

# Abstract

Dipl.-Psych. Michael Hanke:

*Advancing the understanding of brain function with multivariate pattern analysis*

Decoding patterns of neural activity onto cognitive states is one of the central goals of functional brain imaging. Standard univariate fMRI analysis methods, which correlate cognitive and perceptual function with the blood oxygenation-level dependent (BOLD) signal, have proven successful in identifying anatomical regions based on signal increases during cognitive and perceptual tasks. Recently, researchers have begun to explore new multivariate techniques that have proven to be more flexible, more reliable, and more sensitive than standard univariate analysis. Drawing on the field of statistical learning theory, these new multivariate pattern analysis (MVPA) techniques possess explanatory power that could provide new insights into the functional properties of the brain.

However, unlike the wealth of software packages for univariate analyses, there are few packages that facilitate multivariate pattern classification analyses of fMRI data. This in turn prevents the adoption of these methods by a large number of research groups to fully assess their potential with respect to cognitive neuroscience research. Here, a novel, Python-based, cross-platform, and open-source software framework, called PyMVPA, for the application of multivariate pattern analysis techniques to fMRI datasets is introduced. PyMVPA makes use of Python's ability to access libraries written in a large variety of programming languages and computing environments to interface with the wealth of existing machine-learning packages. The framework is presented in this thesis, and illustrative examples on its usage, features, and programmability are provided.

In addition, this thesis provides an overview of promising strategies for the application of MVPA to neuroimaging datasets. While various possibilities are reviewed based on previously published studies, the primary focus lies on the *sensitivity analysis* technique that is shown to provide interesting additional information which are readily available as part of any typical MVPA-based study. Moreover, this technique is eminently suited for modality-independent data analysis, a feature that is demonstrated by an example of a uniform analysis of datasets from four different brain imaging domains.

The thesis concludes with a discussion about the challenges that have to be faced to establish MVPA as a standard analysis procedure, including the statistical evaluation of results, as well as potential pitfalls in their interpretation.

# Zusammenfassung

Dipl.-Psych. Michael Hanke:

*Advancing the understanding of brain function with multivariate pattern analysis*

Die Entschlüsselung neuronaler Aktivitätsmuster und ihre Zuordnung zu kognitiven Prozessen ist eines der Hauptziele des Einsatzes von bildgebenden Verfahren in den kognitiven Neurowissenschaften. In der Vergangenheit ist es gelungen, durch die Anwendung klassischer, univariater Analysemethoden bei der Auswertung von Daten aus der funktionellen Magnetresonanztomographie (fMRT), eine Vielzahl von spezialisierten Hirnarealen zu lokalisieren. Diese Verfahren nutzen bereits seit langem verfügbare Erkenntnisse über das zeitliche Verhalten des neurovaskulären Systems in Reaktion auf erhöhte neuronale Aktivität, um daraus modellhafte Signalverläufe der hirnhysiologischen Antwort auf eine gegebene Stimulation zu gewinnen. Diese Modelle werden anschließend mit den tatsächlich gemessenen Daten verglichen, um bei hinreichender Ähnlichkeit auf eine Assoziation eines bestimmten Hirnareals mit den, für die jeweilige Aufgabe nötigen, kognitiven Verarbeitungsprozessen schließen zu können.

Erst kürzlich haben Wissenschaftler damit begonnen, die Eignung von neuen multivariaten Analysetechniken zu evaluieren, die ohne besondere a priori Modellannahmen auskommen. In ersten Studien erweisen sich diese Methoden als deutlich flexibler, reliabler und vor allem wesentlich empfindlicher bei der Identifizierung der kognitiven Korrelate im fMRT-Signal. Die fMRT-Datenanalyse mit diesen alternativen Verfahren aus der statistischen Lerntheorie hat das Potential, unser Wissen über die Funktionsweise des Gehirns signifikant zu erweitern.

Jedoch stehen einer breiten Anwendung und Evaluation in den kognitiven Neurowissenschaften einige Probleme entgegen. Das drängendste ist derzeit wohl die mangelnde Verfügbarkeit von Analysesoftware, die auf neurowissenschaftliche Datensätze und insbesondere fMRT-Daten spezialisiert ist. Dem gegenüber stehen jedoch eine Fülle von verfügbaren Algorithmen aus dem Gebiet der wissenschaftlichen Forschung zum maschinellen Lernen (ML).

Ein Hauptaspekt dieser Dissertation ist die Vorstellung von *PyMVPA*, einer neuen Softwareumgebung für die Analyse von neurowissenschaftlichen Datensätzen mit den Methoden der multivariaten Musterklassifikation. PyMVPA ist eine Python-basierte, plattform-unabhängige, Open-Source Softwarelösung, die es erlaubt, die verfügbaren ML-Algorithmen in höchst flexibler Weise auf Datensätze verschiedener Herkunft anzuwenden. Besonderes Augenmerk bei der Entwicklung von PyMVPA lag dabei auf die Wiederverwendbarkeit von bereits existierenden Implementierungen, um durch eine vergrößerte Nutzerbasis die Entwicklung beziehungsweise Wartung von qualitativ hochwertiger Software zu begünstigen. Diese Dissertati-

onsschrift umfasst sowohl eine Darstellung des allgemeinen Aufbaus von PyMVPA, wie auch eine Reihe von illustrativen Beispielen, die gleichzeitig die konzeptuelle Überlegenheit, wie auch die praktische Einfachheit verdeutlichen.

Da sich die Anwendung dieser neuen Methoden substantiell von den klassischen Verfahren unterscheidet, liegt der zweite Schwerpunkt dieser Dissertation auf einem Überblick möglicher Analysestrategien. Dabei wird zunächst eine Reihe von Prozeduren dargestellt, die sich bereits in der existierenden Literatur als erfolgreich erwiesen haben. Das Hauptaugenmerk liegt jedoch auf der Darstellung der sogenannten *Sensitivitätsanalyse*. Dabei handelt es sich um ein Verfahren, das die Analyse der durch die ML-Algorithmen extrahierten Modellparameter zu Gegenstand hat. Obwohl die zugrunde liegenden Informationen in vielen Studien zur Verfügung standen, fanden diese bisher nur zögerlich Eingang in die Analyse- und Interpretationslogik. Die vorliegende Arbeit demonstriert, welches Potential in einer genaueren Betrachtung der Ergebnisse liegen kann und führt exemplarisch einige interessante Befunde auf. Insbesondere soll erwähnt werden, dass die Sensitivitätsanalyse gleichermaßen auf eine Vielzahl unterschiedlicher neurowissenschaftlicher Datenmodalitäten anwendbar ist und fruchtbare Ergebnisse liefern kann. Diese Eigenschaft wird hier durch eine Beispielanalyse vorgeführt, die eine identische Prozedur (identisch bis hin zur Ebene des Quellcodes) auf vier unterschiedliche Datenmodalitäten anwendet. Dabei kommen Datensätze der funktionellen Magnetresonanztomographie, der Elektroenzephalografie, der Magnetenzephalographie und schließlich auch von extrazellulären Ableitungen zur Verwendung. Sämtliche dargestellten Auswertungen wurden dabei vollständig in PyMVPA durchgeführt.

Den Abschluß der Arbeit bildet eine Diskussion der noch offenen Fragen und Herausforderungen bezüglich der Etablierung von multivariaten Musterklassifikationsverfahren als Standard-Analysemethoden für neurowissenschaftliche Daten. Diese umfaßt sowohl Aspekte der inferenzstatistischen Absicherungen von Ergebnissen, wie auch deren Interpretation vor dem Hintergrund einer konkreten Forschungsfragestellung. Darüber hinaus wird sowohl ein Ausblick auf die zukünftige Weiterentwicklung von PyMVPA selbst gegeben, als auch mögliche weitere verwandte Verfahren aufgeführt, die in der Lage sind, zusätzliche interessante Aspekte in der funktionellen Struktur von Hirnaktivitätsmustern zu erfassen.