Neurocomputing 129 (2014) 107–113

Contents lists available at ScienceDirect

Neurocomputing

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

An integrated Gaussian mixture model to estimate vigilance level based on EEG recordings ☆

Jing-Nan Gu *, Hong-Tao Lu, Bao-Liang Lu

MOE-Microsoft Laboratory for Intelligent Computing and Intelligent Systems, Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

ARTICLE INFO

Article history:
Received 31 January 2012
Received in revised form
5 July 2012
Accepted 19 October 2012
Available online 6 November 2013

Keywords:
EEG
Vigilance estimation
Labeling
Gaussian mixture model

ABSTRACT

Vigilance level estimation can be used to prevent disastrous accident occurring frequently in high-risk tasks. Electroencephalograph (EEG) based Brain Computer Interface (BCI) is one of the most important tools for detecting one's brain electrical activities. Unfortunately, several problems including its sensitivity to artifacts, inaccurate labels and the great diversity of patterns within EEG signals present great challenges to predict vigilance level reliably. In this paper we propose an integrated approach to estimate vigilance level, which incorporates an automatically artifact removing preprocess, a novel vigilance labeling method and finally a Gaussian Mixed Model (GMM) to discover the underlying pattern of EEG signals. Extensive off-line experiments are conducted on 12 groups of data sets to show the effectiveness of our integrated approach in the real-time application. A reasonably high classification performance (88.46% over 12 data sets) is obtained with low delay by employing only one channel in the frontal lobe, which is in accordance with the conclusions of brain science and is of significance in practice.

1. Introduction

Vigilance refers to the ability of human being to sustain attention over a period of time, which is a central feature of human cognition [2,3]. Previous studies [3,4] have shown that vigilance will decline dramatically when people are required to perform a tedious task, meaning that maintaining high vigilance level throughout a boring job is usually very difficult, if not impossible. Consequently, identification of vigilance level of the operator is essential for guaranteeing security and reliability in some high-risk works such as driving a car or operating a train.

Research in BCI has received much interest in the past few decades [5–7] and has led to many useful applications, such as spelling words or controlling a cursor [8–10]. A BCI provides users with a new communication channel between the human brain and the computer. The central element in each BCI is a translation algorithm that converts electrophysiological input from the user into output that controls external devices. Specifically, the EEG-based BCIs that provide an effective way to predict the vigilance level will be of our particular interest [7]. And these systems have been proven to be powerful in avoiding serious accidents incurred by loss of concentration. EEG reflects the electrical activity of the brain along the scalp [11] and can be well interpreted for both clinical applications and psychological studies [9,10]. Though some techniques with high spatial resolution (<1 mm) such as Magnetic Resonance Imaging (MRI) and X-ray Computed Tomography (CT) are optional alternatives for these tasks [12], EEG-based techniques can provide real-time performance with millisecond-range temporal resolution that is not possible with MRI or CT. Besides, the data acquisition process is relatively more convenient in our daily life for the vigilance estimation uses. Also, because EEG signals measure one's intrinsic physiological status before one's extrinsic behavior, monitoring systems based on EEG are more preferable than the video-based detecting techniques [13].

EEG-based vigilance estimation task is a typical supervised learning problem that every instance of estimation is on the basis of a model learned from a well-labeled training set. This classical pattern recognition task usually involves 3 successive steps, including data preprocessing, feature extraction and finally determining a discrete or continuous vigilance value to the corresponding data [14]. According to whether the estimation value is discrete or not, the problem can be further subdivided into a classification task or a regression task. A regression result always contains more plentiful and comprehensive information and can be easily converted to discrete categories in our vigilance estimation scenario. Both the classification and regression task will be discussed under a generative model [15] in later sections. Unfortunately, several undesired properties related to EEG signals present great challenges to predict vigilance level. The difficulties

☆This paper is an extension of a ICONIP2011 paper [1]. We add some supporting graphs and tables and also give a more clear explanation about the experiment result.

* Corresponding author.
E-mail address: albertguwn@gmail.com (J.-N. Gu).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.10.042

involved in processing real world EEG signals consist of its sensitivity to artifacts, ineffective learning due to inaccurate labels and the great diversity of patterns between different people or even between different time with a same person [16].

Studies in [17] have demonstrated that disposal of artifacts plays an important role to guarantee the robustness of vigilance prediction. Artifact avoidance, rejection and removal are the 3 mostly used strategies to handle artifacts [18]. An efficient algorithm aimed at detecting and removing artifacts automatically that have been developed in [19]. The work will lay a foundation to improve and ensure our classification performance. With a relatively clean EEG recording, many researchers focused on feature extraction process intending to reveal neurophysiological phenomena. Time series information and power spectral density are generally accepted features involved in EEG signals [20]. Both of them will be explored in our study to provide adequate information for subsequent analysis. Also many researchers engaged in specific learning algorithms. Linear discriminant analysis (LDA), support vector machine (SVM) and neural networks are the most popular methods to classify the EEG signals [21]. Specifically, an infinite Gaussian mixture model based on Bayes inference is proposed in [22] to avoid overfitting in the training process when dealing with high dimensional data. However, it turns out impractical for its high computational complexity in the process of model inference.

In this study, we propose an integrated approach to analyze EEG signals for the purpose of predicting vigilance levels on-line. To handle the undesired artifacts that made significantly impact on EEG signals, a Blind Source Separation (BSS)-based artifact removal approach is used for preprocessing [19]. Since there are few acceptable rules for labeling vigilance scales, we develop a novel and reasonable vigilance labeling method. Both the spatial and spectral information are implemented for our feature extractor. Finally the GMM is explored to approximate the class conditional probability to accomplish both classification and regression tasks [23]. The posterior probability calculated by the generative model presents the reliability of each classification rather than only a result of category information. To verify the expressive ability provided by GMM, we compared classification performance with 2 other popular methods SVM and LDA over 12 data sets. Regression performance based on posterior probability calculated by GMM is also investigated to prove the ability of density estimation of Gaussian mixture model.

The rest of this paper is organized as follows. In Section 2 we explain the data acquisition process along with the labeling mechanism in detail. The main framework of our integrated approach and a deep discussion about the GMM is presented in Section 3. Experimental results with comparison to SVM and LDA are illustrated for both classification and regression tasks in Section 4. Followed with conclusions and directions for future work in Section 5.

2. Data acquisition and labeling

To verify the performance and reliability of the integrated approach we proposed, we conducted extensive pre-experiments to collect adequate data based on our simulated driving system. Each experiment trial collects one data set once a day at noon. Every subject is asked to perform a monotonous task sustained about 1 h after lunch with inadequate sleep in the previous night [24]. We develop a new and reasonable approach to label the vigilance scales associated to EEG signals. Having these EEG signals and their corresponding labels, we could go deep into the core part of our study, which is discussed in Section 4.

2.1. Data acquisition

Data acquisition will account for surprisingly large part of the cost of our study. To collect a sufficiently large and representative set of samples for the purpose of subsequent training and testing phase, we invited 10 healthy volunteers to participate in our experiment. The age of the volunteers range from 18 to 35, of which there are 6 males and 4 females. Each of them accomplished 2 experimental trials with interval more than 7 days. It took us nearly a month to complete this group of experiments.

2.1.1. Simulated driving system

Our simulated driving system mainly consists of a software-based simulator, a EEG collection system and other necessary facilities (e.g. a 19 in. LCD, a comfortable chair, etc.) [24]. The whole system is located in an isolated noise-free room which has normal brightness and constant temperature between 24 °C and 26 °C.

During the experiment, the simulator software would emit a series of traffic signs randomly on the computer screen. The sign is rectangle or triangle-shaped and the main component will be among one of the 4 colors (red, blue, yellow or green). The sign would be shown every 7 ± 0.5 s for duration of 1 s successively, with the interval the screen appearing as pure black. There is also a rectangle panel with 4 colored buttons that the subject should hold during the experiment. The subject is supposed to push the corresponding button in their panel accurately and promptly once the sign appeared on the screen.

The 62-channel NeuroScan system running at its full capacity sampling at 500 Hz is employed for collecting EEG signals, though the frequency band between 0 and 50 Hz is considered useful in EEG signals. This high-frequency signal could guarantee the integrity of neurophysiological information and thus being available in the future research. Fig. 1 shows the distribution of electrodes on the scalp based on the International 10–20 system [11].

2.2. Labeling strategy

There is no gold standard for scoring vigilance scales [25]. Traditionally labeling process is mainly supervised by the visual inspection and some heuristic rules. Such a subjective criterion lacks a consistent standard and usually leads to an unreliable learning. To make our result more convincing, we must create an

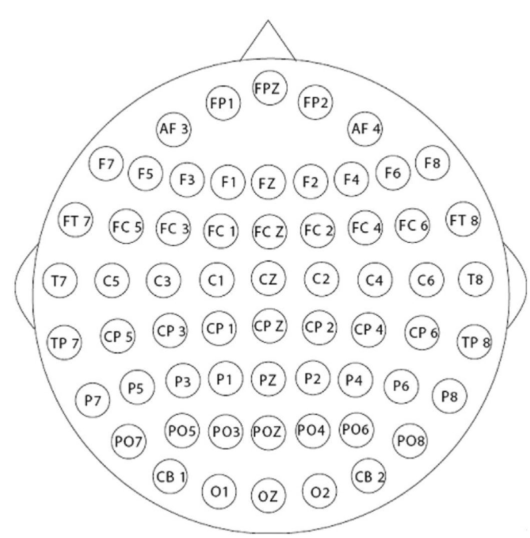


Fig. 1. Locations of electrodes on the scalp based on the International 10–20 system [11].

objective and appropriate strategy to quantize vigilance levels in the light of the performance the subject completed the given task. We propose a reasonable labeling approach considering both the respond time and the error rate the subject pushes the button within a window along 30 s. Our labeling method presents good agreement with the subject's real vigilance state which is monitored simultaneously by a camera.

The mistake credit $mistake_i$ is assigned according to the following rules:

- Case $\mathcal{A}$: normal responding sequence

  sti− > acc− > sti− > acc− > …

  if (diff < 0.2 s) $mistake_i = 1$ //ineffective input.(too fast)
  else if (diff < 2 s) //effective input.
  if (sti! = acc) $mistake_i = 1$ //wrong input.
  else $mistake_i = 0$ //right and prompt input.
  else mistake = 2 //ineffective input.(too slow)

- Case $\mathcal{B}$: abnormal responding sequence (too nervous)

  sti− > acc− > acc− > …− > acc− > sti

  $mistake_i = 0$ //regarded as maintaining high vigilance level.

- Case $\mathcal{C}$: abnormal responding sequence (asleep or distracted)

  acc− > sti− > sti− > …− > sti− > acc

  $mistake_i = 3$ //asleep or distracted

Here *sti* stands for a sign emitting on the screen and the *acc* means the subject's response. *diff* stands for the time between a *sti* and a subsequent *acc*. $mistake_i$ will be assigned to a discrete penalty (0,1,2,3) whenever it encountered with a *sti*. Fig. 2 shows the situation when case $\mathcal{A}$ happened. Note that case $\mathcal{B}$ is rarely happened in real experiment and is reserved only for integrity of our labeling approach.

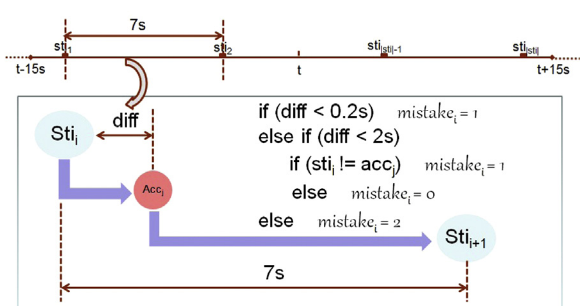


Fig. 2. Case $\mathcal{A}$ happened in our labeling strategy.

The error rate of a particular moment t is defined as follows:

$$Err_t = \frac{\sum_{i=1}^{|sti|}(mistake_i)}{3*|sti|}, \tag{1}$$

where $|sti|$ represents the number of *sti* within a 30 s long window and *Err* always lies in 0–1. Note that the vigilance level has negative correlation with *Err*, with high *Err* indicating low vigilance level which is in danger. The corresponding EEG recordings will be labeled for every 5 s. Low vigilance level would be labeled as 2 if $Err > 0.4$ at the moment, otherwise with high vigilance level labeled as 1.

3. Methods

Given the raw EEG recordings and corresponding labels, our vigilance estimation system mainly entails the following 4 phases as illustrated in Fig. 3. Firstly, the raw signal is filtered to specific frequency bands of our interest and then processed to remove artifacts caused mainly by electrooculography (EOG) and electromyography (EMG). Care must be taken during this preprocessing because EEG signals are relatively weaker than EOG or EMG and useful information contained in EEG is often distorted. Secondly, feature extraction is done to find relevant and effective features for subsequent classification and regression tasks. Both the spatial and spectral information are considered to be promising predominant features and are implemented in our study. Thirdly, we employed the Gaussian Mixture Model to discover the underlying probabilistic representation of the training set. Consequently the approximate model can be further utilized to determine the decision boundaries of different categories and to calculate the posterior probability, which could illustrate the reliability of each classification result and can be interpreted as continuous regression result naturally.

3.1. Filtering and artifact removal

Table 1 listed our prior knowledge about the useful frequency bands of EEG signal. Having known that meaningful information involved in EEG signals are mainly lying in frequency band 1–40 Hz [26], we therefore employ a band-pass filter that passes frequencies within the range 0.1–45 Hz and rejects frequencies outside that range to eliminate unnecessary information within

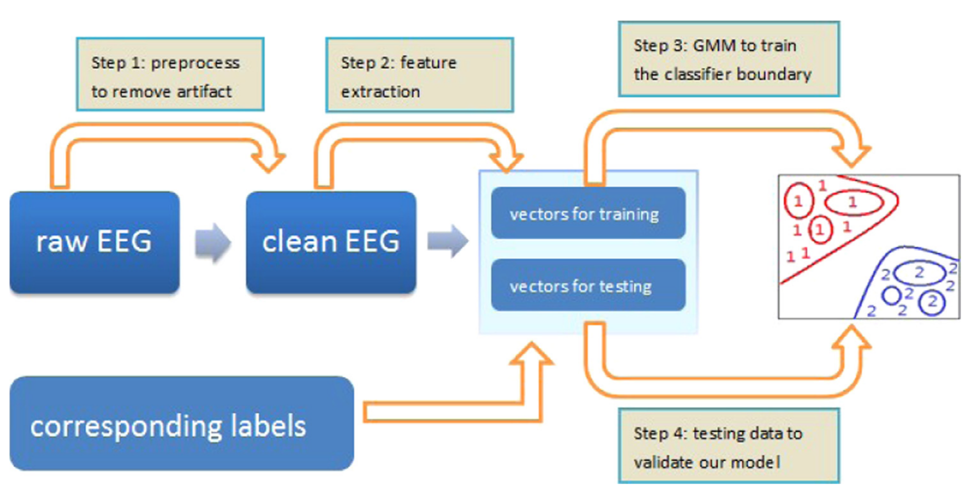


Fig. 3. Flowchart of EEG signal processing. Detailed experimental settings of steps 3 and 4 are discussed in Section 4.

Table 1
Frequency band of EEG signals.

δ	θ	α	β	γ
0.5–3.5 Hz	4–7 Hz	8–13 Hz	14–25 Hz	26–35 Hz

the EEG signals. After this filtering process, we could reduce the sample rate from 500 Hz to 100 Hz to accelerate the speed of computation in the following steps. This downsampling process is out of question as long as the Shannon–Nyquist sampling theorem is maintained. Here we have that the downsampled rate (100 Hz) is greater than double of the highest frequency (45 Hz) of the filtered signal.

Also we know that artifacts, especially the EMG and EOG, would contaminate the original weak EEG signals seriously. It must be detected reliably and the interference is supposed to be eliminated before subsequent data analysis. There are 3 mostly used methods to address artifacts [18]. Artifact avoidance is simply to avoid appearance of artifact by issuing proper instructions to users. It is assumed that no artifact is present in the collected signal and thus needs no extra handling of the signal. However, it is nearly impractical during an on-line use of any BCI system in that blinking eyes or shaking heads is unavoidable. We can also discard the epochs contaminated with artifacts by inspecting manually or automatic approaches. This method is called artifact rejection. But it has a fatal drawback that once the distorted signal is discarded, the system will fail to convert the brain activity in that epoch to corresponding controlling commands. Artifact removal is third approach and being recognized as the most promising strategy in dealing with artifact nowadays. It could remove the artifacts automatically and effectively as well as keeping the EEG-related signals intact.

In our study, we adopt all these 3 underlying ideas to minimize the interference caused by artifact. Firstly the subject usually coordinates with us to configure the experiment setup in the starting few minutes and to conserve the data in the last few minutes. Thus the beginning and last 5 min recordings are removed directly and this operation is recognized as artifact rejection. Then we attempt to avoid the appearance of artifact in the remaining time of experiment in that the subject sits alone in a nearly noise-free room. Finally a BSS-based technique is used to distinguish the EEG-related signals and EEG-unrelated artifacts and then we use SVM to identify and remove the artifacts automatically. It has been verified that this BSS-based artifact removal approach could improve the performance related to a motor imagery task [19].

3.2. Feature extraction

Having obtained the artifact-free EEG signals, we could now devote ourselves to selecting and designing features. Note the fact that changes of one's physiology status would produce corresponding changes of power density of specific spectral band [27]. Thus we use Short Time Fourier Transformation (STFT) to calculate the power spectral density and recognize the density of specific frequency bands as features [28].

Also auto-regression (AR) model could establish the relationship between brain status and time quite effectively. And this spatial information can be fully depicted by the coefficients of AR model [12,29]. To incorporate adequate information involved in EEG signals, we combine both temporal and spatial information as features [20]. Note that the order of AR is a smoothing parameter controlling the dimension of the feature. There is a tradeoff between adequate information and the curse of dimensionality [30]. The order of AR model is finally determined by cross-validation (CV). The detail of this parameter selection process will be explained in Section 4.

3.3. Gaussian mixture model

We employ the generative model to calculate the posterior probability of the kth category [15] by simply applying the Bayes' formula as follows:

$$p(\omega_k|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_k)p(\omega_k)}{\sum_{k=1}^{K} p(\mathbf{x}|\omega_k)p(\omega_k)}. \tag{2}$$

Although the underlying class densities $p(\mathbf{x}|\omega_k)$ of real world EEG data are often difficult to approximate, it would be powerful to both classification and regression problems if we could obtain it. For this reason latent variables are introduced to form a Gaussian mixture model to hope that it could express complicated distribution [23] as well as controlling the complexity of the model:

$$p(\mathbf{x}|\omega_k) = \sum_{i=1}^{|\omega_k|} \pi_i \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \tag{3}$$

where ω_k is the kth category and the class conditional distribution $p(\mathbf{x}|\omega_k)$ is organized as Gaussian mixture model in order to make use of the flexibility of mixture model. We apply the well-known EM algorithm to calculate the unknown parameters $(\pi_i, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i|\omega_k)$ to maximize the log likelihood [31]

$$\ln p(\mathbf{X}_k|\pi_k, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \sum_{n=1}^{N} \ln\left\{ \sum_{i=1}^{|\omega_k|} \pi_i \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) \right\}. \tag{4}$$

Given the class conditional probability above, we could calculate the posterior probability quite easily as given by formula (2). The posterior probability will be used to indicate the reliability of each classification and to be converted to a regression result whenever needed.

4. Experiment and results

In this section, we will validate the effectiveness of the integrated GMM approach we proposed above. Firstly 12 data sets with remarkable features are chosen for subsequent analysis. These data sets either hold their one vigilance level all the while or have slow and steady changes between different states of vigilance level. We expect to obtain a generic model suitable for everyone relying on the great descriptive power of GMM. This strategy can avoid the time-consuming training process necessary for each subject. Consequently, we merged 2 representative data sets of the 12 chosen data sets to train the model. Finally the generic model we approximated would be tested on all 12 data sets. The widely used classification approach involving SVM and LDA is used for the purpose of comparison with GMM [21]. RBF-kernel is utilized for SVM to classify each trial of EEG signals [32].

4.1. Parameter selection

To select optimal parameters with each of the 3 models, we conducted 4-fold CV on the training set. In the case of GMM, the parameters to be determined consist of the order of AR model in feature extraction subprocess, the number of mixtures of each class and the possible combinations of channels. Although we can constrain the order of AR model and number of mixtures (e.g. ≤ 20), the total number of possible parameter is still an astonishing large number, namely $20 * 20^2 * 2^{62}$. Exploring such multi-parameter space would require lots of training runs exponential to the number of possible parameters and is impossible in practice. Considering the fact that too many information is not better than no information at all [30], we decide to investigate relatively small number of channels to accomplish our experiment. This scenario also makes sense in the real-world applications where placing too many electrodes on the scalp will influence the operator's routine work.

In order to discover the relationship between the number of channels and classification performance, we choose 3 groups of channels located in the central part of scalp manually in that the optional channel space (2^{62}) is so large that cannot be searched by the CV. Here a1 means channel {FZ}, b4 means channel {FZ, FC1, FCZ, FC2}, c10 means channel {FZ, FC1, FC2, C3, C4, P3, P4, PO3, POZ, PO4}). We expect to obtain a steady and robust EEG signals using these channel groups. Fig. 4 illustrates an instance of determining the number of mixture component under a1 when the order of AR is 7.

Table 2 presents the best classification performance of CV using single channel after parameter selection process. Between the parentheses is standard deviation within the 4-fold CV, indicating the stability of each channel. This result indicates that it is feasible to estimate vigilance level using single channel. Furthermore, accuracy in the frontal lobe seems higher than in other areas of scalp, which have reasonable interpretation in brain science that frontal lobe is associated with attention and motivation [33]. Locations of electrodes are depicted in Fig. 1.

The precision and cost parameters within RBF-kernel SVM are also determined by 4-fold CV in the same way as above. By restricting each covariance of class Gaussian density being the same, LDA have no adjustable parameter and thus realize a simple but often powerful classifier.

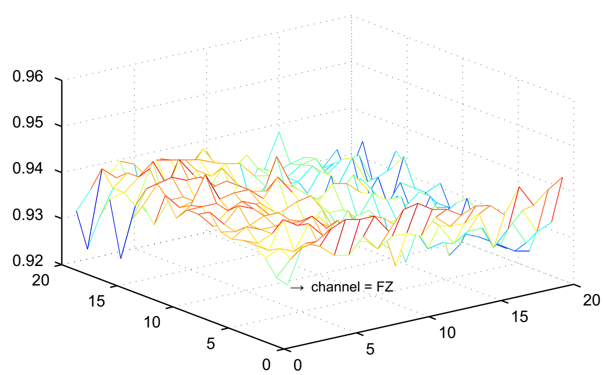


Fig. 4. 20⁎20 possibilities in the parameter space, with channel = FZ and the order of AR is 7. We notice that the accuracy rate remains steady under the square [1,10]⁎[1,10]. It has similar effect when the order of AR varies. To avoid overfitting problems we assign the number of mixtures 4, 4 respectively for subsequent study.

Table 2
Classification accuracy based on single channel using GMM.

AF1	91.29 (0.03)	FP2	92.69 (0.02)	FPZ	92.02 (0.01)	FP2	92.76 (0.01)
AF4	92.09 (0.01)	F7	90.82 (0.07)	F5	91.42 (0.01)	F3	92.89 (0.08)
F1	93.23 (0.05)	FZ	93.30 (0.07)	F2	91.69 (0.01)	F4	92.23 (0.04)
F6	89.75 (0.04)	F8	90.62 (0.07)	FT7	90.95 (0.01)	FC5	90.48 (0.04)
FC3	89.61 (0.06)	FC1	91.82 (0.03)	FCZ	91.29 (0.01)	FC2	90.15 (0.02)
FC4	92.09 (0.07)	FC6	90.21 (0.02)	FT8	88.54 (0.02)	T7	88.07 (0.06)
C5	88.27 (0.02)	C3	90.08 (0.02)	C1	91.89 (0.02)	CZ	70.71 (0.09)
C2	88.27 (0.03)	C4	87.87 (0.02)	C6	82.51 (0.01)	T8	86.86 (0.03)
TP7	87.73 (0.05)	CP5	85.52 (0.02)	CP3	81.43 (0.02)	CP1	88.67 (0.07)
CPZ	79.62 (0.04)	CP2	67.96 (0.14)	CP4	81.79 (0.03)	CP6	87.33 (0.02)
TP8	84.12 (0.04)	P7	86.66 (0.01)	P5	91.15 (0.07)	P3	79.62 (0.09)
P1	79.42 (0.08)	PZ	84.72 (0.03)	P2	93.90 (0.08)	P4	81.50 (0.01)
P6	87.47 (0.08)	P8	80.76 (0.06)	PO7	81.97 (0.09)	PO5	83.45 (0.10)
PO3	78.42 (0.04)	POZ	76.27 (0.16)	PO4	78.55 (0.03)	PO6	79.42 (0.08)
PO8	79.49 (0.10)	CB1	74.93 (0.07)	O1	76.21 (0.03)	OZ	77.55 (0.04)
O2	77.68 (0.11)	CB3	81.30 (0.08)				

4.2. Experimental results

We conducted extensive classification experiments on selected 12 data sets to show the performance of the integrated approach. All the experiments are preprocessed to eliminate artifacts and extract features in the same manner as described in Section 3. Then we compared the capabilities of different classification algorithms and also the effect of using different groups of channels.

Table 3 shows the result that high accuracy rate can be obtained when using GMM with only one channel. It is interesting to note the fact that with increasing number of channels, which means increase of features, the classification rate is degraded for all of the 3 methods consistently. We believe that the difficulty in dealing with artifact is proportional to the number of channels. In fact, more channels indeed provide more useful information while accompanying with predominant artifacts simultaneously. Steady and small number of channels is essential to guarantee our integrated approach to gain a fine result. However, it is a very promising property of our approach that we can employ a small number of electrodes in real-world applications to complete this vigilance estimation task.

It is also interesting to note that the GMM finally lead to the best results compared with SVM and LDA. We know that the GMM could supply unbounded complexity of model only by adjusting the number of mixtures per subclass. Thus the GMM could approximate the distribution of diverse EEG data sets quite well and easy while SVM moves the modeling process from optimizing the parameters to model selection. It will take lots of time to select a suitable kernel to obtain the best classification performance [34]. The effect of this density approximation can also be seen clearly in the following regression experiment. What's more, our original assumption is successful based on this generic model in predicting other subject's vigilance level. This is very valuable in the sense that given adequate training data sets, the GMM can discover the dominating nature underlying EEG signals. And more importantly, the discovered model could recognize and classify new EEG data.

Finally, to predict vigilance level continuously and reliably, we investigate the regression ability based on GMM. The posterior probability of label 2 is supposed to reflect the subject's vigilance level. The posterior probability is averaged over a 30 s long window to avoid extensive oscillation. We obtained a continuous prediction curve which could catch the main trend of vigilance level, as illustrated in Fig. 5. This regression result could also verify the suitability of the density estimation using GMM.

5. Conclusions

In this paper, a new integrated approach is proposed to predict vigilance level sequentially and automatically based on EEG signals. We combined several techniques involved in different component of estimating the vigilance level. With adequate and effective preprocessing, we obtain a reliable and reasonable high classification performance against 2 traditional methods SVM and LDA over 12 data sets. Thus we verified the ability of our integrated approach combined with GMM in dealing with EEG

Table 3
Average classification rate over 12 data sets.

Label	hGMM (%)	SVM (%)	LDA (%)
a1	88.46	80.77	73.19
b4	77.81	72.95	59.28
c10	63.95	58.16	52.41

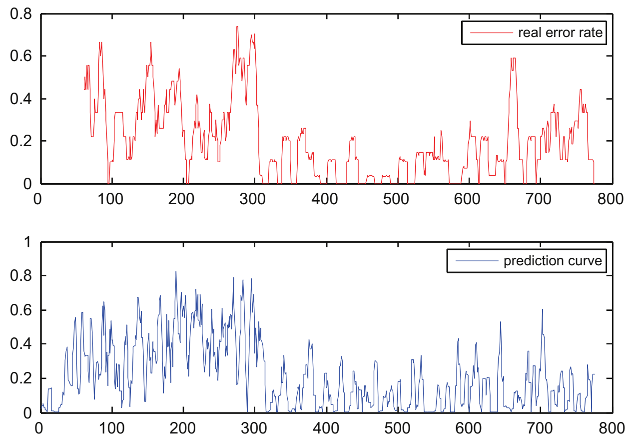


Fig. 5. The vertical axis indicates error rate the subject responded, with low error rate represents high vigilance level. The horizontal axis is the time course of this trial lasting about 5×800 s. This figure shows good agreement between quantified vigilance level and the prediction result.

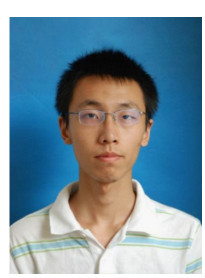
data. The regression performance based on posterior probability of GMM is proven to be promising and need further study to improve its generalization ability. Furthermore, to deal with artifact in EEG recordings more naturally, we will investigate the Dirichlet Mixture Model to improve the robustness of our method.

Acknowledgments

We are very grateful to the anonymous referees. This research was funded by the National High Technology Research and Development Program of China (no. 2008AA02Z310), 973 Program (no. 2009CB320901), and NSFC (no. 60873133).

References

[1] J. Gu, H. Liu, H. Lu, B. Lu, An integrated hierarchical Gaussian mixture model to estimate vigilance level based on eeg recordings, in: Neural Information Processing, Springer, 2011 pp. 380–387.
[2] D. Davies, R. Parasuraman, The Psychology of Vigilance, Academic Press, London, 1982.
[3] K. MacLean, S. Aichele, D. Bridwell, G. Mangun, E. Wojciulik, C. Saron, Interactions between endogenous and exogenous attention during vigilance, Attent. Percept. Psychoph. 71 (2009) 1042–1058.
[4] A. Lutz, H. Slagter, J. Dunne, R. Davidson, Attention regulation and monitoring in meditation, Trends Cognit. Sci. 12 (2008) 163–169.
[5] G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, J. Wolpaw, BCI2000: a general-purpose brain–computer interface (BCI) system, IEEE Trans. Biomed. Eng. 51 (2004) 1034–1043.
[6] B. Blankertz, K. Muller, G. Curio, T. Vaughan, G. Schalk, J. Wolpaw, A. Schlogl, C. Neuper, G. Pfurtscheller, T. Hinterberger, The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials, IEEE Trans. Biomed. Eng. 51 (2004) 1044–1051.
[7] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, T. Vaughan, Brain–computer interfaces for communication and control, Clin. Neurophysiol. 113 (2002) 767–791.
[8] G. Fabiani, D. McFarland, J. Wolpaw, G. Pfurtscheller, Conversion of EEG activity into cursor movement by a brain–computer interface (BCI), IEEE Trans. Neural Syst. Rehab. Eng. 12 (2004) 331–338.
[9] L. Li, Y. Xia, B. Jelfs, J. Cao, D. Mandic, Modelling of brain consciousness based on collaborative adaptive filters, Neurocomputing 76 (2011) 36–43.
[10] C. Lin, M. Hsieh, Classification of mental task from EEG data using neural networks based on particle swarm optimization, Neurocomputing 72 (2009) 1121–1130.
[11] E. Niedermeyer, F. Da Silva, Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, Lippincott Williams & Wilkins, 2005.
[12] J. Wolpaw, N. Birbaumer, W. Heetderks, D. McFarland, P. Peckham, G. Schalk, E. Donchin, L. Quatrano, C. Robinson, T. Vaughan, Brain–computer interface technology: a review of the first international meeting, IEEE Trans. Rehab. Eng. 8 (2000) 164–173.
[13] K. Grauman, M. Betke, J. Lombardi, J. Gips, G. Bradski, Communication via eye blinks and eyebrow raises: video-based human–computer interfaces, Univ. Access Inf. Soc. 2 (2003) 359–373.
[14] R. Duda, P. Hart, D. Stork, Pattern Classification, John Willey & Sons, Singapore, 2001.
[15] K. Mikolajczyk, B. Leibe, B. Schiele, Multiple object class detection with a generative model, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, 1 (2006) 26–36.
[16] E. Derya Übeyli, Statistics over features: EEG signals analysis, Comput. Biol. Med. 39 (2009) 733–741.
[17] H. Nolan, R. Whelan, R. Reilly, FASTER: fully automated statistical thresholding for EEG artifact rejection, Neurosci. Meth. 192 (2010) 152–162.
[18] M. Fatourechi, A. Bashashati, R. Ward, G. Birch, EMG and EOG artifacts in brain computer interface systems: a survey, Clin. Neurophysiol. 118 (2007) 480–494.
[19] G. Bartels, L. Shi, B. Lu, Automatic artifact removal from EEG-a mixed approach based on double blind source separation and support vector machine, in: Proceedings of the IEEE International Conference on Engineering in Medicine and Biology Society, 2010 pp. 5383–5386.
[20] G. Dornhege, B. Blankertz, G. Curio, K. Muller, Combining features for BCI, Adv. Neural Inf. Process. Syst. (2003) 1139–1146.
[21] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, B. Arnaldi, A review of classification algorithms for EEG-based brain–computer interfaces, Neural Eng. 4 (2007) R1.
[22] C. Rasmussen, The infinite Gaussian mixture model, Adv. Neural Inf. Process. Syst. 12 (2000) 554–560.
[23] R. Rosipal, B. Peters, G. Kecklund, T. Åkerstedt, G. Gruber, M. Woertz, P. Anderer, G. Dorffner, EEG-based drivers drowsiness monitoring using a hierarchical Gaussian mixture model, Found. Augmented Cogn. (2007) 294–303.
[24] H. Yu, L. Shi, B. Lu, Vigilance estimation based on EEG signals, in: Proceedings of IEEE/ICME International Conference on Complex Medical Engineering, 2007.
[25] L. Shi, H. Yu, B. Lu, Semi-supervised clustering for vigilance analysis based on EEG, in: Proceedings of the IEEE International Joint Conference on Neural Networks, 2007 pp. 1518–1523.
[26] I. Omerhodzic, S. Avdakovic, A. Nuhanovic, K. Dizdarevic, Energy distribution of EEG signals: EEG signal wavelet-neural network classifier, J. Biol. Life Sci. 6 (2010) 210–215.
[27] A. Belyavin, N. Wright, Changes in electrical activity of the brain with vigilance, Electroenceph. Clin. Neurophysiol. 66 (1987) 137–144.
[28] H. Ramoser, J. Muller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, IEEE Trans. Rehab. Eng. 8 (2000) 441–446.
[29] G. Pfurtscheller, C. Neuper, C. Guger, W. Harkam, H. Ramoser, A. Schlogl, B. Obermaier, M. Pregenzer, Current trends in Graz brain–computer interface (BCI) research, IEEE Trans. Rehab. Eng. 8 (2000) 216–219.
[30] C. Rasmussen, Z. Ghahramani, Occam's razor, Adv. Neural Inf. Process. Syst. (2001) 294–300.
[31] I. Nabney, NETLAB: Algorithms for Pattern Recognition, Springer Verlag, 2002.
[32] C.-C. Chang, C.-J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (2011) 27:1–27:27.
[33] D. Stuss, Functions of the frontal lobes: relation to executive functions, J. Neuropsychol. Soc. 17 (2011) 759–765.
[34] G. Cawley, N. Talbot, On over-fitting in model selection and subsequent selection bias in performance evaluation, J. Mach. Learn. Res. 11 (2010) 2079–2107.



Jing-Nan Gu received the B.S. degree in mathematics and applied mathematics from HeFei University of Technology in 2010. He is currently a M.S. candidate in the Lab of Computational Intelligence, Shanghai Jiao Tong University. His research interests include EEG-based BCI applications, mixture model and model inference.

Hongtao Lu got his Ph.D. degree in electronic engineering from Southeast University, Nanjing, in 1997. After graduation he became a postdoctoral fellow in Department of Computer Science, Fudan University, Shanghai, China, where he spent two years. In 1999, he joined the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, where he is now a professor. His research interest includes machine learning, computer vision and pattern recognition, and EEG processing. He has published more than 60 papers in international journals such as IEEE Transactions, Neural Networks and in international conferences. His papers got more than 400 citations by other researchers.

Bao-Liang Lu is a professor of Computer Science and Engineering at Shanghai Jiao Tong University (SJTU). He received his B.S. degree in instrument and control engineering from Qingdao University of Science and Technology, China, in 1982, the M.S. degree in computer science and engineering from Northwestern Polytechnical University, China, in 1989, and the Dr. Eng. degree in electrical engineering from Kyoto University, Japan, in 1994. From 1982 to 1986, he was with the Qingdao University of Science and Technology. From April 1994 to March 1999, he was a Frontier Researcher at the Bio-Mimetic Control Research Center, the Institute of Physical and Chemical Research (RIKEN), Japan. From April 1999 to August 2002, he was a Research Scientist at the RIKEN Brain Science Institute. Since August 2002, he has been a full Professor at the Department of Computer Science and Engineering, Shanghai Jiao Tong University, China. His research interests include brain-like computing, neural network, machine learning, pattern recognition, brain computer interface, computational linguistics, and computational biology and bioinformatics. He is a senior member of the IEEE.