 DOI:10.3217/978-3-85125-378-8-30

Towards Cross-Subject Workload Prediction

Carina Walter[1], Philipp Wolter[1], Wolfgang Rosenstiel[1], Martin Bogdan[2,1] and Martin Spüler[1]

[1] Department of Computer Engineering, Eberhard-Karls University Tübingen, Tübingen, Germany
Carina.Walter@uni-tuebingen.de
[2] Department of Computer Engineering, University of Leipzig, Leipzig, Germany

Abstract

Developing an electroencephalogram (EEG) based workload (WL) prediction could be utilized in learning environments to adapt instructional material to the students individual WL and thereby support the learning process successfully. In an adaptive learning environment it is not feasible to use data from the same subject and the same task for classifier training and testing. Therefore, the present study examined cross-subject WL prediction, in which one subject's individual WL is predicted based on data from other subjects. EEG-data was recorded from 10 subjects who had to solve math problems with increasing level of difficulty. By applying linear regression within- and cross-subject a precise WL prediction could be reached. For the within-subject WL prediction an average correlation coefficient (CC) of $CC = 0.88$ was achieved. However the cross-subject WL prediction leads to $CC = 0.84$ on average. Since both prediction methods achieved good WL prediction results the cross-subject method is a feasible approach that can be used in an online adaptive system.

1 Introduction

In accordance with Cognitive Load Theory (CLT) [5] the type and amount of workload (WL) during learning is crucial for successful learning and should be held in the individual optimal WL range for each learner. Thus it seems advisable to provide technological support for learners, which adapts instructional materials and tasks to their level of expertise and WL capacity.

As it is not feasible to use data from the same subject and the same task for classifier training and testing in an adaptive learning environment, one can either use data from the same subject but different tasks (cross-task) or from the same task, but different subjects (cross-subject) to calibrate the classifier.

In a previous study [7] we used cross-task classification, where a SVM was trained on electroencephalogram (EEG)-data, recorded while participants had to solve working memory tasks. Subsequently the SVM was tested on EEG-data recorded while solving complex mathematical tasks. This approach lead to classifier accuracies around chance level. Thus we introduced a cross-subject WL prediction in the present study based on linear regression, for a group of 10 subjects. As most of the EEG-based WL classifications are subject-specific, apart from a few exceptions [8], this work is a step towards the development of EEG-based WL classifiers as it would enable training a classifier once that could handle multiple subjects.

2 Materials and Methods

2.1 Participants and Task Design

A total of 10 subjects (17 - 32 years) voluntarily participated in the EEG experiment. The experiment had a within-subject design and comprised two tasks, a WL as well as a vigilance

 DOI:10.3217/978-3-85125-378-8-30

task. In this article we will merely report the results of the WL task. The subjects had to solve 240 math problems with an increasing level of difficulty, by typing the solution in a given time. The presented problems varied in difficulty as measured by the information content (Q) according to Thomas [6], with a Q value ranging from 0.6 (easy) to 7.2 (difficult). As postulated from Kantowitz [2] increasing task difficulty ($\hat{=}$ increasing Q-value) always increases WL, since by definition, an increase of task difficulty demands additional capacity. Further, an increase of WL is characterized in EEG-data by changes in the theta-, alpha- and beta-frequency bands [3]. To avoid the classifier of being based on perceptual-motor confounds the time windows used for analyzing EEG-data should not contain motor events. Therefore, each trial was divided into two phases: First, the calculation phase occurred (= phase for analyzing the EEG-data), where the problem to be solved was shown for 5 sec. Subsequently subjects had 3.5 sec to type in their result, followed by an inter-trial interval of 1.5 sec.

2.2 EEG Recording

A set of 29 active electrodes (actiCap, BrainProducts GmbH), attached to the scalp placed according to the extended International Electrode 10 - 20 Placement System, was used to record EEG-signals. Three additional electrodes were used to record an electrooculogram (EOG); two placed horizontally at the outer canthus of the left and right eye to measure horizontal eye movements and one placed in the middle of the forehead between the eyes to measure vertical eye movements. EOG- and EEG-signals were amplified by two 16-channel biosignal amplifier systems (g.USBamp, g.tec). The sampling rate was 512 Hz. EEG-data was high-pass filtered at 0.1 Hz and low-pass filtered at 100 Hz during the recording. Furthermore a notch-filter was applied between 48 - 52 Hz to filter power line noise.

2.3 Data Processing and Analysis

For further analysis we reduced the number of channels to 17 (FPz, AFz, F3, Fz, F4, FC3, FCz, FC4, C3, Cz, C4, CPz, P3, Pz, P4, Oz, POz), to lower the influence of possible artifacts, which are most prominent on the outer channels. As frequency bands were not consistent and varied between subjects, we used a wide frequency range of 4 - 30 Hz [3].

The power spectra were calculated for each 5 sec window (= calculation phase) by using autoregressive models based on Burgs maximum entropy method [1], using a model order of 32. In addition, the data was z-score normalized along the channels, meaning for each trial the mean of each frequency bin equals zero. The feature selection was conducted during the 10-fold or leave-one-subject-out cross-validation only on the training data. Each feature was correlated with the Information Content (Q) and the 125 features with the highest r^2-values were selected for the regression analysis. For WL prediction we used a linear ridge regression with the regularization parameter set to a fixed value of $\lambda = 0.001$.

As criterion for performance evaluation of the presented prediction method we used the correlation coefficient (CC) to observe the statistical relationship between the actual and the predicted Q-values, as well as the root-mean-squared error (RMSE) to examine the difference between the actual and the predicted Q-values.

Since single-trial prediction is not necessary in a learning environment, the regression output was additionally smoothed to improve prediction accuracy at the expense of increased delay of the system. We used a window-size of 6 trials, which still guaranteed a response time $\leq$ 1 min of the system, which is feasible for the detection of WL.

 DOI:10.3217/978-3-85125-378-8-30

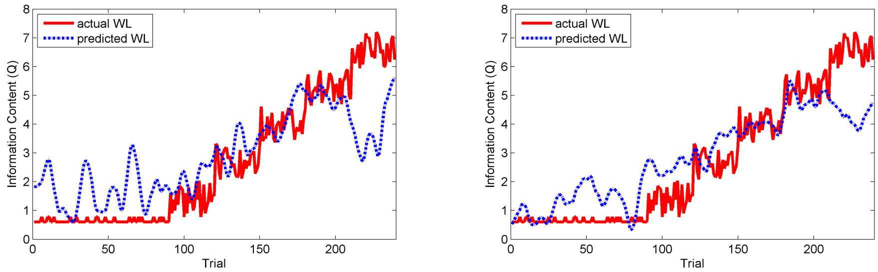

Figure 1: Distribution of actual and predicted WL for subject 1. The continuous red line represents the actual WL and the smoothed dashed blue line the predicted WL for each trial. Left: Within-subject - Using a 10-fold cross-validation; Right: Cross-subject - Regression trained on data of nine subjects and tested with data of the remaining subject 1.

3 Results

The performance for the within-subject WL prediction was calculated by using a 10-fold cross-validation on 240 trials (see Table 1). For the evaluation of the cross-subject WL prediction the regression was trained on data of nine subjects ($= 240 \cdot 9$ trials) and tested with the complete data-set of the remaining subject (see Table 2). The within-subject WL prediction reached on average a correlation coefficient of $CC = 0.66$ and a $RMSE = 1.70$. Using the smoothed regression output for within-subject WL prediction leads to more robust and even better results with an average over all subjects of $CC = 0.88$ and $RMSE = 1.02$ (see Table 1).

As stated in Table 2 the average results over all 10 subjects for cross-subject WL prediction slightly decreased. Applying linear regression to unsmoothed data leads on average to a $CC = 0.58$ and a $RMSE = 1.95$. Using the smoothed data for cross-subject WL prediction improved the CC by 0.26 and the $RMSE$ by 0.37.

The distributions of actual and predicted WL for within- and cross-subjects are exemplary shown for subject 1 in Fig. 1. In both cases the increasing WL was successfully predicted by using the linear regression method. It can be noticed that during the first 90 trials ($Q \leq 0.9$) and the last 30 trials ($Q \geq 6.0$) the actual and predicted WL deviate strongest from each other. This may be due to subjects being unchallenged or overburdened while solving these tasks.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	mean
CC	0.43	0.69	0.68	0.58	0.84	0.58	0.60	0.72	0.65	0.78	0.66
CC (smooth)	0.69	0.93	0.93	0.89	0.94	0.87	0.76	0.91	0.93	0.93	0.88
$RMSE$	2.20	1.61	1.71	2.09	1.11	1.95	1.63	1.60	1.82	1.30	1.70
$RMSE$ (smooth)	1.39	0.86	0.99	1.21	0.68	1.15	1.04	1.04	1.08	0.73	1.02

Table 1: Performance results of the within-subject WL prediction using 10-fold cross-validation.

4 Discussion

The results show that the amount of WL can be predicted using a cross-subject regression method. Compared to earlier studies, where cross-task classification [7] leads to classification

DOI:10.3217/978-3-85125-378-8-30

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	mean
CC	0.67	0.50	0.65	0.63	0.63	0.59	0.39	0.62	0.52	0.57	0.58
CC (smooth)	0.89	0.91	0.88	0.91	0.89	0.86	0.74	0.84	0.71	0.79	0.84
$RMSE$	1.68	2.56	1.69	1.73	1.72	1.77	2.65	1.87	2.04	1.81	1.95
$RMSE$ (smooth)	1.13	2.00	1.25	1.39	1.60	1.44	2.11	1.58	1.57	1.68	1.58

Table 2: Performance results of the cross-subject WL prediction. Regression was trained on data of nine subjects and tested with data from the remaining subject.

accuracies merely around chance level, cross-subject prediction seems to be more robust and recommendable. As the math problems were presented in a fixed order, at advancing levels of difficulty, EEG-signals might change due to non-stationarity (e.g. caused by fatigue) over time. Actually non-stationarity within a session should not have a great influence on the EEG-data while using cross-subject methods, since non-stationaries between subjects are assumed to be larger than within a session. Since we used a very simple method for normalization and thereby alleviated non-stationaries, we expect to further improve the results by using more advanced methods for reduction of non-stationaries [4]. In an upcoming study, these results will be used in an online learning environment to predict the user's WL and adapt the presented exercises accordingly, to support students successfully in their learning process.

Acknowledgments

The research was supported by ScienceCampus Tübingen and the LEAD Graduate School at the University of Tübingen, which is funded by the German Research Foundation.

References

[1] T. M. Cover and J. A. Thomas. *Elements of information theory.* Hoboken,NJ: Wiley-Interscience, 2006.

[2] B. H. Kantowitz. *Human Factors Psychology*, chapter Mental Workload, pages 81–122. North-Holland, 1987.

[3] M. Pesonen, H. Hämäläinen, and C. M. Krause. Brain oscillatory 4-30 Hz responses during a visual n-back memory task with varying memory load. *Brain Research*, 1138:171–177, 2007.

[4] W. Samek, M. Kawanabe, and C. Vidaurre. Group-wise stationary subspace analysis - a novel method for studying non-stationarities. In *Proc. 5th Int. Brain-Computer Interface Conf.*, pages 16–20, 2011.

[5] J. Sweller, J. J. G. Van Merrienboer, and F. Paas. Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3):251–296, 1998.

[6] H. B. G. Thomas. Communication theory and the constellation hypothesis of calculation. *Quarterly Journal of Experimental Psychology*, 15(3):173–191, 1963.

[7] C. Walter, S. Schmidt, W. Rosenstiel, P. Gerjets, and M. Bogdan. Using cross-task classification for classifying workload levels in complex learning tasks. In *Proceedings of the 5th IEEE Humaine Association Conference on Affective Computing and Intelligent Interaction*, pages 876–881, Geneva, Switzerland, 2013.

[8] Z. Wang, R. M. Hope, Z. Wang, Q. Ji, and W. D. Gray. Cross-subject workload classification with a hierarchical bayes model. *NeuroImage*, 59(1):64–69, 2012.