Proceedings of International Joint Conference on Neural Networks, Orlando, Florida, USA, August 12-17, 2007

Semi-Supervised Clustering for Vigilance Analysis Based on EEG

Li-Chen Shi, Hong Yu and Bao-Liang Lu* *Senior Member, IEEE*

Abstract— **Vigilance research is very useful and important to our daily lives. EEG has been proved very effective for measuring vigilance. Up to now, many researches mainly focus on using supervised learning methods to analyze the vigilance. However, the labelled information of vigilance is hard to get and sometimes not reliable. In this paper, we proposed a semi-supervised clustering method for vigilance analysis based on EEG. This method uses the insufficient labeled information to guide the vigilance related feature selection and uses prior knowledge of vigilance state transform to guide the clustering algorithm. The experiment results show that our method can almost correctly distinguish the awake state and the sleeping state by EEG, and can also represent the transform processes of reasonable middle states between the awake state and the sleeping state.**

I. Introduction

During the past few decades, studies on vigilance have shown that vigilance analysis is very useful to our daily lives [1][2]. Vigilance, or sustained attention, refers to the ability of observers to maintain their focus of attention and to remain alert to stimuli for prolonged periods of time. For many human machine interaction systems, the operators should retain vigilance above a constant level. Otherwise, some accidents may occur. In addition, with rapid development and wide applications of robots, in order to offer high quality of service, besides recognizing the object's expressions, the robots also should be able to estimate the objects' vigilance correctly. Therefore vigilance analysis is a very important issue in human machine interaction study.

In the past several decades, various signals were used to analyze the vigilance. Among them, EEG based vigilance analysis is more accurate and faster. In EEG based vigilance research, most existing methods have focused on using supervised learning methods to analyze the vigilance [3]-[9], such as using the evoked potential (EP) response to analyze the vigilance, using group mean performance in a testing environment to analyze the vigilance, and using prior knowledge and experts experiences to analyze the vigilance.

However, till now, there is no uniform standard for vigilance scale labeling, and the existing vigilance labelling methods are complex, expensive and sometimes not reliable. Based on these considerations, we choose clustering method for vigilance analysis. Furthermore, semi-supervised clustering is more powerful than unsupervised clustering, as it can use supervising information to guide clustering algorithms towards a reasonable grouping of data and to guide similarity computing methods [10]–[12]. And in vigilance study, there surely has some labelled information or prior knowledge which can be used.

Many studies show that, the vigilance state transform during a long term is a gradual changing process [2][6]. For example, vigilance states are divided into 4 states from high level to low level. State 1 means clear-headed and state 4 means totally sleeping. The occurrence of vigilance state transform from state 2 to state 3 is more possible than from state 2 to state 4. In addition, the labelled data of clear-headed (state 1) EEG and sleeping (state 4) EEG are easy to be obtained. As a result, we can use these information to supervise the clustering process.

In our study, we use EEG for vigilance analysis. We also divide the vigilance into 4 states from high level to low level. State 1 means clear-headed; state 4 means sleeping; and other states mean middle states between clear-headed and sleeping. Firstly, utilizing insufficient labelled EEG data, we mainly use Common Spatial Patterns (CSP) [13][14] and mutual information based feature selection methods [15] to select the vigilance related features for indirectly guiding the similarity computing. Then, considering the above vigilance state transform property, we design a clustering method combining with some prior knowledge of vigilance states transform to analyze the EEG data. Experimental results show that our method can almost correctly distinguish the awake state and the sleeping state by EEG, and can also represent the transform processes of reasonable and meaningful middle states between awake state and sleeping state.

This paper is structured as follows. In section II, the methods used for vigilance analysis are described. In section III, experimental setup is briefly introduced. In section IV, experimental results are presented. Finally, some conclusions are drawn in section V.

II. Method

We use multi-channels EEG for vigilance analysis. Experiments show that the changing of EEG during vigilance state transform is a continuous process. For example, as shown in Fig. 1, the energy of EEG around 3Hz from clear-headed state to sleeping state is a gradual increasing process. So we can select the features of EEG which can well separate the labelled clear-headed and sleeping data as the vigilance related features. As the features are the inputs of the similarity computing algorithm, this strategy is indirect to supervise the similarity computing, cluster the related features and use the vigilance state transform property to supervise the adjustment of the clustering results.

Asterisk indicates corresponding author. This work was supported in part by the National Natural Science Foundation of China under the grants NSFC 60375022 and NSFC 60473040, and the Microsoft Laboratory for Intelligent Computing and Intelligent Systems of Shanghai Jiao Tong University.
L. C. Shi, H. Yu and B. L. Lu are with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240 China. E-mail: {lch-shi; yuhong1984; bllu}@sjtu.edu.cn.

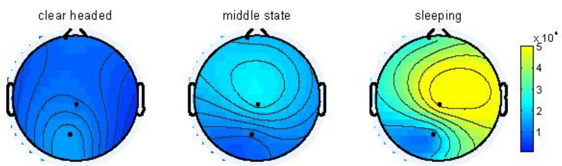


Fig. 1. Distribution of EEG energy around 3Hz on the scalp

The whole process consists of three parts. Firstly, EEG signals are preprocessed for artifact reduction and decomposed for extracting the vigilance related signals. Secondly, based on the preprocessed EEG, related features are extracted, and the appropriate features are selected by using mutual information based method. Finally, by utilizing vigilance states transform property, an extended graph factorization clustering (XGFC) model is proposed for clustering the vigilance states.

A. EEG Preprocessing

The original EEG signals contain a lot of artifacts or unrelated signals. For achieving the goal of analyzing vigilance states correctly and reliably, the artifacts must be rejected and the unrelated signals must be reduced maximally.

Generally speaking, there are two types of artifacts [16]. The first type is extra cerebral source artifact which is recorded together with EEG, such as electrooculogram (EOG), electromyogram (EMG), and electrocardiogram (ECG). The second type is technical artifacts which results from the EEG recording system, such as signal drift and decay.

In our experiments, a 128-channels NeuroScan system was used to record EEG signals. The extra cerebral source artifacts mainly consist of EOG and EMG signals induced by movement. The EOG signals were removed by Scan4.3 software installed in NeuroScan System. And the obvious EMG signals were rejected by hand. For the high performance of NeuroScan system, the technical artifacts could be ignored except the signal drift which could also be corrected by Scan4.3 software.

Besides artifacts, there exist a lot of background signals which are unrelated to vigilance change. Therefore, we need a decomposition method which can minimize the amount of background signals. Suppose we take the background signals as noise signals. As we know, there are a lot of classical or effective decomposition methods. But unfortunately, as the energy of noise signals is much greater than the energy of interested signals, most of them are unavailable for this situation. Here we used a decomposition method based on CSP [13][14] which is effective and specific for EEG signals decomposition.

CSP can seem as a variation of Principal Components Analysis (PCA). By using the CSP method, two kinds of EEG signals are whitened and then projected to the common spatial patterns. After that, the spatial patterns, to which the corresponded variances of the two kinks of EEG signals are most different, are chosen as the projection factors. Finally, the EEG signals are decomposed using the projection factors. CSP projection can be formulated as

$$Z = PV \tag{1}$$

where V denotes the original signals, P denotes the projection matrix and Z denotes the decomposed signals. For example, two kinds of EEG signals are expressed as X_1 and X_2, respectively. Actually, X_i is the combinations of events-related signals S_i and background signals S_i^b, namely

$$X_i = [C_i, C_i^b] \begin{bmatrix} S_i \\ S_i^b \end{bmatrix} \tag{2}$$

where $[C_i, C_i^b]$ is the combination coefficients matrix, $i \in \{1, 2\}$. Assume S_1^b and S_2^b are the same background signals, then CSP can be used to extract the events-related signals S_1 and S_2.

As we see, CSP is only available for labelled two-category problem. However, the vigilance analysis is a multi-category problem and the labelled information is insufficient. To deal with this problem, we propose a new strategy as follows. Firstly, we coarsely divide the EEG signals into three categories, namely clear-headed, sleeping, and others. Then, we use the labelled clear-headed and sleeping EEG signals as two conditions to get the common spatial patterns. Finally, we choose the common spatial patterns to form projection matrix P_{csp} which optimally separates the clear-headed and sleeping EEG signals as the projection factors to transform the whole EEG signals. Let X and Y denote the whole artifact rejected EEG signals and the projected signals, respectively, then

$$Y = P_{\text{csp}} X \tag{3}$$

where X is a matrix with dimension of k by l, and Y is a matrix with dimension of m (number of selected CSP) by l.

As clear-headed state and sleeping state are two terminal states of vigilance and the EEG changing during vigilance state transform is a continuous process, the whole EEG changing process should be reflected on these projected spatial patterns and the middle states of vigilance should be separated by projected to the selected spatial patterns of the two terminal states.

B. Feature Extraction and Selection

Many vigilance researches show that vigilance changing is mainly reflected by Power Spectral Density (PSD) changing of EEG signals [4][9]. So firstly, we use discrete short time Fourier transform to extract the PSD of each CSP projected EEG signal Y and take the PSD bellow 50Hz as the feature information with frequency resolution 1Hz.

$$V_{\text{psd}} = STFT(Y) \tag{4}$$

where $STFT$ denotes short time Fourier transform, and V_{psd} is the PSD matrix with dimension $50m$ by n (number of time window).

Then we use PCA to reduce the dimension of the feature matrix

$$V_R = P_R V_{\text{psd}} \tag{5}$$

where P_R is the matrix of principal spatial patterns with dimension u by $50m$ and V_R is the dimension reduced feature matrix with dimension u by n.

After performing PCA, we use a mutual information based feature selection method [15] to choose a subset S_{mi} of the feature set S. This method selects feature subset by optimizing max-relevance between feature subset and target class, and min-redundancy among the feature subset. Denote the i-th feature of EEG signals by x_i and the vigilance states by c. As we only have two types of labels: clear-headed and sleeping. We just use these labelled information to select the feature subset. The reason for doing like this is just like what is for selecting the projection matrix mentioned above. Let $I(x_i; c)$ denote the mutual information between x_i and c. Then the relevance between feature subset S_k and class c can be defined as

$$D = \frac{1}{|S_k|} \sum_{x_i \in S_k} I(x_i; c) \tag{6}$$

and the redundancy among the feature subset can be defined as

$$R = \frac{1}{|S_k|^2} \sum_{x_i, x_j \in S_k} I(x_i; x_j) \tag{7}$$

where $I(x_i; x_j)$ is the mutual information between x_i and x_j. The criterion of mutual information based feature selection method is to maximize $D - R$. The criterion operator can be defined as

$$\Phi = D - R. \tag{8}$$

Thus, the selected feature subset should maximize Φ.

In practice, we choose the feature subset S_{mi} by an incremental search method as described in [15]. And adjust feature subset S_{mi} according the clustering results from next part. Finally, we get a feature matrix V_{fs} with dimension u_{fs} by n, where u_{fs} is the number of elements in S_{mi}.

C. Extended Graph Factorization Clustering Model

Proper clustering method can mine the intrinsic relations of a given data set. Combined with some supervising information, clustering method can get even better results to interpret the intrinsic relations of the given data set. Here, we propose an extended graph factorization clustering model (XGFC), which is based on graph-factorization clustering method (GFC) [17]. After GFC, it uses vigilance states transform property to adjust the clustering results.

Firstly, we briefly introduce GFC. GFC is based on the pairwise data similarities which assigns data to clusters in a probabilistic way. GFC can also afford the relations among clusters in a probabilistic way. As illustrated in Figs. 2(a) and 2(b), the main idea of GFC is that for any pairwise data relations graph, there exists a latent bipartite graph according to which the data were generated and the pairwise data relations graph were formed. In Figs. 2(a) and 2(b), v_i denotes the observed data, u_i denotes the latent cluster, the edges between two nodes denote the relations of them. The objective of GFC is to estimate the relations between v_i and u_j by which the observed pairwise data relations can be mostly approximated.

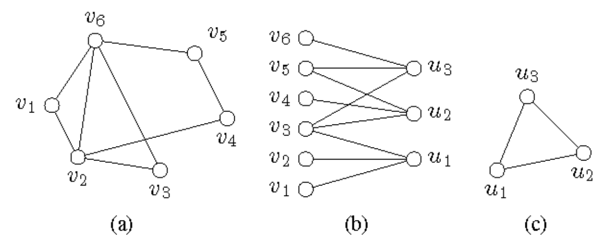


Fig. 2. (a) The original graph representing data relations; (b) The bipartite graph representing data to cluster relations; (c) The induced graph representing clusters relations

The algorithm of GFC is described as follows. Let W denote the matrix of pairwise data relations with dimension n by n, and B denote the matrix of relations between data and clusters with dimension n by k (number of clusters). From the perspective of Markov Random Walks, the relations among V_i can be expressed as

$$\widetilde{W} = (B\Lambda^{-1}B^T), \ \Lambda = diag(\lambda_1, ..., \lambda_k) \tag{9}$$

where $\lambda_j = \sum_{i=1}^{n} B_{ij}$. If we want to get an optimal estimation of B, the divergence between W and $\widetilde{W}$ must be minimized. To make the problem easy to be solved, we replace $B\Lambda^{-1}$ by H. Then the objective function is expressed as

$$min\{l(W, H\Lambda H^T)\}, \ \text{s.t.} \sum_{i=1}^{n} H_{ip} = 1 \tag{10}$$

where $l(\cdot,\cdot)$ is a divergence operator. Let $l(X, Y) = \sum_{i,j}[X_{ij} \log(X_{ij}/Y_{ij}) - X_{ij} + Y_{ij}]$, then the objective function in Equation (10) can be reduced by the following update rule

$$\tilde{H_{ip}} \propto H_{ip} \sum_j \frac{W_{ij}}{(H\Lambda H^T)_{ij}} \lambda_j H_{jp}, \quad \sum_i \tilde{H_{ip}} = 1 \tag{11}$$

$$\tilde{\lambda_p} \propto \lambda_p \sum_{ij} \frac{W_{ij}}{(H\Lambda H^T)_{ij}} H_{ip} H_{jp}, \quad \sum_p \tilde{\lambda}_p = \sum_{ij} W_{ij}. \tag{12}$$

Finally, we get the data cluster relations

$$B = H\Lambda. \tag{13}$$

Then the relations between data and clusters can be considered as the probability that the data belong to the clusters. In Fig. 2(c), the relations W^c among clusters can also be estimated from the perspective of Markov random walks.

$$W^c = (B^T D^{-1} B), D = diag(d_1, ..., d_n) \tag{14}$$

where $d_i = \sum_{j=1}^{k} B_{ij}$. If we consider the above relations in a probabilistic way, then we can get the following results

$$p(v_i, u_j) \propto B_{ij} \tag{15}$$

$$p(u_i, u_j) \propto W^c_{ij} \tag{16}$$

$$p(v_i) \propto d_i \tag{17}$$

$$p(u_j) \propto \lambda_j. \tag{18}$$

Considering the vigilance state transform is a gradual changing process, we propose a state transform model which is shown in Fig. 3. We divide vigilance into 4 states. The edges in this model indicate whether there exists transform probabilities between two states during a short time. For example, in our assumption, there is no edge between state 1 and state 4, this means during a short time state 1 and state 4 can not directly transform to each other. If there exists an edge, we directly use $p(u_i, u_j)$ as the transform probability.

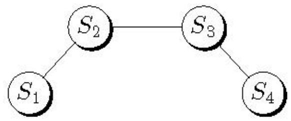


Fig. 3. Each node denotes a vigilance state, and each edge denotes there existing direct transform between these two states

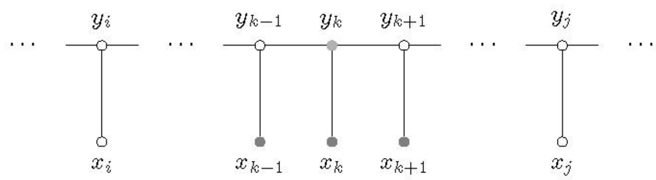


Fig. 4. The final clustering label y_k is determined by x_{k-1}, x_k and x_{k+1}, where x_i denotes observation, y_i denotes clustering label of x_i.

Based on the state transform model, we design a conditional probability model which works after GFC. This model uses the neighbor observations to help the current observation adjust its clustering result as shown in Fig. 4. The detailed algorithm is described as follows.

Let $\{x_i\}$ and $\{y_i\}$ denote the observation sequence and the label sequence, respectively. Then the relevance (x_{i-1}, x_i, x_{i+1}) for $y_i = s_t$ is shown as follows.

$$R(s_t, x_i) = P_v(s_t|x_i) \tag{19}$$

$$R(s_t, x_{i-1}) = \sum_j P_u(s_t|s_j) P_v(s_j|x_{i-1}) \tag{20}$$

$$R(s_t, x_{i+1}) = \sum_j P_u(s_t|s_j) P_v(s_j|x_{i+1}) \tag{21}$$

where P_v is the conditional probability of data to clusters, P_u is the conditional probability of clusters to clusters, and $\{s_i\}$ is the cluster set. P_v and P_u can be calculated by Equations (15–18). As there is no direct link from x_{i-1} (or x_{i+1}) to y_i, the revelance between them is taken by utilizing the path $x_{i-1}y_{i-1}y_i$ (or $x_{i+1}y_{i+1}y_i$). Then we define a criterion function as

$$\begin{aligned} y_i = argmax_{s_t}\{D(s_t|x) = \alpha_{-1}R(s_t, x_{i-1}) \\ +\alpha_0 R(s_t, x_i) + \alpha_1 R(s_t, x_{i+1})\} \end{aligned} \tag{22}$$

where α_j is the coefficient which reflects the contribution of each x_k to the target label y_i. This function considers neighbors' contribution. And the coefficient α_j can be adjusted to get reasonable grouping of data. This process can be seen as a local optimization of conditional random fields (CRF) [19].

In summary, the whole processing of EEG signals can be described as follows.

- Firstly, EEG preprocessing including noise reduction and CSP processing is carried out.
- Secondly, EEG signals are transformed to the frequency domain, then PCA and mutual information are used to perform feature extraction and selection.
- Thirdly, GFC is used to cluster the EEG data so as to get the probability information.
- Finally, conditional probability model (Equation 21-22) is used to adjust clustering results.

III. Experimental Setup

A total of 16 healthy volunteers whose ages are from 19 to 25 took part in our study. Each subject performed at least four turns of experiments. The experiments were carried in a small room with normally illuminated and insulated. The temperature of the room was kept at about 24 degrees and the humidity was kept between 20% and 40%.

During the experiment, the subject was asked to lie on bed, close eyes and try to release until falling asleep. The EEG signals were acquired though the NeuroScan System. 64 channels of signals including 62 channels of EEG and 2 channels of EOG are recorded. Electrodes are arranged based on extended 10/20 system as shown in Fig. 5. Each experiment lasts at least one hour. During this time, a period of soft and short music was presented to the subject several times. The music lasted 10 seconds and the volume of the music was tuned such that the subject would not be disturbed when the subject was sleeping. If the subject listens to the music, the subject should open his or her eyes, which shows that he or she is awake. If not, the subject just does nothing. This means that he or she falls asleep. We also use a DV camera to record the subject's face expressions.

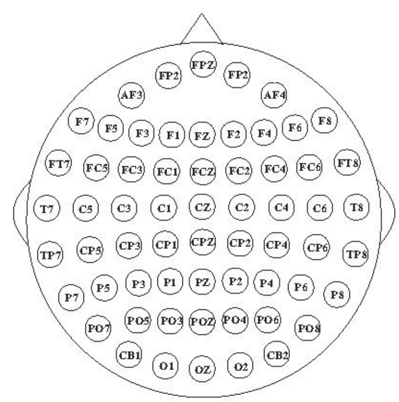


Fig. 5. Electrodes distribution of extended 10/20 system

After each experiment, we used the feedback of the subjects combined with the facial expressions from the video recorded to label the clear-headed EEG and sleeping EEG. The EEG around the period when playing the music was discarded. Only when both sides estimation of vigilance states were the same, the EEG was labelled.

IV. Clustering Results

After acquiring the EEG data, we use K-mean, normalized-cut [18], GFC and XGFC to cluster the EEG data in different situations, then compare and analyze the different clustering results according to the subject's feedback, video and the insufficient labelled data. During clustering the EEG data, we make a decision on the current vigilance state of the subject every 4 seconds. Fig. 6 shows the waveforms of the original EEG data. The sharp peaks in the figure are the EOG signals.

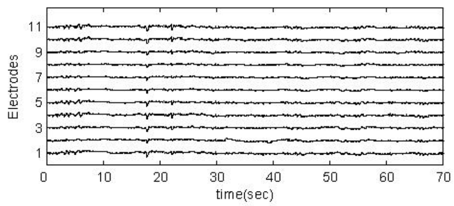


Fig. 6. The original EEG data

A. Directly Clustering the Original EEG Data

We use K-Mean and Normalized-Cut to directly cluster the PSD of the original EEG data into two states: wake (not clear-headed state) and sleep. This EEG data is recorded in one day from subject A. The results are shown in Figs. 7(a) and 7(b). For the resolution limited, there are overlaps in the figure. Here, overlaps do not mean that two states appeared at the same time point. Instead, the two states just emerged alternately during that period of time. From the figure we may conclude that the subject is awake during the first 15 minutes and the last 10 minutes. This is also verified by the subject after the experiments (wake: during the first 20 minutes and the last 10 minutes). However, there are many overlaps. According to the subject's feedback some of them are obvious wrong.

B. Using Vigilance Related Feature Selection

Here, we use the vigilance related features to cluster the EEG data. Figs. 7(c) and 7(d) show the results of clustering the data after vigilance related feature selection. Firstly, we use K-Mean and Normalized-Cut as the clustering algorithms to cluster the same EEG data in Fig. 7(a). Comparing with directly clustering the original EEG data, we can see that many overlaps disappeared. And the results are more close to the subject's feedback and the observations from the recorded video. Thus we can see that the feature selection process can effectively improve the performance of clustering algorithm.

Nextly, we cluster the vigilance related features of EEG data using GFC and XGFC. Figs. 7(e) and 7(f) show the results. From these figures we can see that GFC gets a similar result as K-Mean or Normalized-Cut, while XGFC gets even better results which are almost consistent with the subject's feedback and the observation from recorded video.

Although XGFC can greatly improve the grouping of EEG data, carefully observation on Fig. 7(f) reveals a clear

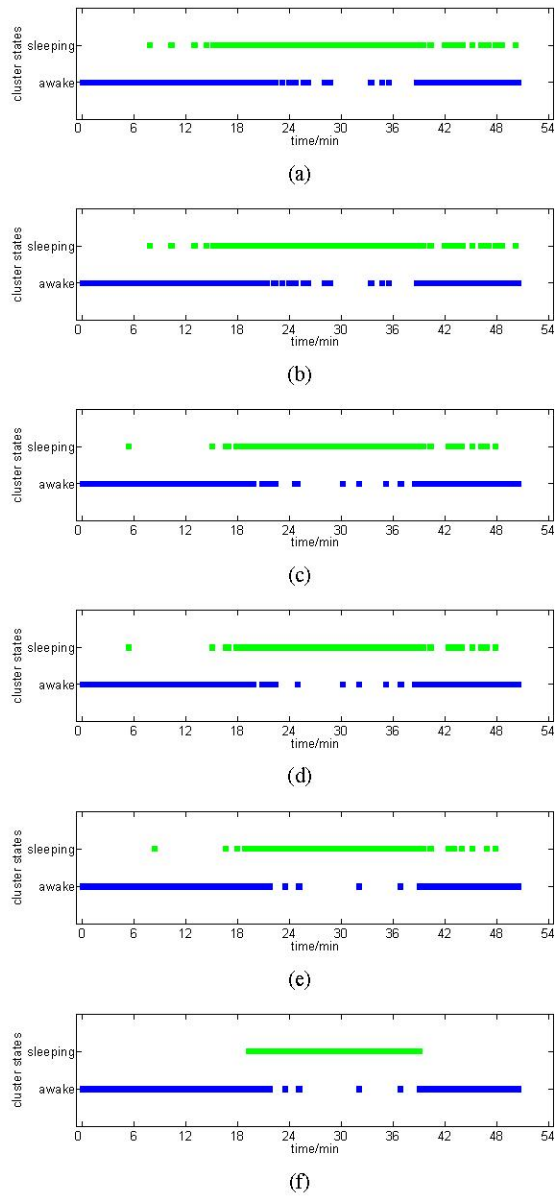
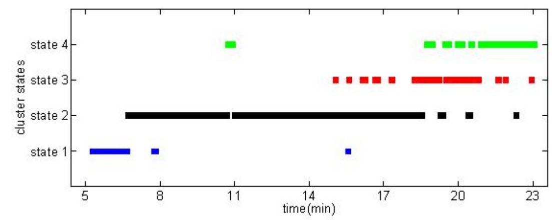


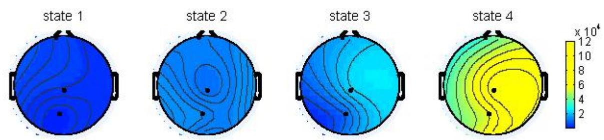
Fig. 7. Clustering results of vigilance states. 1) Cluster the PSD of the original EEG data using (a) K-Mean algorithm; (b) Normalized-Cut algorithm; 2) Cluster the data after vigilance related feature selection using (c) K-Mean algorithm; (d) Normalized-Cut algorithm; (e) GFC algorithm; (f) XGFC algorithm with $\alpha_1 = \alpha_{-1} = 0.4, \alpha_0 = 0.2$

overlap around the time 20 minutes. This may be due to the existence of middle states when falling asleep. To examine this observation, we cluster the data around that particular time into four vigilance states as shown in Fig. 8(a). These four states can be easily distinguished. In order to verify the legitimacy of the clustering result, we calculate the average EEG spectrum of each states around 3Hz as shown in Fig. 8(b). From this figure we can see that the average energy from state 1 to state 4 is gradual increasing. This phenomenon is consistent with physiological results.

Besides these, the feature patterns calculated from one subject in one day combined with XGFC are also applicable to cluster the EEG data from the same subject in other days

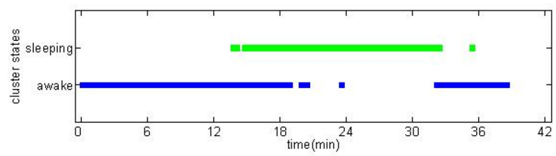


(a)

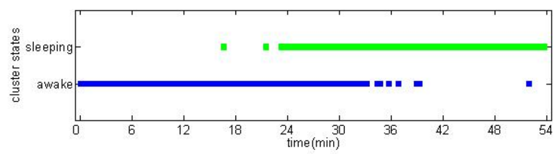


(b)

Fig. 8. (a) Clustering the data from wake to sleep using XGFC; (b) Distributions of different states EEG energy with bandwidth around 3Hz on the scalp



(a)



(b)

Fig. 9. (a) Clustering result of EEG data from the same subject but in another day. (b) Clustering result of EEG data from another subject

or even from different subjects. Fig. 9 shows the results, which are close to the subject's feedback and the observing results from the recorded video.

V. Conclusions

In this paper, we have proposed a semi-supervised clustering method for vigilance analysis based on EEG signals. Firstly, we used the insufficient labelled information to guide the vigilance related feature selection indirectly supervised the similarity computing. Then considering the vigilance states transform property, we proposed the XGFC model for EEG data clustering based on local optimization of CRF. From the experimental results, we can see that vigilance related feature selection process is very helpful to improve the performance of clustering algorithms. In addition, by using condition probability model, the XGFC model can get even better and reasonable grouping of the EEG data. As a result, although labelled information in vigilance studies is very poor, proper semi-supervised clustering can still get meaningful results. In the future, we will continue improving the clustering algorithm and use the clustering results to guide the vigilance labelling and vigilance estimation.

Acknowledgments

The authors would like to thank the volunteers for taking part in our vigilance experiments. The authors also would like to thank Zhong-Yi Zhao, Zhi-Fei Ye, Mu Li, Jia-Wei Fu and Zuo-Fan Li for their helpful work on EEG data acquisition.

References

[1] R. Molloy, R. Parasuraman,"Monitoring an automated system for a single failure: vigilance and task complexity effects," *Human Factors*, vol. 38, pp. 311–322, 1996.

[2] M. Weinger,"Vigilance, Boredom, and Sleepiness," *Journal of Clinical Monitoring and Computing*, vol. 15, pp. 549–552, 1999.

[3] S. Makeig, "Using Feedforward Neural Networks to Monitor Alertness from Changes in EEG Correlation and Coherence," *Advances in Neural Information Processing Systems*, pp. 931–937, 1996.

[4] T. P. Jung, S. Makeig, M. Stensmo and T. J. Sejnowski, "Estimating Alertness from the EEG Power Spectrum," *IEEE Transactions on Biomedical Engineering*, pp. 60–69, vol. 44, 1997.

[5] T. Shimada, T. Shiina and Y. Saito, "Detection of Characteristic Waves of Sleep EEG by Neural Network Analysis," *IEEE Transactions on Biomedical Engineering*, vol. 47, pp. 369–379, 2000.

[6] D. H. Loewy, K. B. Campbell, D. R. de Lugt, M. Elton and A. Kok, "The Mismatch Negativity during Natural Sleep: Intensity Deviants," *Clinical Neurophysiology*, vol. 111, pp. 863–872, 2000.

[7] S. Roberts, I. Rezek, R. Everson, H. Stone, S. Wilson and C. Alford, "Automated Assessment of Vigilance using Committees of Radial Basis Function Analysers," Advances in Medical Signal and Information Processing, pp. 231-237, 2000.

[8] A. Vuckovic, V. Radivojevic, A. C.N. Chen and Dejan Popovic, "Automatic Recognition of Alertness and Drowsiness from EEG by an Artificial Neural Network," *Medical Engineering & Physics*, vol. 24, pp. 349–360, 2002.

[9] S. F. Liang, C. T. Lin, R. C. Wu, Y. C. Chen, T. Y. Huang and T. P. Jung, "Monitoring Driver's Alertness Based on the Driving Performance Estimation and the EEG Power Spectrum Analysis," 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 5738–5741, 2005.

[10] S. Basu, A. Banerjee and R. Mooney, "Semi-supervised Clustering by Seeding," International Conference on Machine Learning, pp. 19–26, 2002.

[11] S. Basu, M. Bilenko and R. J. Mooney, "A Probabilistic Framework for SemiSupervised Clustering," Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 59–68, 2004.

[12] M. Bilenko, S. Basu and R. J. Mooney, "Integrating Constraints and Metric Learning in Semi-Supervised Clustering," Proceedings of the 21st International Conference on Machine Learning, pp. 81–88, 2004.

[13] A. C. K. Soong and Z. J. Koles, "Principal-Component Localization of the Sources of the Background EEG," *IEEE Transactions on Biomedical Engineering*, vol. 42, pp. 59–67, 1995.

[14] Y. Wanga, P. Bergc and M. Scherga, "Common Spatial Subspace Decomposition Applied to Analysis of Brain Responses under Multiple Task Conditions: A Simulation Study," *Clinical Neurophysiology*, vol. 110, pp. 604–614, 1999.

[15] H. Peng, F. Long and C. Ding, "Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1226–1238, 2005.

[16] P. Anderer, S. Roberts, A. Schlogl, G. Gruber, G. Klosch, W. Herrmannh, P. Rappelsbergerc, O. Filz, M. J. Barbanoj, G. Dorffner and B. Saletu, "Artifact Processing in Computerized Analysis of Sleep EEG – A Review," *Neuropsychobiology*, vol. 40, pp. 150–157, 1999.

[17] K. Yu, S. Yu, V. Tresp, "Soft Clustering on Graphs," *Advances in Neural Information Processing Systems*, pp. 1553–1560, 2006.

[18] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 888–905, 2000.

[19] John Lafferty, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," *Proceedings of the Eighteenth International Conference on Machine Learning*, pp.282–289, 2001.