

Giuseppe Spolaore University of Padova, Padova, Italy

Matilde Trevisani University of Trieste, Trieste, Italy

Arjuna Tuzzi Department of Philosophy, Sociology, Education and Applied Psychology, University of Padova, Padova, Italy

Giovanni Urraci University "Ca' Foscari" Venice, Venice, Italy

Abbreviations

AJS	American Journal of Sociology
ASA	American Sociological Association
ASR	American Sociological Review
ATD	Analysis of textual data
CA	Correspondence analysis
CC	Curve clustering
ECU	Elementary context units
ETD	Emerging topic detection
EASP	European Association of Social Psychology
EJSP	European Journal of Social Psychology
EDA	Exploratory data analysis
ETD	Emerging topic detection
FD	Functional data
FDA	Functional data analysis
FPCA	Functional principal component analysis
GCV	Generalized cross-validation
HDP	Hierarchical Dirichlet process
IE	Information extraction
IR	Information retrieval
JASA	Journal of the American Statistical Association
JPSP	Journal of Personality and Social Psychology
KWIC	Keyword in Context
KBS	Knowledge-based system
LNRE	Large number of rare events
LDA	Latent Dirichlet allocation
LSI	Latent semantic indexing
ML	Machine learning
MWE	Multiword expression
MI	Mutual information
NLP	Natural language processing
POS	Part-of-speech

PLSA	Probabilistic latent semantic analysis
PASA	Publications of the American Statistical Association
QASA	Quarterly Publications of the American Statistical Association
RE	Regular expression
RMS	Root mean square
SVD	Singular value decomposition
TM	Text mining
TDT	Topic Detection and Tracking
WSD	Word sense disambiguation