International Journal of Psychophysiology 98 (2015) 157–166

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

Towards an effective cross-task mental workload recognition model using electroencephalography based on feature selection and support vector machine regression

Yufeng Ke [a,b], Hongzhi Qi [a,b,*], Lixin Zhang [a,b], Shanguang Chen [c], Xuejun Jiao [c], Peng Zhou [a,b], Xin Zhao [a,b], Baikun Wan [a,b], Dong Ming [a,b,*]

[a] Department of Biomedical Engineering, College of Precision Instrument and Optoelectronics Engineering, Tianjin University, No. 92 Weijin Road, Nankai District, Tianjin 300072, PR China
[b] Tianjin Key Laboratory of Biomedical Detecting Techniques and Instruments, Tianjin, PR China
[c] National Laboratory of Human Factors Engineering, China Astronaut Research and Training Centre, Beijing, PR China

ARTICLE INFO

Article history:
Received 27 March 2015
Received in revised form 21 August 2015
Accepted 17 October 2015
Available online 19 October 2015

Keywords:
Mental workload
Electroencephalography
Cross-task
Recursive feature elimination
Support vector machine

ABSTRACT

Electroencephalographic (EEG) has been believed to be a potential psychophysiological measure of mental workload. There however remain a number of challenges in building a generalized mental workload recognition model, one of which includes the inability of an EEG-based workload classifier trained on a specific task to handle other tasks. The primary goal of the present study was to examine the possibility of addressing this challenge using feature selection and regression model. Support vector machine classifier and regression models were examined under within-task conditions (trained and tested on the same task) and cross-task conditions (trained on one task and tested on another task) for well-trained verbal and spatial n-back tasks. A specifically designed cross-task recursive feature elimination (RFE) based feature selection was used to handle the possible causes responsible for the deterioration of the performance of cross-task regression model. The within-task classification and regression performed fairly well. Cross-task classification and regression performance, however, deteriorated to unacceptable levels (around chance level). Trained and tested with the most robust feature subset selected by cross-task RFE, the performance of cross-task regression was significantly improved, and there were no significant changes in the performance of within-task regression. It can be inferred that workload-related features can be picked out from those which have been contaminated using RFE, and regression models rather than classifiers may be a wiser choice for cross-task conditions. These encouraging results suggest that the cross-task workload recognition model built in this study is much more generalizable across task when compared to the model built in traditional way.

1. Introduction

1.1. Overview

As an ergonomic concept, mental workload (MW) is usually defined as the mental resources occupied or an intervening variable modulating the tuning between the demands of environment and the capacity of operator (Kantowitz, 1987). It rises following the increasing task difficulty if the operator makes an effort. Maintaining task performance within an acceptable range under such conditions for a prolonged period of time, then becomes increasingly difficult. Participants are usually instructed to devote their efforts to conduct tasks of different demands (difficulties) to induce different MW levels in existing MW studies. MW-based adaptive automation (AA) has shown considerable potential to improve performance in human–machine interaction, especially in instances where high mental demand is required (Kaber et al., 2002; Wilson and Russell, 2007; Zander and Kothe, 2011). However, robustly detecting MW is the sticking point for MW-based AA to be used in practical situations and remains to be a challenge.

1.2. EEG as an index of brain function state

Among the psychophysiological measures, a primary advantage of using brain activity (EEG features) to infer MW is that EEG offers a relatively good temporal resolution of cognitive activity with resolution significantly under a minute and can be used as one of the most direct, nonintrusive and portable measures of the central nervous system (Laine et al., 2002). Studies on EEG spectral variables and other physiological variables based MW recognition have suggested or concluded that EEG is the most sensitive and promising workload indicator (Baldwin and Penaranda, 2012; Berka et al., 2007; Brouwer et al., 2012; Christensen et al., 2012; Taylor et al., 2010; Wang et al., 2012). Several studies reported a correlation between work or memory load

* Corresponding authors.
E-mail addresses: qhz@tju.edu.cn (H. Qi), richardming@tju.edu.cn (D. Ming).

http://dx.doi.org/10.1016/j.ijpsycho.2015.10.004
0167-8760/

and power in certain EEG frequency bands. The evidence of alpha power variations at different workload levels may lie in its link to arousal level, idling and cortical inhibition (Brouwer et al., 2012; Fink et al., 2005; Pfurtscheller et al., 1996). Several papers have reported the association between theta power and mental effort, and task requirements (Esposito et al., 2009; Pesonen et al., 2007). Besides, delta, beta and gamma bands have also been reported to respond to varying MW (Baldwin and Penaranda, 2012; Christensen et al., 2012; Laine et al., 2002; Michels et al., 2010; Pesonen et al., 2007). Subject- and task-specific online workload detectors have been successfully built with power of frequency band 3–15 Hz for auditory task and mental calculation task (Kohlmorgen et al., 2007). Seven bands among 0.5–100 Hz were used by Christensen et al. with considerable results in a functional state classification study (Christensen et al., 2012). Wide frequency range has also been used successfully to classify visuomotor workload in a driving simulator (3–108 Hz) (Dijksterhuis et al., 2013) and to classify memory load across relax and stress contexts (1–90 Hz) (Mühl et al., 2014). As opposed to brain–computer interface studies, in which EEG features are usually focused in fairly narrow bands, spontaneously generated EEG features in studies of the brain's functional state lie in such a wide range. In order to build a robust model with high generalization and invariance, researchers must pick out the most sensitive and the most stable features.

1.3. The challenges of building a robust workload recognition model and the importance of cross-task study

The first crucial step in implementing EEG-based adaptive systems, however, may be the robustness of the MW recognition model. An important but unresolved question is the extent to which classification-based decoding strategies may generalize over time, across subjects and to new situations (Haynes and Rees, 2006). It is impossible to apply in practice and it even may be unreasonable to believe that a MW recognition model has been actually built if a model trained on the data of a specific subject performing a specific task at specific time can only recognize MW level of the same subject performing the same task at the same time. The limitations of a model may imply some unknown scientific questions to be discovered. Just as said, “the ‘holy grail’ of workload classifiers would be able to predict the workload level of any subject performing any task on any given day” (Wang et al., 2012). It clearly points out several challenges to the researchers in this field. However, most of the current physiological-measure-based workload classifiers are subject-dependent, task-dependent and time-limited. It seems that these challenges have not been substantially addressed until very recently.

These three challenges, the problem of variability in the accuracy of machine learning-based pattern classification across tasks, time and people, have been concerned in the three papers of a special issue on neuroergonomics in *NeuroImage* (Parasuraman et al., 2012). However, the cross-task workload classification in which classifiers are trained on EEG features under one task and applied in other tasks seems to be a more knotty challenge. It will be very convenient to build and use a MW estimator in practical applications if it can handle multi-task. Time and resources would be saved if we do not need to train MW estimator for each task we may encounter in daily work. Baldwin and Penaranda have tried to build a cross-task workload classifier using EEG with artificial neural networks across three different working memory tasks, and found that cross-task classification accuracies were significantly lower compared to those within-task (Baldwin and Penaranda, 2012). They further discussed the cross-session and cross-task “costs” to classification, and found that classification performance suffered greater from cross-session than from cross-task (Penaranda and Baldwin, 2012). A recently published cross-task MW study attempted to handle the challenge by training SVM on three working memory tasks (go/no-go, verbal n-back and reading span), inducing nearly the same MW states and types of neural processing as in the two complex learning tasks (working on diagram and algebra problems) used to test the classifier (Walter et al., 2013). However, the cross-task classification performances were not significant over chance level although several satisfactory results were obtained for some subjects. The authors discussed that poor performances may result from the non-stationary patterns caused by advancing levels of difficulty order, the use of different neural structures and executive functions due to the different nature of the tasks, and the varying absolute difficulty across tasks due to the relative MW manipulated within-task.

Throughout the existing MW studies, including within- and cross-task designs, it can be summarized that the degradation of cross-task classification performance may be at least due to (but not limited to) several hypotheses, including:

a) For a specific individual, the workload levels of different tasks should be difficult to exactly match because the psychometrically difficulty levels mismatch between tasks and/or subject's capacity vary between tasks (Baldwin and Penaranda, 2012). So the classifiers that need exact match between workload levels of different tasks and that evaluate performance with classification accuracy would inevitably cause errors under cross-task conditions.
b) EEG patterns of a specific individual may vary significantly across tasks because different tasks relying on different neural structures or types of processing (Penaranda and Baldwin, 2012). Thus, the variations of EEG features should be modulated by the variations of both task type and workload level in cross-task condition. The task type related EEG patterns would strongly affect the performance of cross-task MW recognition.
c) The temporal effect due to the rapid changes in spontaneous EEG results from circadian effects, fatigue and so on (Penaranda and Baldwin, 2012).

The first factor can be partially resolved in a laboratory study with difficulty-well-controlled tasks, but the practicality of such a study may be disputed since task difficulty under real conditions cannot be controlled and would barely match the task used in model building. More importantly, the subject's capacity for different tasks may differ greatly. Therefore, it may be difficult to effectively match workload levels between different tasks. The second factor may be a reflection of the actual different natures of the tasks, but irrespective of these differences, the changes of workload levels remain consistent and there should be similar neuronal mechanisms reflected in on the EEG. It is therefore possible to pick out workload-related features from others with some data mining methods. The third factor may be a reflection of insufficient sampling of a temporally dynamic EEG profile (Penaranda and Baldwin, 2012) or an inability to discover certain features that remain stable over time but change with workload. This can be overcome by sampling sufficient data and extracting exact features. Thus identifying a single set of features that are only workload-related, stable enough over time and applicable for all tasks is highly desired.

1.4. The current study

As mentioned above, it should be a substantial step towards the development of applied EEG-based MW recognition model if a model trained on one task can be enabled to handle another. The primary goal of the present investigation can be concretely described as to determine the probability of cross-task MW recognition based on EEG with well-controlled verbal and spatial n-back tasks. This study was inspired by the result of the novel cross-task study (Baldwin and Penaranda, 2012) which tried to determine the effects of the task type and MW on EEG and whether the performance of cross-task MW recognition can be improved with proper feature selection and modeling method. Based on the three possible causes of the deterioration of cross-task