ARTICLE IN PRESS

YNIMG-08590; No. of pages: 6; 4C:

NeuroImage xxx (2011) xxx–xxx

Contents lists available at SciVerse ScienceDirect

NeuroImage

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

Cross-subject workload classification with a hierarchical Bayes model

Ziheng Wang [b], Ryan M. Hope [a,*], Zuoguan Wang [b], Qiang Ji [b], Wayne D. Gray [a]

[a] *Department of Cognitive Science, Rensselaer Polytechnic Institute, Troy, NY 12180, USA*
[b] *Department of Electrical, Computer, & Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA*

ARTICLE INFO

Article history:
Received 30 April 2011
Revised 26 July 2011
Accepted 27 July 2011
Available online xxxx

Keywords:
Workload classification
EEG
Hierarchical Bayes model
Artificial neural network

ABSTRACT

Most of the current EEG-based workload classifiers are *subject-specific*; that is, a new classifier is built and trained for each human subject. In this paper we introduce a *cross-subject* workload classifier based on a hierarchical Bayes model. The cross-subject classifier is trained and tested with data from a group of subjects. In our work, it was trained and tested on EEG data collected from 8 subjects as they performed the *Multi-Attribute Task Battery* across three levels of difficulty. The accuracy of this cross-subject classifier is stable across the three levels of workload and comparable to a benchmark subject-specific neural network classifier.

Introduction

As cognitive workload increases, maintaining task performance within an acceptable range becomes more difficult. Increased cognitive workload may demand more cognitive resources than the human brain has available (Norman and Bobrow, 1976a,b) and may also result in physiological stress (Gaillard, 1993).Both cognitive overload and stress may result in performance degradation and errors. An objective measure of workload could be used to evaluate alternative system designs, allocate workload appropriately to minimize overload and stress, and intervene in real time before operators become overloaded while performing safety-critical tasks (Byrne and Parasuraman, 1996).

Current state of workload measures

Traditionally, workload is assessed by questionnaires which are quantified through statistical techniques such as factor loading, discriminant analysis, and correlation/covariance analysis (Hart and Staveland, 1988). Although progress has been made, there are no globally accepted methods for detecting and measuring cognitive workload (Lysaght et al., 1989; Noyes and Bruneau, 2007; Rubio et al., 2004). In addition, subjective measures are invasive and cannot be obtained in real-time as they require interrupting the task to complete a questionnaire. As a result, many researchers have investigated physiological measures such as heart rate variability, galvanic skin response, pupillometry and electroencephalography (EEG), to predict workload. EEG promises to provide the applied community with an objective and relatively unobtrusive means for measuring workload. However, cashing in on this promise requires the development of new and innovative quantitative methods for analyzing and interpreting EEG data.

EEG has been used extensively to examine the changes in the brain's electrical activity in response to cognitive activity (Gevins et al., 1998; Gevins and Smith, 2003). The main assumption is that if brain-state classifiers can be found, they can then be used as a brain-computer interface (BCI) or input to adaptive automation that detects operator mental workload in real time (Wilson and Russell, 2007). A number of different classifiers have already been applied to predict workload with EEG data, such as linear discriminant analysis, support vector machines and artificial neural networks. Among these classifiers, artificial neural networks (NN) have shown success discriminating at least two levels of cognitive workload (Wilson et al., 2009, 2010; Wilson and Russel, 2003a,b). However, much work still needs to be done in the development of quantitative methods for analyzing and interpreting EEG data.

Challenges to EEG-based workload classifiers

The 'holy grail' of workload classifiers would be able to predict the workload level of any subject performing any task on any given day. The development of such a classifier would be quite an ambitious feat given the number of challenges that would need to be overcome. Such a classifier would need to be robust enough to handle large variations in input data from a wide variety of sources. Currently, all of the existing workload classifiers are subject-specific, meaning new classifiers are trained for each subject. Classifiers trained on data from one subject do

* Corresponding author at: Department of Cognitive Science, Rensselaer Polytechnic Institute, Carnegie Building (rm 108), 110 8th Street, Troy, NY 12180, USA. Fax: +1 518 276 8268.
E-mail address: hoper2@rpi.edu (R.M. Hope).

1053-8119/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2011.07.094

Please cite this article as: Wang, Z., et al., Cross-subject workload classification with a hierarchical Bayes model, NeuroImage (2011), doi:10.1016/j.neuroimage.2011.07.094

not generalize to other subjects well. Current classifiers also have difficulty generalizing to performance of the same subject on different sessions and days (Wilson et al., 2010).

Overview

In the present article, we introduce a cross-subject workload classifier for a group of subjects based on the hierarchical Bayes model. This work is a substantial step forward in the development of EEG-based workload classifiers as it would enable training a classifier once that could handle multiple subjects. We will compare our proposed model with the dominant technique NN in this field in order to demonstrate its advantage. In the following sections, we will show that our hierarchical Bayes model when trained on data from multiple subjects is comparable to the subject-specific neural network. We will also show that the accuracy of our hierarchical Bayes model is stable across three levels of workload, in comparison with neural networks, which have not been demonstrated capable of accurately classifying more than two levels of workload. The novel part of our proposed model lies in its ability to explicitly model the variations across subjects at each level of workload and apply the captured variation to enhance the classification.

Models

In this section we introduce a hierarchical Bayes model for workload classification. Compared to traditional models such as neural networks and support vector machines, it has the advantage of modeling the inner relationships, incorporating the prior knowledge as well as accounting for uncertainties in data, through probabilistic theory.

The naive Bayes classifier will be discussed first to illustrate how the Bayes and probability theories function in the model. Following this discussion we describe our cross-subject hierarchical Bayes model that can effectively classify workload levels of multiple subjects.

Naive Bayes classifier

The naive Bayes classifier (NB) is a simple classifier based on the Bayes' theorem. Its structure is shown in Fig. 1a, where the C node represents different classes and $X_1, X_2, \cdots, X_n$ represent different components or features of a sample. The causal directed link from C to X_i is captured by $P(X_i|C)$, i.e. the probability of X_i given the observation of a certain class C. For instance, if X_i is a feature of a high workload sample, $P(X_i|C=High)$ would be expected to be greater than $P(X_i|C=Low)$. $P(X_i|C)$ is also called the likelihood, and typically follows the Gaussian distribution $N(\mu,\sigma)$, where μ is the mean and σ is the standard deviation. NB assumes all the feature nodes are independent of each other given the class.

For simplicity, $X_1, X_2, \cdots, X_n$ are often represented by only one node X, as is shown in Fig. 1b. In this case X is a vector consisting of all the features of a sample and $P(X|C)$ follows the multivariate Gaussian distribution. Fig. 1c shows $P(X|C)$ when X is a 2 dimensional vector. We can see that $P(X|C)$ gradually decreases as X goes far away from its center.

Given a test sample $X = (X_1, X_2, \cdots, X_n)$, the classification result C^* is determined by the posterior probability $P(C|X)$, the probability that a test sample X belongs to the class C, with Eq. (1).

$$C^* = \underset{C}{argmax}\, P(C|X) \tag{1}$$

$P(C|X)$ can be further transformed using the chain rule and Bayes' theorem into Eq. (2), where α is a normalization constant.

$$\begin{aligned} P(C|X) &= \alpha P(X|C)P(C) \\ &= \alpha \prod_{i=1}^{n} P(X_i|C)P(C) \end{aligned} \tag{2}$$

In our case, the class node represents three workload conditions and the feature nodes $(X_1, X_2, \cdots, X_n)$ represent the magnitude of EEG frequency bands. Despite its naive design and apparently over-simplified assumptions, NB has worked quite well in many complex real-world situations such as alert correlation (Benferhat et al., 2008), intrusion detection (Axelsson, 2004; Gowadia et al., 2005; P.M. and Patra, 2007; Puttini et al., 2002), text classification (Kamruzzaman and Rahman, 2010) and medical diagnosis (Qu et al., 2010). Many of the properties of NB can be found in Domingos and Pazzani (1997), Rish (2001), Zhang (2004).

In summary, NB works in two simple stages:

1. In the training stage, the likelihood $P(X|C)$ (the probability of the training sample X given their class C) is estimated with respect to the training data;
2. In the testing stage, based on the posterior probability $P(C|X_{test})$, a decision whether the test sample X_{test} belongs to a class C is made, using Eq. (1).

This basic theory can be easily extended to more complex Bayes models.

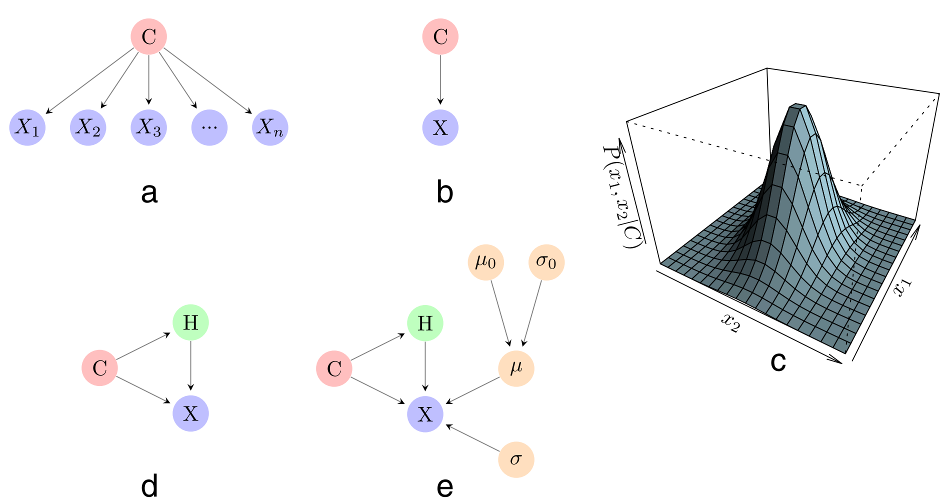

Fig. 1. The evolution from a naive Bayes classifier to the hierarchical Bayes model: (a) a naive Bayes classifier, where C represents the class node and X represents the feature node; (b) A simplified graphical model of naive Bayes classifier; (d) a naive Bayes classifier plus a hidden node; (e) A Hierarchical Bayesian model where μ_0 and σ_0 describes the prior distribution of μ; (c) A visualization of the 2 d Gaussian distribution.

Please cite this article as: Wang, Z., et al., Cross-subject workload classification with a hierarchical Bayes model, NeuroImage (2011), doi:10.1016/j.neuroimage.2011.07.094

Hierarchical Bayes model

In an attempt to deal with the large amount of between-subject variation present in a cross-subject workload classifier we created a hierarchical Bayes model. Figs. 1b–e illustrate the evolution of the naive Bayes classifier into the proposed hierarchical Bayes model. In the following section we will describe step by step, how our proposed model is designed to work, its mathematical details and the intuition behind the design.

In the first step, a discrete hidden node H is embedded into NB, as is shown in Fig. 1d. The value of H ranges from 1 to K, representing K different hidden states. The basic idea is that the EEG signals associated with each level of workload are generated from a certain mixture of hidden states. For example, if H stands for different subjects, EEG signals can be treated as being generated from different subjects with different probabilities. Likewise EEG signals are generated from different trials with different probabilities if H stands for the trial. Generally speaking, H can represent any factor that is likely to cause variations and these factors are typically unknown. Mathematically, the hidden states can be interpreted as transformed features, resulted from projecting all EEG features in all frequencies to a latent space. Hence each dimension in the hidden subspace is a function of all EEG frequencies instead of an individual frequency. Each hidden state does not relate to any specific frequency feature. By connecting C to H, the variations are captured for each level of workload and the generation of the EEG signals can be interpreted as follows:

1. Choose a workload level C with probability $P(C)$;
2. For the chosen workload, choose a hidden state $H=i$ from 1 to K with probability $P(H=i|C)$;
3. Generate EEG signal/sample X from this hidden state and the current workload with a probability $P(X|H=i,C)$.

The likelihood $P(X|C)$ is calculated by marginalizing over all the possible hidden states with Eq. (3).

$$P(X|C) = \sum_{i=1}^{K} P(X|H=i,C)P(H=i|C) \tag{3}$$

We assume that for each hidden state, $P(X|H=i)$ follows the Gaussian distribution $N(\mu_i,\sigma_i)$ and $P(H=i|C)$ follows the multinomial distribution. Thus $P(X|C)$ is a mixture of Gaussian distribution (Xu and Jordan, 1996) in this case.

The between-subject and other unknown variations can be well modeled by decomposing the EEG signal to different hidden states. Intuitively, more hidden states will lead to stronger power in dealing with variations. A typical way to decide the number of hidden states is through cross validation within the training data, which will be discussed in a later section.

While the model gains stronger ability to deal with variability by adding a hidden node, such a model is more sensitive to the noise in the data. In other words, large amounts of noise in the EEG data could be misinterpreted as hidden states as well. It leads us to move to the second step, where a constraint will be imposed upon the variations to alleviate the risk of over-fitting.

The idea is that hidden states should not depart significantly from their shared characteristics, which can be realized by the model shown in Fig. 1e. Extending Fig. 1d, the parameters μ and σ that are used to describe $P(X|H,C)$ are added next to X. Two higher level nodes μ_0 and σ_0 are then pointed to node μ and make the model hierarchical (Gelman and Hill, 2006), denoting that the mean values of each hidden source follows another Gaussian distribution $N(\mu_0,\sigma_0)$ as shown in Eqs. (4) and (5). Intuitively speaking, μ_0 and σ_0 capture the commonalities and the hidden components (μ's) are thus restricted to a certain area close to μ_0.

$$P(X|H) \sim N(\mu,\sigma) \tag{4}$$

$$P(\mu|\mu_0,\sigma_0) \sim N(\mu_0,\sigma_0) \tag{5}$$

Finally, the training procedure of such a hierarchical model will be performed in two steps:

1. Learning the hyper-parameters μ_0 and σ_0. In our case, μ_0 is estimated as the mean vector of all the training data within this workload and the covariance matrix σ_0 is set to be the identity;
2. Learning all the other parameters $\theta=\{\mu,\sigma,P(H|C),P(C)\}$. Since we have unknown hidden nodes, EM algorithm (Dempster et al., 1977; Xu and Jordan, 1996) is applied here to estimate parameters instead of the traditional Maximum Likelihood Estimation. EM algorithm consists of two steps: In the E step, the hidden node is uncovered by their expected posterior probability $P(H|X,\theta)$, given all the training data X; In the M step, parameters are estimated by maximizing the expected posterior probability of the parameters given training data, as shown in Eq. (6), where λ is a constant, controlling the weight of the constraint, and $logN(\mu|\mu_0,\sigma_0)$ is the regularization term.

$$\max_{\theta} \log P(\theta|X) \propto \log P(X|\theta) + \lambda \log N(\mu|\mu_0,\sigma_0) \tag{6}$$

During testing, given a test sample X_{test} the posterior probability $P(C|X_{test})$ is then calculated with Eq. (7).

$$P(C|X_{test}) \propto \sum_{i=1}^{K} P(X_{test}|H=i,C)P(H=i|C)P(C) \tag{7}$$

Experiment

All of the data used in the present article comes from a previously published study, Wilson et al. (2010), which is available upon request.[1]

Participants and stimuli

Eight participants (3 males; mean age 21.1 years) performed the Multi-Attribute Task Battery (MATB) (Comstock and Arnegard, 1992). The MATB is commonly used in laboratory studies of operator performance and workload (Fairclough et al., 2005; Harris et al., 1995; Wilson and Russel, 2003b). It incorporates tasks analogous to activities that pilot perform in flight including a tracking task, monitoring gauges and warning lights, air traffic control communications, and resource allocation tasks (fuel pumps) all of which are performed concurrently in a continually changing task environment (see Fig. 2).

Procedure

Participants performed the MATB on five separate sessions spread over the course of a month. The five sessions were separated by 1 day, 1 week, 3 weeks and 4 weeks. The demands of each subtask were varied so that three levels of overall MATB difficulty were available. In an attempt to reduce learning effects, participants were trained until performance scores reached asymptote with minimal errors. Each day's session consisted of three trials where a trial was comprised of a low, medium and high difficulty block. Each block lasted 5 min and the order of blocks within each trial was random. Three of the participants did not fully complete all of the trials on day 3. For this

[1] To request a copy of the MATB_AF dataset contact Justin R Estepp; email: Justin.Estepp@wpafb.af.mil, address: 711 HPW/RHCP, 2255 H Street, Building 33, Wright-Patterson AFB, OH 45433–7022, United States.

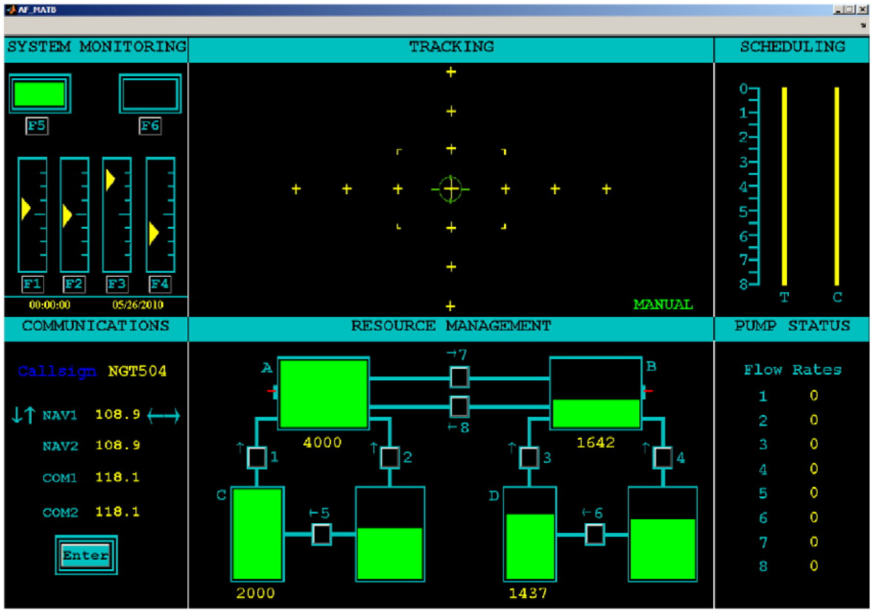

Fig. 2. An example of the MATB environment. Starting in the top-left corner moving clockwise are the following subtasks: system monitoring, tracking scheduling, communications, resource management, pump status.

reason, day 3 was excluded resulting in 12 complete trials for each participant.

EEG

Nineteen EEG channels, using the International 10–20 montage (Jasper, 1958), were collected using the MICROAMPS system from SAM Technologies, Inc. MICROAMPS has default high-pass and low-pass filters at 0.05 Hz and 100 Hz, respectively, and a sampling rate of 256 Hz. The 19 EEG channels were referenced to the left mastoid. Additional VEOG and HEOG channels were also recorded. VEOG was a bipolar channel with electrodes placed above and below the left eye. HEOG was also a bipolar channel with electrodes placed outside the outer canthus of each eye. Impedances for the EEG channels were all below 5 kΩ and impedances for the VEOG and HEOG channels were all below 15 kΩ.

Feature selection

The EEG data was down-sampled to 128 Hz and no artifact correction or rejection procedures were performed prior to analysis. Discrete-time short-term Fourier transform (STFT) was performed on the down-sampled EEG data using 40 second windows with 35 s of overlap. No taper function was applied to the windows. The magnitude of the 5 standard clinical bands (delta [2–4 Hz], theta [5–8 Hz], alpha [9–13 Hz], beta [14–32 Hz] and gamma [33–43 Hz]) as well as two expanded gamma bands ([33–57 Hz] and [63–100 Hz]) from the 19 sites were used resulting in 133 input features to the classifiers.

Classifiers

Five different classifiers are compared in our experiment: the neural net and naive Bayes classifiers trained and tested on individual subjects (NN1 and NB1), and the neural net, naive Bayes classifier and the hierarchical Bayes classifier trained on multiple subjects and tested on individual subjects (NN8, NB8 and HNB8). All the models are multi-class classifiers trained on all the three workload levels.

The neural network classifier used for comparison in our experiment was based on the same setup as the neural network used in Wilson et al. (2009). The neural network had one hidden layer with the number of hidden nodes equal to the number of features use as input to the model. The output layer contained three nodes, one for each workload level. The parameters of neural network were also tuned so that the results reflected the best performance possible.

Functionally, alpha oscillations are interpreted as an idling rhythm that diminishes during mental activity and have been found to be negatively correlated with the BOLD response in the parietal and frontal cortex regions (Laufs et al., 2003). In other words, alpha magnitude can be thought of as being inversely proportional to workload.

Model training

Models were trained and tested using a fivefold cross-validation setup. For the purpose of testing the performance of the models with little training data, only one fifth, instead of four fifths, of the EEG data from each trial was randomly sampled for training. The data not selected for training was used for testing. Data was sampled evenly across workload blocks, and for the models including multiple subject data, evenly across subjects. This procedure was repeated for each trial. For the neural network classifiers, one fifth of the training data was randomly selected as the validation set and Scaled Conjugate Gradient algorithm (Moller, 1993) was used to train the net.

Results

In the following results, NN1 and NB1 represent the neural net and naive Bayes classifiers trained and tested on individual subjects. NN8 and NB8 represent the neural net and naive Bayes classifiers trained on multiple subjects and tested on individual subjects. HNB8 represent hierarchical Bayes model trained on multiple subjects and tested on individual subjects.

Hidden states

In order to determine the optimal number of hidden states for the HNB8 model, a validation set was randomly sampled from the training data. The accuracy rate on the validation set of HNB8 model was then measured as the number of hidden states was increased from 1 to 40 (see Fig. 3). For the hierarchical Bayes model trained on data from

Please cite this article as: Wang, Z., et al., Cross-subject workload classification with a hierarchical Bayes model, NeuroImage (2011), doi:10.1016/j.neuroimage.2011.07.094

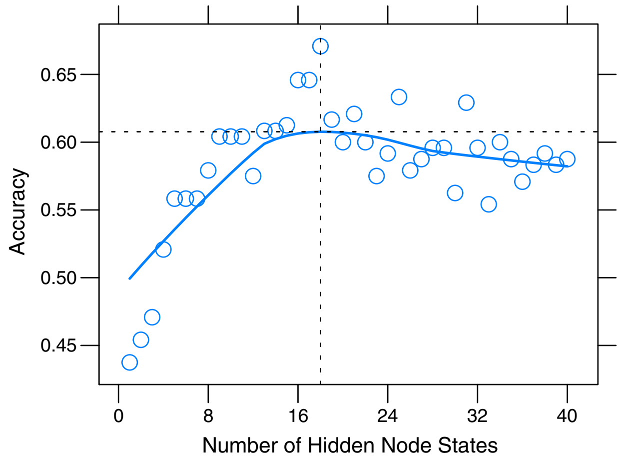

Fig. 3. The optimal number of hidden states was estimated by testing (cross validation within the training data) the accuracy of the HNB8 model across a range of hidden states. For 8 subjects, the HNB8 model performs best with 18 hidden states.

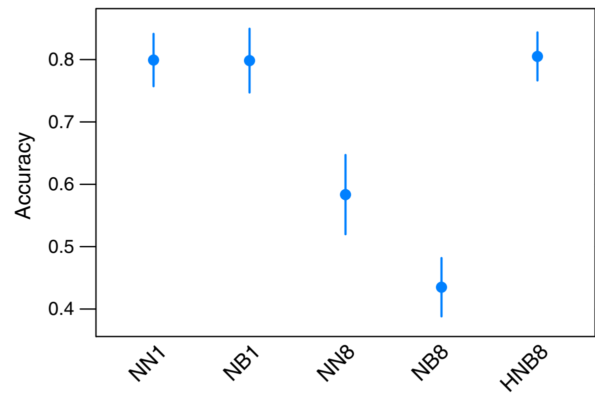

Fig. 4. Mean classification accuracies, error bars represent 95% confidence intervals.

8 subjects, the best performance was achieved using 18 hidden states. This value was used for the remainder of the analyses.

Classification accuracy

Classification accuracy means and standard deviations for each combination of model and workload level can be seen in Table 1. A one-way ANOVA was performed to compare the effect of model on classification accuracy. There was a significant effect of model on classification accuracy, $F(4,35) = 65.46$, $p < .001$, $\eta^2 = 0.88$.

Post hoc comparisons using the Tukey HSD test indicated that the NB8 model has a lower classification accuracy than every other model and the NN8 model had a lower classification accuracy than the NN1, NB1 and HNB8 models. There was no significant difference between then NN1, NB1 and HNB8 models. These comparisons can be seen in Fig. 4.

NN1 and NB1 perform reasonably well for subject-specific workload classification. NN8 and NB8, however, perform poorly for workload classification across subjects. In contrast, HNB8 performs well on all subjects, demonstrating its ability to effectively model the between subject variations. Tukey's honestly significant difference test was performed for all pairwise comparisons. NN1, NB1, and HNB8 had significantly higher mean classification accuracies than NN8 and NB8. The mean classification accuracy for the NN1, NB1 and HNB8 models, for each of the 3 workload levels is shown in Fig. 4. We can see that the performance of HNB8 is consistent across different workloads.

Discussion

The present experiment tested the performance of classifiers on three levels of workload. Except for the cross-subject naive Bayes model (NB8), the performance of the remaining classifiers were stable across all three workload conditions. This result is different from much of the past research on EEG-based workload classification which only had success classifying 2 levels of workload. There are two possible explanations for this difference. One possible explanation for this is that the task manipulations in previous experiments did not generate more than 2 actual levels of workload. The other explanation is related to choice of features used as input to the classifiers. Most, if not all of the previous research has used EEG power as an input to their workload classifiers. However, in our paper we chose to use magnitude. While the subjective perception of many psychological phenomenon (such as hearing) increase non-linearly, there has been no evidence to support the idea that the amplitude of brain waves increase non-linearly. For this reason we decided to use EEG magnitude as our classifier inputs. This choice may have led to a more stable classifier across workload conditions. On the other hand, the performance of our classifiers might be improved by using EEG power. A future comparison of magnitude versus power for classifier input could answer this question.

The classification accuracy of both the single subject naive Bayes (NB1) and cross-subject hierarchical naive Bayes classifier (HNB8) when aggregated across workload levels was 80% and equal to the performance of our benchmark single subject neural network classifier (NN1). These results are consistent with previous research involving classifiers of general cognitive workload. Additionally, this is an important result as this is the first report of a Bayes classifier being used for workload classification. The comparable performance of the NB1 and HNB8 models to the NN1 model shows that Bayes classifiers are a viable alternative to neural networks for EEG based workload classification.

When both naive Bayes and neural network classifiers were trained and tested on multiple subjects (NN8 and NB8), their performance was significantly worse than the subject-specific classifiers of the same type (i.e., NN8<NN1 and NB8<NB1). The performance of NB8 was actually no better than chance. This indicates that when not appropriately dealt with, the large amounts of variation that exists between different subjects can lead to significant decreases in performance. However, we were able to restore the accuracy of the cross-subject naive Bayes classifier by introducing a hidden node which through the use of hidden states and additional constraints, was able to account for the between subject variation. Compared to the NB1 and NN1 classifiers, which can only handle one subject at one time, the HNB8 classifier could handle data from multiple subjects without losing any performance.

In order to decide the optimal number of hidden states for the hierarchical model, its accuracy rates were tested with varying hidden states by cross-validation within the training data. The performance increased to a peak at 18 hidden states and gradually falls as more states are added. As expected, a large improvement in performance is seen when we increase the number of hidden states at first, which conforms to our assumption that more hidden states lead to strong

Table 1
Classification accuracy means and standard deviations for each combination of model and workload level.

Model	Workload		
	L	M	H
NN1	0.79 (0.04)	0.77 (0.06)	0.84 (0.06)
NB1	0.79 (0.07)	0.77 (0.09)	0.83 (0.06)
NN8	0.55 (0.09)	0.55 (0.11)	0.65 (0.15)
NB8	0.30 (0.16)	0.27 (0.12)	0.73 (0.16)
HNB8	0.83 (0.05)	0.79 (0.03)	0.79 (0.08)

Please cite this article as: Wang, Z., et al., Cross-subject workload classification with a hierarchical Bayes model, NeuroImage (2011), doi:10.1016/j.neuroimage.2011.07.094

power in dealing with variations. However, adding too many hidden states becomes detrimental to performance.

Conclusion

We have demonstrated that a cross-subject classifier can achieve performance comparable to subject-specific classifiers trained on individual subjects. This was accomplished by a hierarchical Bayes model that captures the between-subject variation with a latent variable with imposed constraints. Instead of using multiple subject-specific classifiers, we built one classifier that can handle multiple subjects. The novelty of this model is its ability to explicitly model the variations between subjects which has been a tremendous problem in EEG-based workload classification. These results take EEG-based classification one step closer to being able to discriminate workload for a novel subject which was not trained on.

Future directions

In the current article, we built an EEG-based workload classifier for a group of subjects which achieves performance comparable to subject-specific classifiers. This work was done with the goal of working towards a completely subject-independent workload classifier. Currently the data of all the subjects appear in both the training and testing data. We plan on improving our hierarchical Bayes model to handle novel subjects.

Since this work focuses on a study of cross-subject workload classification, the data used for both training and testing came from the same trial. In the future we plan to extend our method to cross-trial or even cross-day workload classification. This could be achieved by introducing new hidden nodes, additional hidden states to current hidden node, or a combination of both.

Our proposed model was compared with the dominant technique neural network in EEG-based workload classification in the current work. In the future, we plan on comparing it with other state of the art classifiers including SVM.

Different from other models such as NN and SVM, our proposed model applies the Bayesian approach to workload classification, which enables us to flexibly incorporate different forms of prior knowledge that is available but typically overlooked with the data to further improve the performance. For example, currently, all samples are treated as independent estimates of a workload level. However, state transitions between workload levels are slow; data points closer in time should be more similar than data points farther in time. We anticipate that adding temporal components into our models could further improve the classification accuracies. We also plan on investigating different circumstantial knowledge and applying them to improve workload classification using Bayes models.

We also plan on improving the overall classification accuracy of our hierarchical Bayes model by using better features and adding temporal information into the model. Currently, all samples are treated as independent estimates of a workload level. However, state transitions between workload levels are slow; data points close in time should be more similar than data points far away in time. We anticipate that adding temporal components into our models that classification accuracy will be further improved.

Acknowledgment

The work was supported, in part, by grant N000141010019 to Wayne Gray from the Office of Naval Research, Dr. Ray Perez, Project Officer.

References

Axelsson, S., 2004. Combining a Bayesian classifier with visualisation: understanding the ids. Proceedings of the 2004 ACM workshop on Visualization and data mining for computer security. VizSEC/DMSEC '04. ACM, New York, NY, USA, pp. 99–108.

Benferhat, S., Kenaza, T., Mokhtari, A., 2008. A naive Bayes approach for detecting coordinated attacks. Computer Software and Applications Conference, Annual International 0, pp. 704–709.

Byrne, E., Parasuraman, R., 1996. Psychophysiology and adaptive automation. Biol. Psychol. 42, 249–268.

Comstock, J. Raymond, Arnegard, R. J., 1992. The multi-attribute task battery for human operator workload and strategic behavior research. NASA Technical Memorandum 104174, Langley Research Center.

Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the em algorithm. J. R. Stat. Soc. 39, 1–38.

Domingos, P., Pazzani, M., 1997. On the optimality of the simple Bayesian classifier under zero-one loss.

Fairclough, S., Venables, L., Tattersall, A., 2005. The influence of task demand and learning on the psychophysiological response. Int. J. Psychophysiol. 56, 171–184.

Gaillard, A.W., 1993. Comparing the concepts of mental load and stress. Ergonomics 36, 991–1005.

Gelman, A., Hill, J., 2006. Data Analysis Using Regression and Multilevel/Hierarchical Models, 1st ed. Cambridge University Pression.

Gevins, A., Smith, M.E., 2003. Neurophysiological measures of cognitive workload during human–computer interaction. Theoretical Issues in Ergonomics Science, 4, pp. 113–131.

Gevins, A., Smith, M.E., Leong, H., McEvoy, L., Whitfield, S., Du, R., G.R., 1998. Monitoring working memory load during computer-based tasks with eeg pattern recognition methods. Hum. Factors 40, 79–91.

Gowadia, V., Farkas, C., Valtorta, M., 2005. Paid: a probabilistic agent-based intrusion detection system. computers and security 24(7):529-545. In Computers and Security, pp. 529–545.

Harris, W.C., Hancock, P., Arthur, E., Caird, J., 1995. Performance, workload, and fatigue changes associated with automation. Int. J. Aviat. Psychol. 5, 169–185.

Hart, S.G., Staveland, L.E., 1988. Development of nasa-tlx (task load index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (Eds.), Human mental workload. North-Holland, Amsterdam, pp. 139–183.

Jasper, H.H., 1958. Report of the committee on methods of clinical examination. Electroencephalogr. Clin. Neurophysiol. 10, 370–375.

Kamruzzaman, S.M., Rahman, C.M., 2010. Text categorization using association rule and naive Bayes classifier, 9.

Laufs, H., Kleinschmidt, A., Beyerle, A., Salek-Haddadi, A., Preibisch, C., Krakow, K., 2003. EEG-correlated fMRI of human alpha activity. NeuroImage 19 (4), 1463–1476.

Lysaght, R.J., Hill, S.G., Dick, A., Plamondon, B.D., Linton, P.M., Wierwille, W.W., Zaklad, A.L., Alvah, C., Bittner, J., Wherry, R.J., 1989. Operator workload: comprehensive review and evaluation of operator workload methodologies. AFI Technical Report, 851. U.S. Army Research Institute.

Moller, M.F., 1993. A scaled conjugate gradient algorithm for fast supervised learning. Neural Netw. 6 (4), 525–533.

Norman, D.A., Bobrow, D.G., 1976a. On data-limited and resource-limited processes. Cogn. Psychol. 7.

Norman, D.A., Bobrow, D.G., 1976b. On the analysis of performance operating characteristics. Psychol. Rev. 83.

Noyes, J.M., Bruneau, D.P.J., 2007. A self-analysis of the nasa-tlx workload measure. Ergonomics 50, 514–519.

P.M., Patra, M.R., 2007. Network intrusion detection using naïve Bayes.

Puttini, R.S., Marrakchi, Z., Me, L., 2002. Bayesian classification model for real-time intrusion detection. In 22th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering.

Qu, L., Vetter, V., Bird, G., Qiu, H., White, P., 2010. A naive Bayes classifier for differential diagnosis of long qt syndrome in children. Bioinformatics and Biomedicine (BIBM), 2010 IEEE International Conference on, pp. 433–437.

Rish, I., 2001. An empirical study of the naive Bayes classifier. International Joint Conference on Artificial Intelligence.

Rubio, S., Diaz, E., Martin, J., Puente, J.M., 2004. Evaluation of subjective mental workload: a comparison of swat, nasa-tlx, and workload profile methods. Appl. Psychol. Int. Rev. 53, 61–86.

Wilson, G.F., Russel, C.A., 2003a. Operator functional state classification using psycho-physiological features in an air traffic control task. Hum. Factors 45, 381–389.

Wilson, G.F., Russel, C.A., 2003b. Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. Hum. Factors 45, 635–643.

Wilson, G.F., Russell, C.A., 2007. Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. Hum. Factors 49, 1005–1018.

Wilson, G.F., Estepp, J., Davis, I., 2009. A comparison of performance and psychophys-iological classification of complex task performance. Human Factors and Ergo-nomics Society Annual Meeting Proceedings 53, 141–145.

Wilson, G.F., Estepp, J., Christensen, J.C., 2010. How does day-to-day variability in psychophysiological data affect classifier accuracy. Human Factors and Ergonomics Society Annual Meeting Proceedings, 54, pp. 264–268.

Xu, L., Jordan, M., 1996. On convergence properties of the em algorithm for gaussian mixtures. Neural Comput. 8, 129–151.

Zhang, H., 2004. The optimality of naive Bayes. The International FLAIRS Conference, Vol. 17.