Workload classification from EEG: Comparison of LDA and CNNs

Jelena Jovanovic[1], Luka J. Bojovic[1], Pavle Mijovic[1], Bogdan Mijovic[1]

[1] mBrainTrain LLC, Belgrade, Serbia

Introduction

Accurate, real-time and objective mental workload (MWL) assessment of the operators could help accidents prevention in safety critical operations. Although the miniaturization of the neuroimaging devices largely accelerated the neuroergonomics research, the currently obtained results in classifying the MWL based on the electroencephalographic (EEG) data are still far from reliable usage in the real-world applications.

The convolutional neural network (CNN) models provide promising advancement for timely estimation of the operators MWL, with the goal of developing a neuroadaptive system. However, current state-of-the-art in the EEG research still depends on the classical signal processing techniques for extracting the features that are further used for the MWL classification, mainly by using the linear models, such as the linear discriminant analysis (LDA, Zhang et al. 2018). Standard processing steps and human feature design are a limiting factor in the traditional approaches.

This study compares the performance of CNN based architectures to LDA, on a 4-class workload level classification.

Experimental Setup

A total of 43 healthy subjects (26 female, 21-38 years old (25.8 ± 5.3 years)) performed the NASA Multi-Attribute Task Battery II (MATB-II). Each participant performed one training session to get familiar with the task a day before the experiment. No physiological data were acquired during the training session. The following day, the participant performed the experiment that contained two sessions. Each session contains 3 segments, which are further divided in blocks of different durations (1, 2.5 or 5 minutes) and different levels of workload presented in a randomized order (Passive watching - PW, Low load - LL, Medium load - ML and High load – HL). Workload levels were modulated by different task demands in MATB II based on task structure in (Smith et al. 2001). Importantly, PW intervals were of a shorter duration then the others and in total 8.6h of PW was recorded and 21.5h of each of the remaining load levels. During the experiment, EEG data was recorded using the 24-channel Smarting MOBI device.

EEG data pre-processing

EEG data were processed in EEGlab (Delorme and Makeig 2004). The data were first band-pass filtered from 1 - 40Hz and then standardized (z-scored) across time and channels from each session. For LDA-based approach, the individual alpha frequency (IAF) was calculated for each subject, based on which we extracted the features: relative band power in theta [IAF-6; IAF-2], alpha [IAF-2; IAF+2] and beta [IAF+2; IAF+20] frequency bands across all 24 EEG electrodes.

For both LDA and CNN approach, the processed EEG data were divided into windows of 10 seconds and each window is labeled based on the difficulty block it belongs to.

Models

The CNN based model consists of 9 1D convolutional layers. The average pooling and softmax in the last layer were used to obtain the class predictions. Windowed 24 EEG channels data were brought to the input, and the 1D convolution was applied only across the time dimension. Batch normalization, GELU activation functions and dropout were employed between layers, and the model was trained using the standard cross-entropy loss function.

LDA classifier is trained on a total of 72 features (alpha, theta and beta bands from 24 EEG electrodes).

Evaluation

In order to make the results consistent, a 10-fold cross validation was performed for both models, and the training and test datasets for both models were extracted in exactly the same way. In each fold 8 randomly chosen sessions from different subjects were used for test and 78 sessions for training (out of 86 sessions in total). The models were trained on 4-class classification (PW, LL, ML and HL). We also report the results for 2 classes where PW+LL and ML+HL were grouped in order to compare it to the results from the literature (e.g. Zhang et al. 2018). However, the 2-classes results are also trained on 4-classes and grouped only after the classification.

In addition, the LDA analysis was performed on a per-subject basis, where the first session was used as a training set, while the second session was used as a test set. In per-subject analysis, CNN model was first pre-trained on all subjects and then the parameters were fine-tuned with a lower learning rate (5e-4) on each individual subject, in order to obtain a customized model for each subject independently. As with LDA, one session was used for fine-tuning, and the other one for validating and testing.

Results and Discussion:

The results of the corresponding two models on a 4- and 2- class classification of MWL are shown in Table 1. Confusion matrices for both models on all subjects are shown in Fig 1.

	LDA on all subjects	CNN on all subjects	LDA per subject	CNN fine-tuned on one subject
4 classes	47.7% ± 1.7%	58.6% ± 3.6%	50.9% ± 8.4%	62.7% ± 7.2%
2 classes	75.7% ± 2.6%	85.5% ± 5.3%	80.8% ± 8.7%	88.8% ± 6.6%

Table 1. Accuracies of the respective models

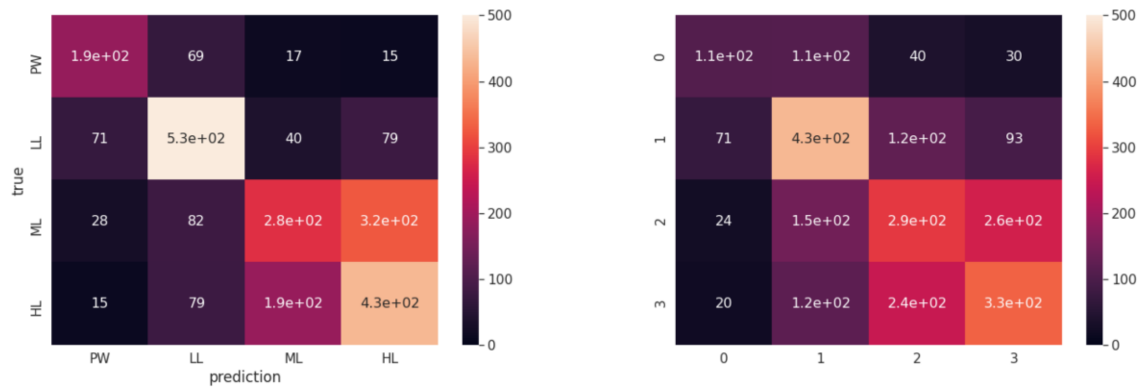

Figure 1. Comparison of the confusion matrices obtained for the CNN model (LHS) and the LDA model (RHS)

Noticeably, the confusion matrices show that, for the CNN model, the misclassification most frequently occurs between the ML and HL classes. There is also a level of LL class being misclassified instead of ML or HL, but to a less significant level. The reason for this misclassification may lie in the fact that the task difficulty of a block may vary for different 10-seconds windows within the same task difficulty block. LDA classifier confusion matrix shows significantly higher overlap between classes compared to CNN model.

In the per-subject analysis (Table 1) CNN also outperforms LDA for both 4- and 2-class cases.

Conclusion

The results clearly demonstrate that neural networks achieve superior performance compared to the traditional linear models, for both all-subjects and per-subject classification approaches.

To test the hypothesis that the differences in the task difficulty between the windows within the same block influence the classification performance, future work will have to consider additional tests, such as training the classifiers on larger windows or exploring methods to extract task difficulty from each window separately.

References

Delorme, A. and Makeig, S. (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Jour. of neurosc. met.*, *134*(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Smith M.E., Gevins A., Brown H., Karnik A., and Du R. (2001) Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction. Hum. Fact. 2001;43(3):366-80. doi: 10.1518/001872001775898287. PMID: 11866193.

Zhang, X., Krol, L.R. and Zander, T.O. (2018). October. Towards task-independent workload classification: shifting from binary to continuous classification. In *2018 IEEE Int. Conf. on Sys., Man, and Cyber.* 556-561. IEEE.