
Praktikumsbericht

Timotheus Berg
372964
Mathematik

Betreuer: John Iversen
Institution: Swartz Center for Computational Neuroscience, UCSD

05.07.2022

Inhaltsverzeichnis

1	Das Unternehmen	3
2	Beschreibung meiner Tätigkeiten und Erläuterung der Resultate	3
2.1	Datenanalyse	3
2.2	Effekt von Datenreinigung auf eine ICA Dekomposition	3
2.3	Datenvisualisierung	4
3	Fazit und Ausblick	4

1 Das Unternehmen

Das Swartz Center for Computational Neuroscience (SCCN) ist ein Forschungszentrum welches an die University of California, San Diego (UCSD) angegliedert ist. Das SSCN entwickelt Datenanalyse Methoden und die zugehörige Software für die Forschung mit Elektroenzephalografie (EEG). Als Entwickler der meistgenutzten (Matlab) Toolbox für EEG-Datenanalyse 'eeglab' ist das Zentrum weit bekannt in der EEG-Forschungsgemeinschaft. Ich arbeite am Fachgebiet für Biopsychologie und Neuroergonomie an der TU Berlin welches eine enge Kolaboration mit dem SCCN führt. Im Rahmen meiner Tätigkeit am Fachgebiet hatte ich die Möglichkeit für ein sechs monatiges Forschungspraktikum am SCCN. Das Ziel dabei war Daten, welche ich zuvor in dem Labor an der TU Berlin erhoben habe, zu analysieren und dabei entstehende Analyse Methoden in eeglab zu integrieren oder zu publizieren. Mein Forschungsaufenthalt dort lief von 09/2021 bis 02/2022.

2 Beschreibung meiner Tätigkeiten und Erläuterung der Resultate

2.1 Datenanalyse

Wie oben erwähnt, war das primäre Ziel die Analyse von Daten eines Experimentes welches ich in Berlin erhoben habe. Die Herausforderung hierbei bestand in der schlechten Signal-to-Noise ratio der Daten. Diese entsteht durch eine große Menge an Störsignalen die vor allem durch Muskelaktivität erzeugt werden. Um dennoch wissenschaftliche Hypothesen mit den Daten testen zu können, ist eine aufwändige Säuberung der Daten vonnöten. Meine Aufgabe bestand also darin mich zunächst in die bestehenden Methoden zur Datensäuberung einzulesen und diese auf meine Daten anzuwenden. Eine Herausforderung war die Findung eines guten Kompromisses zwischen Reproduzierbarkeit und Sauberkeit. Damit meine Datensäuberung reproduzierbar und ggf. auf andere Datensätze übertragbar ist, war es wichtig automatisierte Methoden zu verwenden. Diese Algorithmen funktionieren jedoch nicht auf allen Datensätzen gleich gut weshalb es notwendig sein kann manuelle Datensäuberung zu betreiben. Nach dem Testen verschiedener Ansätze habe ich eine Reihe Methoden gefunden, welche passend für meinen Datensatz und gleichzeitig automatisiert waren. Hierbei folgende Punkte wichtig:

- Finden von EEG-Kanälen welche zu verrauscht sind und diese durch umliegende Kanäle interpolieren
- Entfernen von Netzstörgeräuschen
- Entfernen von Zeit-segmenten mit über-durchschnittlich starkem Rauschen

Da das letztendliche Ziel eine Dekomposition der Daten mithilfe einer Independent Component Analysis (ICA) ist, wurde der Effekt von verschiedenen Methoden zur Datensäuberung anhand der Dekompositionsqualität gemessen. Da die Berechnung einer ICA auf der großen Datenmenge rechnerisch zeitaufwändig ist, konnte ich nur bedingt schnell über die Methoden iterieren um diese zu testen und Parameter zu optimieren. Glücklicherweise ist das SCCN an das San Diego Super Computer Center angeschlossen. Die Nutzung des Supercomputers hat es mir ermöglicht innerhalb des Praktikums eine akzeptable Reinigung der Daten zu finden. Dennoch gestaltete sich diese Aufgabe als zeitaufwändiger als zunächst angenommen, weshalb ich die eigentliche Analyse der Daten erst in Berlin beginnen konnte.

2.2 Effekt von Datenreinigung auf eine ICA Dekomposition

Im Rahmen meiner Suche nach geeigneten Methoden zur Datenreinigung kam ich in eine Diskussion mit einem Kollegen über den Effekt von der Entfernung von verrauschten Zeitsegmenten auf die anschließende ICA. Für die Identifikation von 'schlechten' Zeitsegmenten gibt es verschiedene Methoden welche jedoch noch nie systematisch verglichen wurden. Deshalb starteten wir ein Projekt um eben genau diesen Vergleich durchzuführen. Hierbei suchten wir zunächst frei verfügbare EEG Datensätze. Dabei achteten wir darauf, dass sich die zugrunde liegende experimentale Paradigma in dem Maß an Bewegung des Probanden unterscheiden, da durch Bewegung induzierte Muskelaktivität eine Hauptquelle für Störsignale ist. Als Reinigungsmethoden wurden zwei verschiedene Algorithmen ausgewählt. Der erste schneidet das Signal in kleine Zeitsegmente und vergleicht die mittleren Amplituden und deren Varianz mit den anderen Segmenten. Zeitsegmente mit stark abweichenden mittleren Amplituden oder ungewöhnlich hoher Varianz werden entfernt. Die zweite Methode ist Teil des verwendeten ICA Algorithmus names Adaptive Mixture ICA (AMICA)[1]. Die AMICA lernt zunächst ein Modell der Daten und berechnet anschließend die log-likelihood von jedem Sample bezüglich diesem Modell. Sample mit einer zu geringen log-likelihood werden entfernt. Diese Prozess wird mit einer vorgegebenen Anzahl an Iterationen wiederholt. Beide Methoden sowie deren Kombination wurden auf die Daten angewandt und der Effekt anhand einer Reihe Qualitätsmaße evaluiert. Die Analyse dieser Effekte ist aktuell im Gange und wird hoffentlich in naher Zukunft publiziert.

2.3 Datenvisualisierung

Zu Beginn der Datenreinigung war es wichtig einen Eindruck davon zu bekommen wie verrauscht die Daten sind und wie die Störsignale aussehen. Dazu ist es am hilfreichsten die Daten in ihrer Rohform zu betrachten. Leider ist die vorhandene Software der eeglab Toolbox dafür sehr langsam. Deshalb versuchte ich eine eigene, schnellere Software zu schreiben welche EEG Daten visualisiert. Gleichzeitig war meine Motivation meine Julia Kenntnisse zu verbessern. Ich verbrachte einige Zeit mit diesem Projekt und bin zufrieden mit dem Prototypen [2] der online zu finden ist. Da dieses Projekt jedoch nicht meine oberste Priorität hatte, war ich leider nicht in der Lage aus dem Prototypen eine vollständige Software zu bauen.

3 Fazit und Ausblick

Für mich war die Zeit in San Diego sehr bereichernd. Persönlich konnte ich mich viel weiter entwickeln und die USA besser kennen lernen. Akademisch war das Praktikum auch sehr wertvoll. Durch das SCCN bin ich mit einer Reihe unterschiedlicher Forscher aus verschiedenen Bereichen in Kontakt gekommen und habe einen tiefen Einblick in die Forschung bekommen. Die ersten beiden Projekte sind noch aktiv in Arbeit und werden hoffentlich in naher Zukunft publiziert. An dieser Stelle möchte ich mich bei Prof. Gramann meinem Professor hier an der TU, John Iversen meinem Betreuer in San Diego und dem PROMOS Team der TU Berlin welche das Praktikum finanziell unterstützt haben bedanken.

Literatur

- [1] J. Palmer, K. Kreutz-Delgado, and S. Makeig, “Amica: An adaptive mixture of independent component analyzers with shared components,” 01 2011.
- [2] T. Berg, “eegraw explorer,” 7 2022. [Online]. Available: https://github.com/timo-berg/eegraw_explorer