33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

An EEG Workload Classifier for Multiple Subjects

Ziheng Wang, Ryan M. Hope*, Zuoguan Wang, Qiang Ji and Wayne D. Gray

***Abstract*— EEG data has been used to discriminate levels of mental workload when classifiers are created for each subject, but the reliability of classifiers trained on multiple subjects has yet to be investigated. Artificial neural network and naive Bayesian classifiers were trained with data from single and multiple subjects and their ability to discriminate among three difficulty conditions was tested. When trained on data from multiple subjects, both types of classifiers poorly discriminated between the three levels. However, a novel model, the naive Bayesian classifier with a hidden node, performed nearly as well as the models trained and tested on individuals. In addition, a hierarchical Bayes model with a higher level constraint on the hidden node can further improve its performance.**

I. INTRODUCTION

Adaptive automation technologies promise to meliorate the demands made on mental capabilities by modern automation and computerized systems [7], [11]. A critical aspect of these adaptive technologies is accurate and reliable assessment of operator workload. Traditionally, workload is assessed by questionnaires which are quantified through statistical techniques such as factor loading, discriminant analysis, and correlation/covariance analysis [6]. Although progress has been made, there are no globally accepted methods for measuring and predicting workload [10], [13], [12]. 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 moved towards using electrophysiological measures to predict workload [5], [4]. In particular, electroencephalography (EEG) has been used extensively to examine the changes in the brain's electrical activity in response to cognitive activity. The main assumption is that if brain-state classifiers can be found, then they can be used by a brain-computer interface (BCI) in real-time to detect operator mental workload [17].

While a number of different classifiers have been used with EEG data, such as linear discriminant analysis, support vector machines and artificial neural networks, it is not clear which method is superior (see [1] and [9] for extensive reviews). Artificial neural networks (NN) are very popular classifier and have shown success discriminating at least two levels of cognitive workload [14], [15], [16], [19], [18]. In this paper we present the first application of Bayesian networks to the detection of cognitive workload and compare these to NNs.

In principle, EEG provides an objective and relatively unobtrusive means for measuring workload. In practice, much work needs to be done in the development of quantitative methods for analyzing and interpreting EEG data. Training classifiers is time consuming and requires a lot of data, especially in situations that involve multiple subjects. Currently, the standard practice is to train a new classifier for each subject. Recent research suggests it might even be necessary to train new classifiers each day [18]. One way to potentially reduce overall training time is to train one model across subjects. The large variability between subjects poses a significant challenge to building a common classifier, which has not previously been investigated. With traditional techniques such as NN, it seems likely that the classifier would not separate signal from noise across multiple subjects. However if measures can be incorporated to account for between subject variation, such a classifier might produce more robust and stable classifications.

The present paper has two goals. The primary goal is to investigate the effect of between subject variability on workload classifier accuracy. The secondary goal is to introduce a cross-subject Bayesian network and compare its performance with NN. In order to accomplish these goals, we first compare the accuracy of a NN, a standard naive Bayes classifier, a novel naive Bayes classifier with a hidden node and a hierarchical Bayes classifier when each are trained on EEG data from multiple subjects and tested on individual subjects. We then compare these classifiers to the performance of NNs and standard naive Bayes classifiers which were trained and tested on single subjects. The current paper is organized as follows: section II presents the graphical models we used to handle between-subject variations. Following the models we will introduce the experiments in section III and show the results in section IV. Finally we will discuss the results and the future directions.

II. MODELS

A. Naive Bayes Classifier

Naive Bayes Classifier (NB) is a very 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. NB assumes all the feature nodes are independent of each other given the class, and typically, the feature variables are assumed to have Gaussian distribution if they are continuous. Despite its naive design and apparently over-simplified assumptions, NB has worked quite well in many

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

Ziheng Wang, Zuoguan Wang and Q. Ji are with the Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute

Ryan M. Hope and Wayne D. Gray are with the Department of Cognitive Science, Rensselaer Polytechnic Institute.

*Corresponding author. Email address: hoper2@rpi.edu

complex real-world situations. Compared to other complex graphical models, it requires smaller amount of training data to accurately estimate the parameters necessary for the classification [20]. For the view of simplicity, the feature nodes are always grouped into one node X and the structure of NB is plotted as Fig. 1b.

The classification results are determined by the posterior probability $P(C|X_1, X_2, ..., X_n)$, which can be transformed using the chain rule and Bayes' theorem into (1), where α is a normalization constant.

$$\begin{aligned} &P(C|X_1, X_2, \cdots, X_n) \\ =&\alpha P(X_1, X_2, \cdots, X_n|C)P(C) \\ =&\alpha \prod_{i=1}^{n} P(X_i|C)P(C) \end{aligned} \tag{1}$$

Given a test sample $(X_1, X_2, \cdots, X_n)$, the class is determined by (2):

$$C^* = \underset{C}{argmax} \prod_{i=1}^{n} P(X_i|C)P(C) \tag{2}$$

In our case, the class node represents three workload conditions and the feature nodes $(X_1, X_2, \cdots, X_n)$ represent the EEG frequency features.

B. Modeling Between-subject Variations with a Hidden Node

The between subject variations pose a big challenge when training a common classifier for use on multiple subjects. To deal with the variations we propose a novel naive Bayes Classifier with a hidden node (NB-HN).

Despite a large amount of variation between subjects, it is reasonable to assume that there exists some commonalities in their brain signals in response to the task demands. By introducing a hidden node to the standard NB model we can account for both the common aspects of each subject's as well as the individual differences. A graphical model of NB-HN classifier is shown in Fig. 1c, where an additional node H is connected to each feature node and the class node. Node H is used to model the inter-subject differences. For easier understanding, H can represent different subjects in the simplest case. However in our model it is not restricted to any specific meaning and may stand for any factor that can cause variations. The expectation maximization (EM) algorithm is used to uncover the hidden node H [3]. No a priori information is needed during the training stage to compute the hidden node. Additionally, the value of the hidden node is not needed at the testing stage. The likelihood can be computed by marginalizing over the hidden node H with (3).

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

The hidden node may be a discrete node or continuous node. Intuitively, the larger its size is, the more information it contains. For this experiment, we used a discrete hidden node of size 18.

C. Constraining Variation with a Hierarchical Model

Despite the advantage of the hidden node to capture the large unknown variations, it also brings the risk of overfitting. That is, variations as well as noises are both captured by the hidden node. We propose a hierarchical model to alleviate this risk by imposing some constraints on the hidden states. The assumption is that variations should not depart significantly from the common part. Fig. 1d shows this hierarchical model (HNB-HN). X represents the feature vector. H is still the hidden node. μ and σ are the parameters of $P(X|C, H)$. The hyper parameters μ_0 and σ_0 are used to represent the commonality among different subjects. By introducing this top node μ_0, σ_0 which is connected to the parameter μ, the variations are restricted to follow a Gaussian distribution $N(\mu_0, \sigma_0)$ and thus are constrained to a certain area close to the shared commonality.

The training procedure of this hierarchical model is divided into 2 steps. Firstly, The commonality (parameters μ_0 and σ_0) among different subjects is learned using the training data as well as the prior knowledge from the expert domain. Secondly the hidden node H are uncovered with EM under these constraints.

Mathematically, the model can be described by (4) - (5):

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

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

III. METHOD

The EEG data used in the present article comes from a previously published study by [18] in which eight participants (3 males; mean age 21.1 years) performed the Multi-Attribute Task Battery (MATB) [2] 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 MATB includes monitoring, communication, and resource allocation tasks which are performed concurrently in a continually changing task environment. 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 five minutes 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 reason, day 3 was excluded resulting in 12 complete trials for each participant.

A. Feature

Nineteen EEG channels were recorded using the International 10-20 montage [8]. Mastoids were used as reference and ground with electrode impedances 5K ohms or less. The EEG data was down sampled to 128 samples per second 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 seconds of overlap. No taper

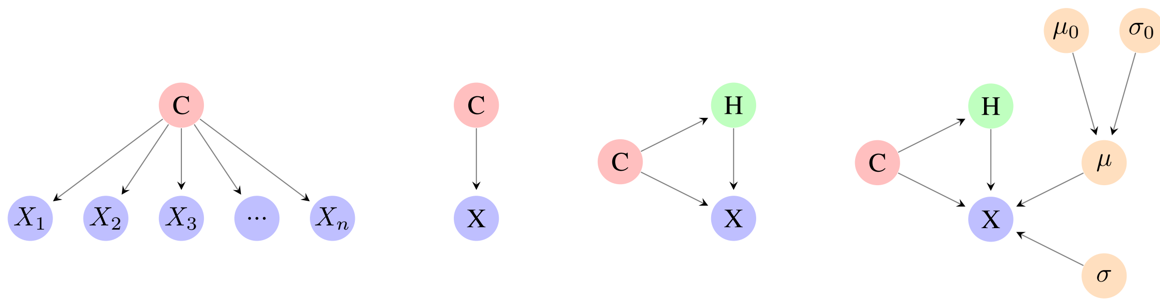

Fig. 1. The Evolution from A Naive Bayes Classifier to the Hierarchical Bayes Model: (a) A Naive Bayes Classifier, where C represents the class node and X represents the feature node; (b) A simplified Graphical Model of Naive Bayes Classifier; (c) A Naive Bayes Classifier plus a Hidden Node; (d) A Hierarchical Bayesian model where μ_0 and σ_0 describes the prior distribution of μ.

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.

B. Classifiers in the Present Paper

Six different classifiers are compared in our experiment, namely, the neural net and naive Bayes classifiers trained and tested on individual subjects (NN-1 and NB-1), the neural net and naive Bayes classifiers trained on multiple subjects and tested on individual subjects (NN-8 and NB-8), naive Bayes classifier with a hidden node (NB-HN-8) and the hierarchical hidden node naive Bayes classifier (HNB-HN-8) both of which are trained on multiple subjects and tested on individual subjects.

For the neural network classifier used for comparison in our experiment, we followed the same setup in [19]. The hidden layer has the same number of nodes as the input layer and the output nodes are the three conditions (Low, Medium and High). One fifth of the training data was randomly selected as the validation set and Scaled Conjugate Gradient algorithm was used to train the net. The parameters of neural network were also tuned so that the results reflected the best performance we can achieve.

The size of the hidden node of NB-HN-8 and HNB-HN-8 are 12 and 18 respectively. Besides, all the models are multi-class classifiers trained on all the three workloads instead of binary ones.

C. Experimental Setup

Models were trained and tested using a fivefold cross-validation setup. For the purpose of testing the algorithm with little amount of training data, only one fifth of the EEG data from each trial was randomly sampled for the purpose of training the models. The data not selected for training was used for testing. Data was sampled evenly across workload blocks, and for the models including multiple subjects' data, evenly across subjects. This procedure was repeated for each trial.

IV. Results

The mean classification accuracy for each $model \times training \times workload$ combination is shown in Fig. 2. See Table I for results of Tukey's honestly significant difference test. NN-1, NB-1, NB-HN-8 and HNB-HN-8 had significantly higher mean classification accuracies than NN-8 and NB-8. The classification accuracy of NN-8 was significantly higher than NB-8. There were no significant differences between NN-1, NB-1, NB-HN-8 and HNB-HN-8. Our proposed common classifier HNB-HN-8 for multiple subjects achieves performance comparable to the individually trained classifiers NN-1 and NB-1.

V. Discussion

We had two goals in the present paper, explore the affects of between-subject variability on classifier accuracy, and compare the performance of artificial neural networks and Bayesian networks. In order to ensure fair comparisons, identical features were used for all models. We had no interest in feature selection or optimization. Therefore, the accuracy levels achieved do not necessarily represent the best possible performance of any of the models. However the advantage of such our proposed model HNB-HN for handling multiple subjects remains.

Previous research has demonstrated that artificial neural networks can achieve high classification accuracy rates. It is worth noting that in our experiments, the individually trained naive Bayesian classifiers (NB-1) and the naive Bayesian classifiers trained on multiple subjects (NB-HN-8 and HNB-HN-8) had mean classification accuracies that are comperable to the individually trained neural network classifiers (NN-1).

As expected, both the neural net and standard naive Bayes classifiers trained on multiple subjects (NN-8 and NB-8 respectively) performed worse than the individually trained classifiers of the same type (i.e., NN-8 $<$ NN-1 and NB-8 $<$ NB-1). The performance of NN-8 and NB-8 were both poor, but NB-8 was essentially no better than chance. These classifiers were not able to pick out the signal from the noise when presented with data from multiple subjects. However,

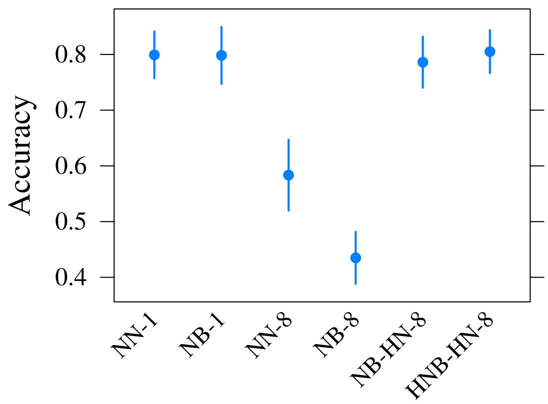

Fig. 2. Mean classification accuracy from testing on each of the 8 subjects for each of the five classifiers. The error bars represent 95% confidence intervals.

TABLE I
TUKEY MULTIPLE COMPARISONS OF MEANS 95% FAMILY-WISE CONFIDENCE LEVEL

		diff	*sig*
NB-HN-8	HNB8	-0.0189	
NB1	HNB8	-0.0067	
NB8	HNB8	-0.3701	*
NN1	HNB8	-0.0058	
NN8	HNB8	-0.2216	*
NB1	NB-HN-8	0.0122	
NB8	NB-HN-8	-0.3512	*
NN1	NB-HN-8	0.0131	
NN8	NB-HN-8	-0.2027	*
NB8	NB1	-0.3634	*
NN1	NB1	0.0009	
NN8	NB1	-0.2149	*
NN1	NB8	0.3643	*
NN8	NB8	0.1485	*
NN8	NN1	-0.2158	*

when a constrained hidden node that performed expectation-maximization was introduced to the standard naive Bayes classifier (HNB-HN-8), its performance increased to that of a individually trained naive Bayes classifier.

VI. CONCLUSION

In this paper, we demonstrated that a classifier trained on multiple subjects can achieve performance comparable to classifiers trained on individual subjects. This was accomplished by adding a hidden node to a naive Bayes classifier. The hidden node in this case used the expectation-maximization algorithm to account for between subject variations. The performance is further improved by imposing constraints on the hidden node. These results take EEG classification one step closer to being able to discriminate workload levels on subjects that the classifier was not trained on. Since this work focuses on a study of cross-subject workload classification, the data used for both training and testing is limited to the same trial. One possible future work is to extend our method to cross-trial or even cross-day workload classification. This can be achieved by either introducing additional hidden nodes or additional hidden states to current hidden node or a combination of two.

REFERENCES

[1] Ali Bashashati, Mehrdad Fatourechi, Rabab K Ward, and Gary E Birch. A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. *Journal of Neural Engineering*, 2007.
[2] Jr. Comstock, J. Raymond and Ruth J. Arnegard. The multi-attribute task battery for human operator workload and strategic behavior research. NASA Technical Memorandum 104174, Langley Research Center, 1992.
[3] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39:1–38, 1977.
[4] A. Gevins and M. E Smith. Neurophysiological measures of cognitive workload during human–computer interaction. *Theoretical Issues in Ergonomics Science*, 4:113–131, 2003.
[5] A. Gevins, M. E. Smith, H. Leong, L. McEvoy, S. Whitfield, R. Du, and Rush G. Monitoring working memory load during computer-based tasks with eeg pattern recognition methods. *Human Factors*, 40:79–91, 1998.
[6] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, editors, *Human mental workload*, page 139183. Amsterdam: North-Holland, 1988.
[7] T Inagaki. Adaptive automation: Sharing and trading control. In Erik Hollnagel, editor, *Handbook of Cognitive Task Design*, pages 147–169. Lawarence Erlbaum Associates, Inc., 2003.
[8] H. H. Jasper. Report of the committee on methods of clinical examination. *Electroencephalography and clinical Neurophysiology*, 10:370–375, 1958.
[9] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for eeg-based brain-computer interfaces. *Journal of Neural Engineering*, 4(2), June 2007.
[10] Robert J. Lysaght, Susan G. Hill, A.O. Dick, Brian D. Plamondon, Paul M. Linton, Walter W. Wierwille, Allen L. Zaklad, Jr. Alvah C. Bittner, and Robert J. Wherry. Operator workload: Comprehensive review and evaluation of operator workload methodologies. AFI Technical Report 851, U.S. Army Research Institute, 1989.
[11] N Moray, T Inagaki, and M Itoh. Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. *Journal of Experimental Psychology: Applied*, 6:44–58, 2000.
[12] J. M. Noyes and D. P. J. Bruneau. A self-analysis of the nasa-tlx workload measure. *Ergonomics*, 50:514–519, 2007.
[13] S. Rubio, E. Diaz, J. Martin, and J. M. Puente. Evaluation of subjective mental workload: A comparison of swat, nasa-tlx, and workload profile methods. *Applied Psychology An International Review*, 53:61–86, 2004.
[14] G. F. Wilson and F Fisher. Cognitive task classification based on topographic eeg data. *Biological Psychology*, 40, 1995.
[15] G. F. Wilson and C. A. Russel. Operator functional state classification using psychophysiological features in an air traffic control task. *Human Factors*, 45:381–389, 2003.
[16] G. F. Wilson and C. A. Russel. Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. *Human Factors*, 45:635–643, 2003.
[17] G. F. Wilson and C. A Russell. Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. *Human Factors*, 49:1005–1018, 2007.
[18] Glenn F. Wilson, Justin Estepp, and J. C. Christensen. How does day-to-day variability in psychophysiological data affect classifier accuracy. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 54:264–268, 2010.
[19] Glenn F. Wilson, Justin Estepp, and Iris Davis. A comparison of performance and psychophysiological classification of complex task performance. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 53:141–145, 2009.
[20] Harry Zhang. The optimality of naive bayes. In *The International FLAIRS Conference*, volume 17, 2004.