Online Workload Recognition from EEG Data during Cognitive Tests and Human-Machine Interaction

Dominic Heger, Felix Putze, and Tanja Schultz

Cognitive Systems Lab (CSL)
Karlsruhe Institute of Technology (KIT), Germany
{dominic.heger,felix.putze,tanja.schultz}@kit.edu

Abstract. This paper presents a system for live recognition of mental workload using spectral features from EEG data classified by Support Vector Machines. Recognition rates of more than 90% could be reached for five subjects performing two different cognitive tasks according to the flanker and the switching paradigms. Furthermore, we show results of the system in application on realistic data of computer work, indicating that the system can provide valuable information for the adaptation of a variety of intelligent systems in human-machine interaction.

1 Introduction

Machines play an important role in our everyday lives as matters of communication, work, and entertainment. However, in the interaction with humans, machines widely neglect different internal states of their users with the consequence of unnatural interaction, inadequate actions and inefficient adaptation to the user. Research in Human Machine Interaction and Affective Computing approaches this problem by developing systems that sense the current situation of the user and adapt to it. In this paper we contribute a system for automatic assessment of mental workload on short segments of EEG data for real-time adaptation of intelligent systems.

Terms, such as mental workload, task demand, engagement, vigilance, and others are often imprecisely used in literature to describe a human internal state of mental effort. Throughout this paper we use the term 'workload' as the amount of mental resources that are used to execute a current activity.

Numerous types of applications can strongly benefit from the ability to adapt themselves according to a detected workload level of their users. For example, in industrial production an optimal productivity of the operators could be maintained by monitoring their level of workload. Intelligent assistants, such as interactive driver assistance systems in future cars, could attempt to delay communication in difficult traffic situations to more suitable points of time in order to shift and balance the user's workload and improve safety while driving. In the near future, humanoid robots will be commercially available as household robots or to assist in elderly care. Therefore, social skills of robots in the interaction with humans will become essential. The robot could adapt the strategies of

R. Dillmann et al. (Eds.): KI 2010, LNAI 6359, pp. 410–417, 2010.

its spoken dialog system, for example by using shorter utterances in situations where its owner has a high workload. Thus, workload recognition can be a first step towards a social and empathic behavior.

The assessment of a user's workload must have a high temporal resolution, i.e. a workload prediction made from a short segment of data, so that a system can instantaneously react to it by adaptation. In contrast to other biosignals, such as heart rate variability, or skin conductance, EEG is a direct measure of the electric activity of the brain. Therefore, EEG can directly reflect mental processes and is the most suitable measure for workload estimations on the basis of time slices as short as one second length. In many situations, especially in professional working environments, sensors can be integrated into professional clothing (e.g. helmet), which attenuates concerns on wearing EEG devices.

2 Related Work

The analysis of workload from EEG data has a tradition in the psychological community. However, there is still no consensus of workload effects on the EEG. This section outlines systems that have a focus on the automatic computational assessment of workload based on EEG data.

Gevins and Smith [1] evaluated data from subjects performing different tasks of computer interaction and sequential memorization of stimuli (n-back tests). They used spectral features of the theta and alpha frequency bands from chunks of four second length and applied subject-specific multivariate functions and neural networks for discrimination of three different workload levels.

Berka et al [2] developed the B-Alert system for monitoring alertness and cognitive workload in different operational environments using a wireless EEG sensor headset. They applied the system to several tasks, such as sleep deprivation studies, military motivated monitoring (Warship Commander Task), reaction and digit sequence identifaction, as well as image memorization tasks. They predicted four states of alertness using a linear discriminant function on spectral features out of the range 3-40Hz derived from one second long epoches of data.

Kohlmorgen, et al. [3] measured workload of car drivers. They applied a highly parameterized feature extraction optimized for each subject and classification by Liner Discriminant Analysis. Parameters include data segment length (10-30 seconds), channel selection, spatial filters, and frequency bands (within 3-15Hz). Workload was induced by a secondary auditory reaction task and a tertiary task, which consisted of mental calculation or following one speaker in a recording of simultaneous voices. The authors could show that mitigation of high workload situations for the driver based on online detection of workload leads to improvements in reaction time for most subjects.

Honal and Schultz [4] analyzed task demand from EEG data recorded in lecture and meeting scenarios. They used Support Vector Machines (SVMs) and Artificial Neural Networks for classification and regression of features from short time Fourier transform (2 second epoches). To make brain activity measurements less cumbersome, they also evaluated a comfortable headband in addition to a standard EEG cap.

Putze, Jarvis, and Schultz [5] proposed a multimodal recognizer for different levels of cognitive workload. They recorded EEG data in addition to skin conductance, pulse, and respiration and classified it on one minute windows in a person independent way using SVMs. For their evaluation they used data recorded from subjects performing a lane change task in a driving simulator, while solving visual and cognitive secondary tasks.

The current paper contributes an online workload classification system based on one second long segments of EEG data and presents post-processing by voting and temporal smoothing of the prediction results. Results for training on one cognitive test paradigm and application of the system on the data of another are presented, which shows robustness in task variation of the proposed classifier. Finally, we show the abilities of the system to estimate a person's workload during realistic data of computer work. This indicates that a system trained on standardized data, such as cognitive tests and resting periods, can successfully be applied for detection of workload in realistic scenarios.

3 System Setup

In this section we describe the system setup of the proposed system. Figure 1 shows a block diagram of the processing stages involved.

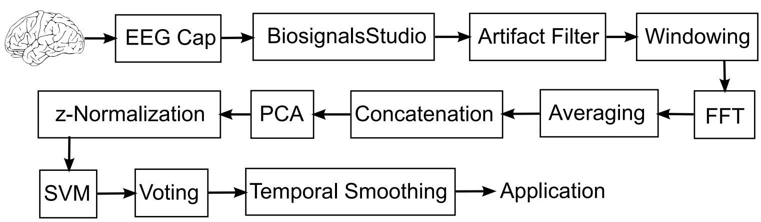

Fig. 1. Processing stages of the workload recognition system

3.1 Recording Setup

An active EEG-cap (Brain Products actiCAP) has been used for recording of EEG data. 16 active electrodes placed at positions FP1, FP2, F3, F4, F7, F8, T3, T4, C3, Cz, C4, P3, P4, Pz, O1, and O2 according to the international 10-20 system [6] have been recorded with reference to FCz. The impedance of each electrode was kept below 20 kΩ during each session. Amplification and A/D-conversion has been done using a 16 channel VarioPort biosignals recording system by Becker Meditec using a sampling rate of 256 Hz.

For data recording and stimulus presentation for the cognitive tests we used BiosignalsStudio (BSS) [7], which is a flexible framework for multimodal biosignal recording that has recently been developed at the Cognitive Systems Lab in Karlsruhe. BSS is also used as input layer for real-time EEG data acquisition during the online application of the proposed workload detection system.

3.2 Pre-processing and Feature Extraction

Contamination by artifacts is a severe problem for a reliable estimation of workload from EEG signals. Predominantly eye movement artifacts and muscular artifacts can be found in EEG signals recorded under non-laboratory conditions. Therefore, we applied a heuristic approach for artifact detection using thresholds on the signal power and its slope to identify artifacts in each data segment. Contaminated data segments are dropped and not used for classification or training.

The incoming stream of raw data from each electrode is cut into segments of one second length overlapping by 0.5 seconds. Each segment is multiplied by a Hamming window function to reduce spectral leakage. Afterwards, the windowed data segments are transformed to frequency domain using FFT. Three adjacent frequency bins are combined by averaging, which reduces noise in the data and lowers the dimensionality of the feature space. Thus, each coefficient in the resulting vector represents a frequency range of 3 Hz. The spectral features of each channel are concatenated to a final feature vector. As EEG data from neighboring channels and frequencies are highly correlated, we apply Principal Component Analysis (PCA) to further reduce the dimensionality of the feature space. A reduction from 224 dimensions (4-45Hz, 16 channels) to 100 dimensions explains 79.5-86.3% (mean 83.3%) of variance in the training data. The final vector for classification consists of z-scores of the PCA coefficients, i.e. normalization of each coefficient by subtracting the mean of each feature and dividing by its standard deviation determined on the training data.

3.3 Recognition and Post-Processing

For recognition we apply Support Vector Machines (SVMs) with linear kernels (LibSVM implementation [8]) to discriminate between low and high workload conditions. The penalty constant to control misclassification is set to $C = 1$ for each subject (parameter optimization by cross-validation did not improve the results). For each time step we calculate a majority vote using k previous predictions. This way the stability of recognition results can be increased, by the cost of temporal resolution. For the experiments presented in this paper we used $k = 3$, which gives a still high temporal sensitivity as the SVM delivers predictions every 500ms, while eliminating outliers and noise in the estimations.

Adaptation of an intelligent interaction system using binary workload estimations is usually unsuitable, as rapid changing of system behavior would appear unnatural to the user. Therefore, workload estimations of longer duration are required in addition to predictions gathered only from small data segments. To provide such information of workload trends with increased stability, we calculate the following workload index, which is a simple linear temporal smoothing of the prediction results:

$$load\_index(x) = \sum_{t=x-(l-1)}^{x} \frac{pred(t)}{l},$$

where x is a particular point of time, l is the smoothing length, and $pred(t)$ is the voted binary prediction by the SVM at time t. The temporal integration results in a workload index value in the range between 0 and 1. For the evaluations on computer work data we chose $l = 30$, i.e. linear smoothing on a time period of 15 seconds.

4 Evaluation

4.1 Cognitive Test Data

For training and evaluation of the system, workload and resting conditions from five subjects have been recorded. All of them are male students or employees of the Karlsruhe Institute of Technology (KIT).

We used the flanker paradigm and the switching paradigm for induction of workload. In both of these cognitive tasks subjects repeatedly react to stimuli presented on a display by pressing one of two keys on a keyboard using their left and right index finger.

During the flanker test, different horizontal arrays of five arrows are displayed (e.g. <<><<). Subjects respond as quickly as possible to the orientation of the middle arrow by pressing the corresponding left or right key. During the switching task, digits are presented on the screen surrounded by a dashed or solid square. A dashed square requires the subjects to indicate whether the stimulus is greater or lower than 5, while subjects need to decide whether the digit is odd or even, when a digit is surrounded by a solid square. Both tests require concentration and alertness and are especially demanding on visual-perceptual information processing. Workload is enhanced by the executive control required to overcome the interference of the presented stimuli in the flanker task and task switching decisions due to multitasking in the switching task.

We decided for these tasks for workload induction, because they allow a widely standardized evaluation of the system as the same stimuli are presented to each person in a controlled fashion. They only require little physical activity that could lead to artifacts caused by musclular activation (EMG) or eye movements. Moreover, they have some behavioral patterns similar to usual computer work (e.g. multitasking) and induce a constant level of workload, which is required to derive features from short time slices of data.

In contrast to the workload data, two resting conditions have been recorded, where the subjects were asked to relax without any activity keeping their eyes open. The workload recognition system is trained on the two classes of data: workload, i.e. subjects performing the flanker or switching task, and resting, i.e. subjects performing no particular task.

Due to temporal effects, such as high correlation of neighboring feature vectors calculated from small segments of EEG data, we did not use cross-validation for our evaluations. Instead, we used one cognitive task and one resting period as training data and evaluated the system on the other cognitive task and the second resting period recorded for each person. These evaluations also show the task robustness of the proposed system. Therefore, two systems were trained and evaluated for each subject by switching training and evaluation data.

Approximately six minutes of resting and workload data have been used for training of the system, with a balanced number of samples for each condition.

Table 1 shows the recognition results for different frequency bands that are usually used in EEG analyses [9], including theta, alpha, beta, and gamma band, as well as a full frequency range (4-45 Hz). Lower and higher frequencies are left out because they are more sensitive to artifacts.

Table 1. Recognition results on different frequency bands

Frequency Band / Subject ID	θ 4-7 Hz	α 8-13 Hz	β 14-38 Hz	γ 38-45 Hz	$\theta - \gamma$ 4-45 Hz
1	69.6%	74.4%	88.4%	96.4%	95.6%
2	81.6%	89.1%	84.5%	85.9%	86.0%
3	57.4%	83.9%	99.8%	99.4%	99.8%
4	50.5%	48.8%	74.4%	77.3%	75.2%
5	63.6%	82.3%	93.2%	94.7%	98.3%
Mean Recognition Rate	64.5%	75.7%	88.1%	90.7%	91.0%
Standard Deviation	11.9	15.9	9.5	9.0	10.3

The results indicate, that especially high frequencies are relevant for the classification task. This hypothesis is supported for example by [10], however we cannot confirm high recognition performance using theta and alpha activity (e.g. as proposed in [1]).

In addition to these person dependent results we used a leave-one-out cross-validation scheme for person independent evaluation of the system, i.e. both constellations for training and evaluation of each person are left out once from the training data for evaluation. Spectral features from frequency range 4-45 Hz again gave the best recognition results. The system delivered a recognition rates between 67.4% and 87.4% (mean 72.2% sd=9.0) using the data of the flanker and switching experiments.

4.2 Computer Work Data

In addition to the cognitive tests, we qualitatively evaluated the system on more realistic data in which subject 3 performs different tasks of common computer work. The classifier is trained on the same data as above, i.e. using data from a cognitive task and a resting period. The recording session consists of the following tasks of computer work:

It starts with a resting period of three minutes, i.e. no interaction with the computer. Next, a one minute long period of mail reading shows moderate workload classification results. Then, six minutes of writing an email have been recorded which show quite high workload. Next, starting approximately at minute ten, the subject has done undemanding internet surfing until minute 13, where the subject has started computer programming until approximately minute 16. Finally, another resting period of approximately one minute length has been recorded.

Figure 2 shows the time course of the load index during the session with the manually marked boundaries between the different periods. The blue curve gives the workload index as predicted from the workload recognizer. The results indicate that boundaries between workload periods can be identified with reasonable accuracy. Smoothing of the binary classification results, reveals not only relaxed and high workload conditions, but also can show moderate workload as in the mail reading condition, where a non extremal worload level is maintained over a period of several minutes. This can be explained by the uncertainty of the classifier in situations of medium workload or situations that repeatedly require high workload for short amounts of time.

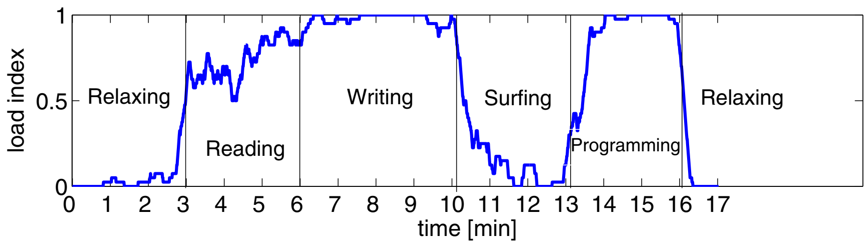

Fig. 2. Smoothed recognition results during different tasks of computer work

We also successfully applied the proposed workload recognizer to data of solving an test exam in computer science - showing high workload. In addition, the system has been applied to EEG recordings in a driving simulator. The results indicate that driving on a highway results in a rather relaxed user state, while demanding traffic situations are indicated by high workload.

5 Conclusion

In this paper we have proposed a system for assessment of workload from EEG data that can easily run in real-time on standard desktop computer or laptop. Excellent recognition results could be shown for person dependent classification for training and evaluation on two standardized cognitive tasks paradigms. Furthermore, we showed in a qualitative evaluation that a system can be trained using cognitive tests and applied to realistic data, such as computer work. The proposed system has been successfully demonstrated live several times, with different subjects and different scenarios, which indicates that it can provide valuable information for the adaptation of a variety of intelligent systems in human-machine interaction.

This work has been supported by the Deutsche Forschungsgemeinschaft (DFG) within Collaborative Research Center 588 “Humanoid Robots - Learning and Cooperating Multimodal Robots” [11].

6 Outlook and Future Work

In this paper we have shown the applicability of a system to detect workload in realistic scenarios. More detailed analyses and systematic evaluations are needed to get more insights of the capabilities and limitations of the system and the proposed workload index.

For a real application in human-machine interaction a major drawback of the proposed system is the fact that the acceptance of users to wear an EEG-cap outside experimental scenarios is rather low. Therefore, we experimented with less invasive wearable devices, which are more fashionable and, for example, integrated in clothes such as a hat or a headband. Furthermore, new electrode technologies, which do not require the use of conductive gel could strongly improve the usability and acceptance of the system.

Additional work needs to address the discrimination of different cortical activation patterns to identify used and available resources of a person. Such information could be helpful for an intelligent system to find a suitable adaptation scheme. For example, when intensive visual but low auditive processing is recognized, a system might provide the same information through an acoustic signal or speech synthesis, instead of using the visual communication channel.

References

1. Gevins, A., Smith, M.: Neurophysiological measures of cognitive workload during human-computer interaction. Theoretical Issues in Ergonomics Science 4(1), 113–131 (2003)
2. Berka, C., Levendowski, D., Cvetinovic, M., Petrovic, M., Davis, G., Lumicao, M., Zivkovic, V., Popovic, M., Olmstead, R.: Real-time analysis of EEG indexes of alertness, cognition and memory acquired with a wireless EEG headset. International Journal of Human-Computer Interaction 17(2), 151–170 (2004)
3. Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Müller, K., Curio, G., Hagemann, K., Bruns, A., Schrauf, M., Kincses, W.: Improving human performance in a real operating environment through real-time mental workload detection. In: Toward Brain-Computer Interfacing, pp. 409–422
4. Honal, M., Schultz, T.: Determine task demand from brain activity. In: International Conference on Bio-inspired Systems and Signal Processing (2008)
5. Putze, F., Jarvis, J., Schultz, T.: Multimodal Recognition of Cognitive Workload for Multitasking in the Car. In: Accepted for International Conference on Pattern Recognition 2010 (2010)
6. Jasper, H.: The 10-20 electrode system of the International Federation. Electroencephalography and Clinical Neurophysiology 10, 371–375 (1958)
7. Heger, D., Putze, F., Amma, C., Wand, M., Plotkin, I., Wielatt, T., Schultz, T.: BiosignalsStudio: A flexible Framework for Biosignal Capturing and Processing. In: Dillmann, R., et al. (eds.) KI 2010. LNCS (LNAI), vol. 6359, pp. 33–39. Springer, Heidelberg (2010)
8. Chang, C., Lin, C.: LIBSVM: a library for support vector machines (2001)
9. Zschocke, S.: Klinische Elektroenzephalographie. Springer, Heidelberg (2002)
10. Koles, Z., Flor-Henry, P.: Mental activity and the EEG: task and workload related effects. Medical and Biological Engineering and Computing 19(2), 185–194 (1981)
11. CRC588: Collaborative research center 588 humanoid robots - learning and cooperating multimodal robots, `http://www.sfb588.uni-karlsruhe.de/`