Using machine learning in aging neuroscience

By

Christian Johannes Gölz

A thesis presented for the degree of
Doctor rerum naturalium
(Dr. rer. nat)

Paderborn University
Faculty of natural sciences
2022

Contents

Acknowledgement

Abstract

Aim: Apply data science methods to questions in aging Neuroscience
Methods: Supervised and unsupervised methods in different settings
Results: Novel Data Driven insights
Coclusion: ML rocks!

Figures

Tables

List of Abbreviations

Publications and other scientific contributions

Chapter 1

Introduction

- ML as the next frontier in science
- Open questions in aging neuroscience
- What can ML tell us?
- Age related changes occur at different scales and are manifestet at several levels.
- There is a wide variety in how this changes occur
- Changes are e.g. neural dedifferentiation and compensatory mechanisms (see Reuter Lorenz et al. 2010) and are noticable brain network level and dynamics
- NOTE: Check what EEG studies said about this...
- The idea is to model these changes with tools from datascience to answer questions in aging neuroscience
- First study is about detecting dedifferentiated and compensatory mechanisms with EEG
- Tools used are DMD and Machine learning
- Main idea: Study classification performance as proxy for age related changes in different motor control tasks
- Expertise as possible way of builing a reserve:
- Higher individuality

- Dynamics of dedifferentiation and how do they relate to fitness
- Basic for targeted interventions
- How much and what (relate to Julia)
- Background of ML
- ML as tool
- novel insights
- Problem: Data is multidimensional and we have often limited data
- Solution: Use DMD to reduce Complexity and ”model” evolution of signal
- Dynamic Mode Decompsition
- DMD extracts coupled spatio-temporal modes and is able to kind of model the evolution of the signal
- Backgrouund + Papers
- Mathematical Formulation
- What can ML tell us?
- ML applied in aging Neuroscience
- Formulating Aims and goals
- Formulation expectred outcomes

Chapter 2

Theoretical Background

2.1 A data driven way to study the aging brain

"If our goal as a field is to use data to solve problems, then we need to move away from exclusive dependence on data models and adopt a more diverse set of tools" [1]

The complex interplay of the internal and external factors leads to the dynamics of brain network changes over the lifetime. However, this is a prerequisite to differentiate healthy from pathological changes and to develop and verify treatments as well as targeted interventions. Besides classical statistical approaches tools from data science, i.e., machine learning, provide new opportunities to understand multidimensional data in a naturalistic way. Rather than assuming that data is generated by a given stochastic process an algorithmic view treats data mechanisms as unknown [1]. This view is mainly used in the field of data science, which emerged as an independent scientific discipline at the beginning of the 21st century and has adopted concepts from computer science and statistics, especially from the field of machine learning [10]. In chapter 2.1 machine learning and its usage in neuroscience will be introduced. Chapter 2.2 gives a broad overview of age related changes of the brain and summerizes use cases of machine learning in aging neuroscience.

2.2 Machine learning

Machine learning emerged in the 1950s as a subbranch of artificial intelligence to enable computers to learn without being explicitly programmed [11]. It is defined by algorithms that automatically extract patterns and trends from data or in other words learn from data [4]. With big datasets being openly available and availability of computing power, machine learning is the basis of many advances in science and engineering including public transportation, e.g., autonomous vehicles [6], the medical sector, e.g. diagnostic imaging [7] or basic science, e.g. understanding protein structure [5]. With possibility for the automatic analysis of complex multimodal data machine learning bears great potential for science [2].

2.2.1 Categories of machine learning

Machine learning can be subdivided into three categories, supervised, unsupervised and reinforcement learning. In supervised machine learning the goal is to learn a model representing the relationship between data and associated information or description, a so-called label. This model can than be used to predict the label of new data that not have been used during model creation. If the labels are categorical, this process is called classification; for continuous labels, the term is regression. The goal of unsupervised machine learning is to find hidden structure in data without taking into account associated labels. This could be grouping similar data points, i.e. clustering, or uncovering a meaningful low dimensional representation of the data, i.e. dimensionality reduction. Rather than a specific label in reinforcement learning the goal is to learn optimal actions to solve a certain problem by maximizing the reward linked to that action.

2.2.2 Applications in Neuroscience

Aiming at building an artificial general intelligence machine learning algorithms and technology were inspired by the working principle of the brain leading to advances of neural networks [8] and neuromorphic computing [3]. On the other side machine learning offers great tools for a wide array of problems addressed in a variety of areas of neuroscience.

- Clinical: identify disease, develop biomarkers, characterize patients, epilepsy detection/prediction - Basic: Understand working principle of processing, e.g. visual system, working memory - Cognitive: Identify brain states and study brain behavior interaction

Summary: - Solving engineering problems as well as understanding brain processing - Investigate high dimensional representations with classification/regression/model selection - Uncover underlying processes with dimensionality reduction

2.3 Age related reorganization of the brain

Age related reorganization processes are detectable at the whole body. This is underpinned by multiple interacting biological systems operating on several spatial and temporal scales contributing to the complexity of the phenomenon [9]. At the behavioral level these processes are noticeable in changes in cognitive, motor and sensory functioning [QUELLE]. Aging is one of the biggest risk factors for neurodegenerative diseases such as dementia, including Alzheimer's disease, as well as Parkinson's disease making the brain as one of the target systems to study. Patterns of reorganization of the brain are highly individual as they are subject to genetic and environmental influences [QUELLEN]. At the same time, however, overarching, generalizable patterns can be detected [QUELLE]
On a structural level aging has been associated with a reduction in gray matter with an onset early in life

2.4

The complex interplay of the aforementioned factors leading to the dynamics of brain network changes over the lifetime perspective is not fully understood. However, this is a prerequisite to differentiate healthy from pathological changes and to develop and verify treatments as well as targeted interventions. This requires uncomplicated, easy-to-use, and cost-effective methods and novel analyses to quantify changes in brain organization. Several methods are available to study the brains' structure and function including functional and structural Magnetic Resonance Imaging (s/fMRI) as well as Magnetoencephalography (MEG) and Electroencephalography (EEG). In addition to cost-intensive imaging techniques such as fMRI, EEG, which meets these criteria, is particularly suitable. Moreover, EEG measures neuronal activity directly with high temporal resolution, which would allow us to gain new insights into the reorganization of brain networks over the lifespan in health and disease. However, it is unclear how the changes described above are reflected in electrophysiological markers. Furthermore, EEG signals are temporally and spatially highly dimensional and have a low signal-to-noise ratio, which makes the detection and visualization of brain networks and their dynamics difficult and requires advanced signal analysis methods. Advanced methods such as methods from the field of artificial intelligence with machine learning, as a subbranch, are of special interest in this context. Methods from supervised and unsupervised machine learning are possible candidates. In unsupervised machine learning, the goal is to find structure in the data. This includes methods for dimensionality reduction and clustering. Dimensionality reduction for example allows us to describe the structure of high dimensional data in fewer properties [20]. Two common methods in the analysis of neurophysiological data are the principal component analysis (PCA) and independent component analysis (ICA). These allow the detection of spatial patterns in the data that represent the underlying network characteristics of neurophysiological data [8]. In addition, with dynamic mode decomposition (DMD), Brunton, Johnson, et al. [21] apply for the first time a method to electrophysiological data that allows us to map both the spatial and temporal structure of the network structure of neurophysiological data. In supervised machine learning, models are created that can predict a certain outcome based on input data. This method is used to detect neuronal representations of the environment or certain behaviors as well as group memberships and to identify relevant markers [22]. In the context of lifespan changes, complex brain network behavior based on EEG data could be extracted and visualized using dimensionality reduction. Supervised machine learning methods could be used to detect representations of the environment and behavior and to draw conclusions about the differentiation of brain networks. Automatic detection of group membership could further provide new predictors of nervous system states.

Chapter 3

Aims and scope

Chapter 4

General methodology

Chapter 5

Publications

5.1 Paper 1

5.2 Paper 2

5.3 Paper 3

5.4 Paper 4

Chapter 6

General discussion

Chapter 7

Bibliography

Bibliography

[1] Leo Breiman. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16(3):199–231, 2001.

[2] Bingni W. Brunton and Michael Beyeler. Data-driven models in human neuroscience and neuroengineering. *Current Opinion in Neurobiology*, 58:21–29, 2019. Computational Neuroscience.

[3] Chanyeol Choi, Hyunseok Kim, Ji-Hoon Kang, Min-Kyu Song, Hanwool Yeon, Celesta S Chang, Jun Min Suh, Jiho Shin, Kuangye Lu, Bo-In Park, et al. Reconfigurable heterogeneous integration using stackable chips with embedded artificial intelligence. *Nature Electronics*, 5(6):386–393, 2022.

[4] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction, Second Edition*. Springer-Verlag, New York, second edition, 2009.

[5] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, Aug 2021.

[6] John J Leonard, David A Mindell, and Erik L Stayton. Autonomous vehicles, mobility, and employment policy: the roads ahead. *Massachusetts Inst. Technol., Cambridge, MA, Rep. RB02-2020*, 2020.

[7] Yuan Liu, Ayush Jain, Clara Eng, David H. Way, Kang Lee, Peggy Bui, Kimberly Kanada, Guilherme de Oliveira Marinho, Jessica Gallegos, Sara Gabriele, Vishakha Gupta, Nalini Singh, Vivek Natarajan, Rainer Hofmann-Wellenhof, Greg S. Corrado, Lily H. Peng, Dale R. Webster, Dennis Ai, Susan J. Huang, Yun Liu, R. Carter Dunn, and David Coz. A deep learning system for differential diagnosis of skin diseases. *Nature Medicine*, 26(6):900–908, Jun 2020.

[8] Tom Macpherson, Anne Churchland, Terry Sejnowski, James DiCarlo, Yukiyasu Kamitani, Hidehiko Takahashi, and Takatoshi Hikida. Natural and artificial intelligence: A brief introduction to the interplay between ai and neuroscience research. *Neural Networks*, 144:603–613, 2021.

[9] Kathleen M. Mooney, Amy E. Morgan, and Mark T. Mc Auley. Aging and computational systems biology. *WIREs Systems Biology and Medicine*, 8(2):123–139, 2016.

[10] Erik Lee Nylen and Pascal Wallisch. *Neural Data Science*. Academic Press, 2017.

[11] A. L. Samuel. Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3):210–229, 1959.

Chapter 8

Statutory Declaration