

Feature-Engineering-Strategien für Eye-Tracking-Daten – Einfluss von Preprocessing auf die Merkmalsqualität

Untersuchung, wie verschiedene Preprocessing-Schritte (z. B. Missing-Value-Handling, Normalisierung, Signal-Glättung, Outlier-Handling) die Qualität und Aussagekraft abgeleiteter Features beeinflussen.

Schwerpunkte:

- Vergleich von roh extrahierten vs. vorverarbeiteten Features (z. B. Fixationsdauer, Sakkadengeschwindigkeit, Pupillendurchmesser).
- Einsatz von Feature-Importance-Analysen (Random Forest, Permutation Importance) zur Bewertung.
- Entwicklung einer optimierten Pipeline, die Merkmalsqualität maximiert und irrelevante Merkmale reduziert.

Verwendete Techniken / Methoden

- Feature-Sets: Fixationsdauer-/Anzahl, Sakkaden-Geschwindigkeit/Beschleunigung, Pupillen-Features, Scanpath-Metriken.
- Feature-Selektion/-Bewertung: Permutation Importance, Random-Forest-Importance.
- Vergleich roh vs. vorverarbeitet (mit/ohne Smoothing/Imputation/Normierung).
- Klassifikatoren: Logistic Regression, Random Forest, XGBoost; Validierung subject-wise.
- Robuste Pipeline mit minimaler Leckage.

Zeitraum

- Wochen 1–3: Literaturrecherche, Datenakquise, Setup, Data-Profiling.
- Wochen 4–6: Pipeline-Implementierung & Baselines.
- Wochen 7–9: Experimente & Auswertung.
- Wochen 10–12: Ergebnisanalyse, Schreiben, Replikation.

Literaturempfehlungen (Startpunkte)

- Bekler, M. et al. (2024). Assessing Feature Importance in Eye-Tracking Data. Applied Sciences.
- Birawo, B. A. et al. (2025). Performance Analysis of Eye Movement Event Detection Neural Network Models with Different Feature Combinations. MPDI
- Dar, A. H. et al. (2020/2021). REMoDNaV: robust eye-movement classification for dynamic stimulation. BRM.
- Wagner, A. S. (2019). multimatch-gaze: The MultiMatch algorithm for gaze path comparison in Python. JOSS
- Devlin, S. P. et al. (2022). Scan-based eye tracking measures are predictive of workload transition performance. Applied Ergonomics.

Optional, sehr technisch: