Proceedings of the 6th International Brain-Computer Interface Meeting, organized by the BCI Society DOI: 10.3217/978-3-85125-467-9-184

Hyperparameter Optimization for Machine Learning Problems in BCI

A. Meinel[1], K. Eggensperger[2], M. Tangermann[1*], F. Hutter[2]
[1]Brain State Decoding Lab, [2]Machine Learning for Automated Algorithm Design Group
(both:) Cluster of Excellence BrainLinks-BrainTools, Dept. Computer Science, University Freiburg, Germany

*Albertstr. 23, 79104 Freiburg, Germany. E-mail: michael.tangermann@blbt.uni-freiburg.de

Introduction: Pipelines for BCI data analysis comprise several building blocks, such as signal preprocessing, feature extraction, decoding of features and output shaping for the BCI application at hand. These components contain many hyperparameters, such as frequency bands, time intervals, regularization factors, adaptation parameters, etc., which need to be chosen carefully in order to obtain optimal overall performance. As even simple BCI setups comprise tens of mutually dependent (discrete or continuous) hyperparameters, the search space is too large for a full grid search. Even though experts can tune most parameters based on experience, the inter-subject variability inherent to BCIs is likely to reward a subject-dependent optimization strategy.

Material, Methods and Results: 20 healthy volunteers participated in a cued isometric force task of the hand (SVIPT, [1]) while EEG signals were recorded (64 passive channels, BrainAmp DC). A full setup is described in [2]. Comparable to a motor imagery processing pipeline, oscillatory EEG bandpower of a narrow frequency band within a (pre-trial) time interval was investigated. The analysis aimed at extracting supervised EEG subspaces (SPoC, [3]) which maximize the predictive squared correlation of their bandpower with continuous labels. The latter were obtained from a task-related motor performance metric. Overall, the processing pipeline comprised one nominal and three integer continuous hyperparameters resulting in 242,730 possible configurations. We investigated the performance and time requirements of four automatic methods for hyperparameter learning (SMAC [4], TPE [5], Spearmint [6], and random search). The first three perform Bayesian optimization, which iteratively fits and updates a probabilistic model to predict the performance of all parameter settings and uses this model to determine promising configurations to evaluate next. The mentioned methods ran 4 hours of CPU time, enough to evaluate up to 300 configurations (almost a factor of 1,000 less than the overall number of configuration). The results of the automatic methods are given in Fig. 1. Except for very short time budgets, the optimizer SMAC delivered best correlation values (lowest loss), followed by TPE and random search.

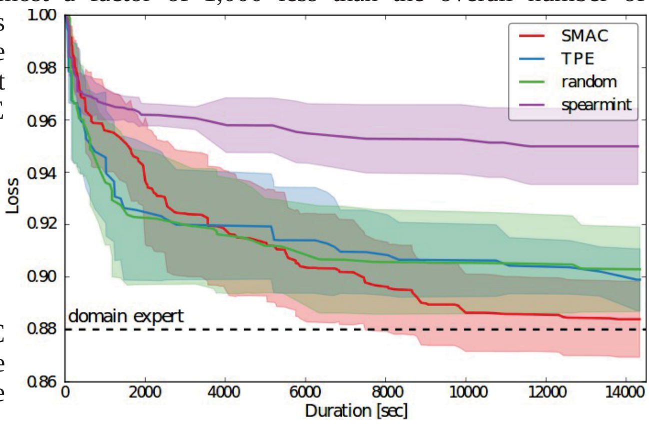

Figure 1. *Comparison of four hyperparameter optimization strategies wrt. the best obtained parameter set. The y-axis represents the (1-r²)-loss minimized (mean +/- stddev), the x-axis shows the runtime (CPU time budget).*

The best hyperparameter sets discovered by SMAC were very plausible and closely resembled those determined by the expert, with frequencies in the alpha band and short time windows within [-1000ms, 0ms] relative to the go cue of the hand force task.

Discussion and Significance: Since modern hyperparameter learning strategies outperform grid search by orders of magnitude and do not require human expert time, they can substantially facilitate progress in BCI. Although their application can be expensive for involved machine learning pipelines, they can also be parallelized for inclusion in an online system.

Acknowledgements: This work was (partly) supported by BrainLinks-BrainTools, Cluster of Excellence funded by the German Research Foundation (DFG), grant number EXC 1086.

References
[1] Reis J, Schambra HM, Cohen LG, Buch ER, Fritsch B, Zarahn E, Krakauer JW. Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. *PNAS*, *106*(5), 1590-1595, 2009.
[2] Meinel A, Castaño-Candamil JS, Dähne S, Reis J, Tangermann M. EEG band power predicts single-trial reaction time in a hand motor task. In *Proceedings* of *Neural Engineering (NER), 7th International IEEE/EMBS Conference*, 182-185, 2015.
[3] Dähne S, Meinecke F, Haufe F, Höhne J, Tangermann M, Müller K-R, Nikulin V. SPoC: A novel framework for relating the amplitude of neural oscillations to behaviorally relevant parameters, *NeuroImage*, 86:(111–122), 2014.
[4] Hutter F, Hoos H, Leyton-Brown K. Sequential model-based optimization for general algorithm configuration. *In Proceedings of the fifth conference on Learning and Intelligent Optimization*, 507–523, 2011.
[5] Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. In *Proceedings of the Neural Information Processing Systems Conference*, 2546–2554, 2011.
[6] Snoek J, Larochelle H, Adams R. Practical Bayesian optimization of machine learning algorithms. In *Proceedings of the Neural Information Processing Systems Conference (NIPS)*, 2960–2968, 2012.