

Aufgabenblatt 2

Daten fenstern

Unterteilt die nach Labels segmentierten Abschnitte weiter in Fenster fester Länge (z.B. 2 Sekunden, sollte ein konfigurierbarer Parameter sein). Jedes dieser Fenster soll später einem Sample für das Training bzw. die Evaluierung des Workload-Klassifikators dienen.

Vorverarbeitung

Informiert euch in diesem Paper <https://www.frontiersin.org/articles/10.3389/fninf.2015.00016/full> und in weiterer Literatur (z.B. https://neuro.inf.unibe.ch/AlgorithmsNeuroscience/Tutorial_files/BaselineCorrection.html) über sinnvolle Arten der EEG-Vorverarbeitung. Implementiert ggf. entsprechende Methoden zur Normalisierung, Datenselektion, Filterung, Re-Referenzierung, etc. (nicht alle diese Methoden sind unbedingt notwendig).

Merkmalsextraktion

Entwickelt eine Methode zur Extraktion von Frequenz-basierten Merkmalen, sodass als Ausgabe der Methode einem Fenster aus der vorhergehenden Teilaufgabe ein Merkmalsvektor zugeordnet werden kann. Lasst euch bei der genauen Definition der Merkmale von der Literatur und weiteren Quellen inspirieren. Eure Merkmale können z.B. pro Elektrode das Betragsspektrum für einen bestimmten Frequenz-Bin ausdrücken, oder ihr fasst diese noch einmal in die bekannten EEG-Bänder zusammen.

Klassifikation: Literatur

Informiert euch im Paper <https://iopscience.iop.org/article/10.1088/1741-2552/aab2f2/meta> über Methoden des maschinellen Lernens im BCI-Bereich. Welche Besonderheiten fallen euch auf im Vergleich zu anderen Bereichen (z.B. Bildverarbeitung)? Wählt einen Klassifikator aus, mit dem ihr eure Analysen beginnen wollt.

Klassifikations-Framework

Nutzt das Framework scikit-learn (<https://scikit-learn.org/stable/index.html>), um eine Schleife zur Klassifikation eurer Samples zu definieren. Dabei sollt ihr eine Kreuzvalidierung verwenden und den Code so aufbauen, dass er mit allen aktuellen und zukünftigen Sessions auf einmal arbeiten kann. Die Schleife soll dabei auf der Menge fertig extrahierter Merkmalsvektoren und Labels arbeiten und Training und Auswertung eines Klassifikators durchführen. Dabei sollt ihr zwei Varianten umsetzen:

- Personen-abhängig: Es werden für die Kreuzvalidierung nur die Daten einer Person verwendet. Dabei kann eine stratifizierte k-fold Kreuzvalidierung eingesetzt werden. Diese Auswertung soll aber für jede Person durchgeführt werden, um die Ergebnisse dann zusammenzuführen.
- Personen-unabhängig: Es wird eine Kreuzvalidierung auf den Daten aller Personen durchgeführt. Hierbei sind die Personen in Test- und Trainingsdatenmenge voneinander verschieden. Hier bietet sich eine Leave-one-person-out Kreuzvalidierung an (d.h., in jeder Iteration werden die Daten genau einer Person als Testdaten verwendet). Hier kann Normalisierung eine wichtige Rolle spielen.

Betrachtet die Daten dabei als 3-Klassen-Problem mit den drei verschiedenen Workload-Stufen.

Auswertung und Visualisierung

Bei der Auswertung sollen geeignete Auswertungsmetriken verwendet werden. Eine Auswahl möglicher Maße findet ihr hier: https://scikit-learn.org/stable/modules/model_evaluation.html

Stellt die Ergebnisse der Auswertung außerdem in Form einer Konfusionsmatrix dar: https://scikit-learn.org/stable/modules/model_evaluation.html#confusion-matrix

Das Ergebnis der Klassifikationsauswertung inklusive der Streuung soll grafisch aufbereitet werden, insbesondere zum Vergleich verschiedener Konfigurationen (siehe folgende Aufgaben). Geeignete Visualisierungen können z.B. Boxplots oder Violinenplots sein. Überlegt euch bei Bedarf auch andere Formen der Visualisierung.

Fenstergröße

Nutzt das implementierte Klassifikationsframework zur Bestimmung des Einflusses der Fenstergröße, indem ihr verschiedene Werte testet und mit den Methoden aus der vorhergehenden Aufgabe grafisch gegenüberstellt.

Feature/Elektroden-Auswahl

Untersucht, welche Elektroden bzw. welche Merkmale besonders zur Klassifikation beitragen. Dafür können verschiedene Ansätze genutzt und gegenübergestellt werden, z.B.:

- https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SelectKBest.htm
- <https://www.samuelstaylor.org/articles/feature-importance-for-any-model.html>
- https://scikit-learn.org/stable/modules/permutation_importance.html
- <https://www.sciencedirect.com/science/article/pii/S1053811913010914>