World Academy of Science, Engineering and Technology
International Journal of Computer and Systems Engineering
Vol:16, No:04, 2022

Exploring Data Leakage in EEG Based Brain-Computer Interfaces: Overfitting Challenges

Authors : Khalida Douibi, Rodrigo Balp, Solène Le Bars

Abstract : In the medical field, applications related to human experiments are frequently linked to reduced samples size, which makes the training of machine learning models quite sensitive and therefore not very robust nor generalizable. This is notably the case in Brain-Computer Interface (BCI) studies, where the sample size rarely exceeds 20 subjects or a few number of trials. To address this problem, several resampling approaches are often used during the data preparation phase, which is an overly critical step in a data science analysis process. One of the naive approaches that is usually applied by data scientists consists in the transformation of the entire database before the resampling phase. However, this can cause model' s performance to be incorrectly estimated when making predictions on unseen data. In this paper, we explored the effect of data leakage observed during our BCI experiments for device control through the real-time classification of SSVEPs (Steady State Visually Evoked Potentials). We also studied potential ways to ensure optimal validation of the classifiers during the calibration phase to avoid overfitting. The results show that the scaling step is crucial for some algorithms, and it should be applied after the resampling phase to avoid data leackage and improve results.

Keywords : data leackage, data science, machine learning, SSVEP, BCI, overfitting
Conference Title : ICDSTA 2022 : International Conference on Data Science, Technologies and Applications
Conference Location : Boston, United States
Conference Dates : April 21-22, 2022