EEG-Based Measure of Cognitive Workload during a Mental Arithmetic Task

Brice Rebsamen¹, Kenneth Kwok¹, and Trevor B. Penney²

¹ Temasek Laboratories, National University of Singapore, Singapore

² Dept of Psychology, National University of Singapore, Singapore
{tslrbr, kenkwok, penney}@nus.edu.sg

Abstract. We collected EEG data from 16 subjects while they performed a mental arithmetic task at five different levels of difficulty. A classifier was trained to discriminate between three conditions: relaxed, low workload and high workload, using spectral features of the EEG. We obtained an average classification accuracy of 62%. A continuous workload index was obtained by low-pass filtering the classifier's output. The average correlation coefficient between the resulting workload index and the difficulty level of the task was 0.6.

Keywords: EEG, mental workload, arithmetic, cognitive state.

1 Introduction

Excessive cognitive workload can lead to delayed information processing, response errors, or a failure to respond, but a too low cognitive workload can also lead to errors due to boredom. Thus, the ability to measure cognitive workload in real-time might be valuable in enforcing industrial safety, in improving human-computer interfaces, and designing appropriate adaptive automation strategies [1, 2].

Analysis of workload using EEG data has been attempted in several studies. Gevins and Smith [3] discriminated three workload levels in a task involving sequential memorization of stimuli using spectral power in the theta and alpha frequency bands. Kohlmorgen et al. [4] designed an online system to measure workload from EEG, used it to manage high workload situations during driving and showed that this procedure improved the reaction times of drivers. Putze, Jarvis and Schultz [5] proposed a multi-modal approach, taking into account biological signals such as skin conductance, pulse and respiration and EEG, to classify several levels of workload during a driving task.

In this paper we present the results of EEG data classification of a dataset obtained from 16 participants who performed a mental arithmetic task with five levels of difficulty. A continuous workload index based on the classifier's output is proposed and evaluated.

2 Method

Data were collected from 16 students, 9 males and 7 females, with a mean age of 21.5 (SD=1.5) years.

EEG data was recorded from 16 channels (FP1, FP2, F3, Fz, F4, T7, C3, Cz, C4, T8, P3, Pz, P4, O1, Oz, O2) at 256 Hz with an ActiveTwo system from Biosemi, referenced to the nose.

The task consisted of summing mentally two figures presented on the computer screen and comparing the result to a proposed answer, also presented onscreen. There were 5 possible difficulty levels. At level 1, the problems consisted of summing two one-digit figures, and each subsequent level included an extra digit such that at level 5 the problems consisted of summing two three-digit figures.

The experiment was segmented in blocks of one minute, with a thirty-second rest period between each block. The rest period was introduced to avoid excessive fatigue, so that differences between block would be on workload only and not fatigue. Within a block, all problems were of the same difficulty level (1 to 5).

After three repetitions of the six difficulty levels, a slide-show of landscape pictures (one picture every thirty seconds) was presented for five minutes to allow the participant to relax. Pictures were favored upon a fixation cross to avoid boring the subject too much. The whole session was then repeated (18 blocks of mental arithmetic and 5 minutes of relaxation).

3 Results

3.1 Behavior Analysis

For each subject, we observed an increase of the mean response time with the difficulty level. Similarly, subjects made more errors for higher levels of difficulty. Although there were clear performance differences between subjects, they all showed the same behavior: response time and error rates increased with the difficulty level.

We also investigated the effect of time on task on the response time and the error rate. For each subject we computed the difference between the first and last third of the trials. No statistically significant difference was found for either response time ($p=0.06$) or error rate ($p=0.26$). Hence, subjects did not seem to be affected by time on task (fatigue, boredom, etc.).

3.2 Classification

EEG analysis was done using PyMVPA [6], a multivariate pattern analysis package for Python. The continuous EEG was first epoched in three second windows, with a two second overlap between consecutive windows. Epochs were labeled with either the difficulty level (LVL1 to LVL5) or as RELAXATION. This yielded approximately 2000 epochs, around 300 per label (900 for the RELAXATION condition).

The power spectrum for each epoch was computed using the fast Fourier transform (FFT) and a Hanning window. The power spectrum was reduced to the total power in the following five frequency bands: delta (1Hz-4Hz), theta (4Hz-8Hz), alpha (8Hz-14Hz), beta (14Hz-35Hz) and gamma (35Hz-45Hz). The resulting feature vector hence contained 80 features (16 channels and 5 frequency bands).

A SMLR (Sparse Multinomial Logistic Regression) classifier [7] was trained to discriminate between three classes: relaxed, low workload (LVL1 and LVL2), and

high workload (LVL3 to LVL5). For each class, a workload value was assigned: 0 for the relaxed class, 2.5 for the low workload class, and 5 for the high workload class.

The system was evaluated using a two-fold validation procedure: the classifier was first trained on the first half of the data (up to after the first five minute relaxation period), and applied to the second half; it was then trained with the second half and applied to the first half. Across subjects, the mean accuracy was 62% (SD=4.3), which is significantly higher than the 33% chance level ($p < 0.001$).

The classifier assigned to each epoch a discrete workload value corresponding to the output class. These discrete values were then low passed filtered over the previous 10 epochs, which yielded a continuous index of workload, ranging from 0 to 5 (an example of the result is provide in Fig. 1). We measured the correlation between the continuous workload index, and the task difficulty, which was taken as the difficulty level for the task condition and as 0 for the other conditions. Across subjects, the mean correlation coefficient was 0.6 (SD=0.08), which is significantly higher than chance ($p < 0.001$).

Fig. 1. An example of the resulting workload index: the task difficulty level is in black, the workload index in blue

4 Discussion

We classified EEG data from 16 participants who were performing a mental arithmetic task with 5 levels of difficulty. For each participant, a classifier was trained to classify spectral features into 3 classes: relaxed, low workload, and high workload. The resulting accuracy was 62% on average across participants, which was significantly higher than chance.

A continuous workload index was created by assigning a workload value to each class and low-pass filtering the result. The correlation between the task difficulty level and the resulting workload index was 0.6 on average across participants, which was also significantly higher than chance.

These results constitute a first step toward a real-time system to measure cognitive workload from EEG. However, development of an operational workload measurement system requires much additional work. In particular the following points will require our attention:

- ▲ intra-subject consistency: once the system has been trained for a participant, does the prediction accuracy degrade with time?
- ▲ inter-subject consistency: can a system trained with data from one participant be used with another participant? How does it affect the performance?
- ▲ task independence: the current results were obtained with participants performing mental arithmetic in controlled conditions. What will the system give when the participant is performing a real-world task?
- ▲ robustness to noise: how will the system behave in a non-controlled situation with more artifact sources (movements, sweat, etc.)?

References

1. Parasuraman, R., Mouloua, M., Molloy, R.: Effects of adaptive task allocation on monitoring of automated systems. *Human Factors* 38, 665–679 (1996)
2. Wiener, E.L.: Beyond the sterile cockpit. *Human Factors* 27, 75–90 (1985)
3. Gevins, A., Smith, M.: Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science* 4(1), 113–131 (2003)
4. 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. MIT Press, Cambridge (2007)
5. Putze, F., Jarvis, J., Schultz, T.: Multimodal Recognition of Cognitive Workload for Multitasking. In: *The Car. Int. Conf. on Pattern Recognition* (2010)
6. Hanke, M., Halchenko, Y.O., Haxby, J.V., Pollmann, S.: Statistical Learning Analysis in Neuroscience: Aiming for Transparency. *Frontiers in Neuroscience*, 4 (2010)
7. Krishnapuram, B., Carin, L., Figueiredo, M.A.T., Hartemink, A.J.: Sparse multinomial logistic regression: Fast algorithms and generalization bounds. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 957–968 (2005)