Semi-Supervised Clustering for Vigilance Analysis Based on EEG

Li-Chen Shi, Hong Yu and Bao-Liang Lu
Department of Computer Science and Engineering
Shanghai Jiao Tong University
800 Dong Chuan Rd., Shanghai 200240, China
Email: {lch-shi, yuhong1984, bllu}@sjtu.edu.cn

***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 labeled 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 inadequate labeled information to guide the vigilance related feature selection and uses prior knowledge of vigilance state transform to guide the clustering algorithm. Analyzing results show that our method can almost correctly distinguish the awake and sleeping EEG. The results can also present reasonable middle states transform processes.**

I. Introduction

During the past few decades, studies on vigilance have shown that vigilance analysis is very useful to our daily lives [1][2][3]. 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 quick development and wide applications of robots, in order to offer high quality service, besides recognizing the object's expressions, the robots also should be able to estimate the objects' vigilance correctly. So vigilance analysis is a very important issue in human machine interaction field.

Up to now, many signals were proposed to analyze the vigilance. Among them all, EEG based vigilance analysis is more accurate and faster. In EEG based vigilance research, most past methods have focused on using supervised learning methods to analyze the vigilance [3]-[13]. Such as using the evoked potential (EP) response to analyze the vigilance, using group mean performance in a testing environment to analyze the vigilance, or using prior knowledge and experts experiences to analyze the vigilance.

However, till now, there is no uniform standard for vigilance scale labeling. As the existing vigilance labeling 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 the clustering algorithm towards a reasonable grouping of data and to guide the similarity computing method [14]-[17]. And in vigilance study, there is still some labeled information or prior knowledge to be used, so we propose a semi-supervised clustering method for vigilance analysis.

Many studies show that, during a long term vigilance states transform is a gradual changing process [2][9]. 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 states transform from state 2 to state 3 is more possible than from state 2 to state 4. In addition, the labeled data of clear-headed (state 1) EEG and sleeping (state 4) EEG is easy to get. So we can use these information to supervise the clustering process.

In our study, we use EEG for vigilance analysis. We divided 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. Firstly, by utilizing the inadequate labeled EEG data, we mainly use Common Spatial Patterns (CSP) [18][19]and mutual information based feature selection methods [20][21]to select the vigilance related features for indirectly guide the similarity computing. Then, considering the above vigilance states transform property, we design a clustering method combined with some prior knowledge of vigilance states transform to analyze the EEG data. Analyzing results show that just in several seconds our methods can almost correctly distinguish the awake and sleeping EEG. And the results can also present reasonable and meaningful middle states transform processes.

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, analyzing results are presented. Finally, some conclusions are drawn in section V.

II. Methods

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

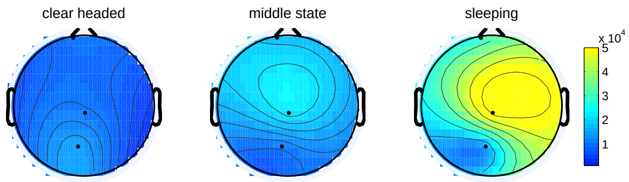


Fig. 1. Distribution of EEG energy around 3Hz on the scalp

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, extended graph factorization clustering (XGFC) model is proposed for clustering the vigilance states.

A. EEG Preprocessing

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 [22].the first type is extra cerebral source artifact which is recorded together with EEG, such as electrooculogram (EOG), electromyography (EMG), and ECG. The second type is technical artifact which results from the EEG recording system, such as signal drift and decay.

In our experiments, 128-channels NeuroScan System was used to record EEG signals. The extra cerebral source artifacts mainly consisted of EOG and EMG 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. So 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 Common Spatial Patterns (CSP) [18][19]which is effective and specific for EEG signals decomposition.

CSP can seem as a variation of Principal Components Analysis (PCA). The method of CSP is that 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, use the projection factors to decompose the EEG signals. 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, denote the two kinds of EEG signals as X_a and X_b. Both of them are the combinations of events-related signals and background signals.

$$X_a = [C_{a1}, C_{a2}] \begin{bmatrix} S_a \\ S_{c1} \end{bmatrix}, \; X_b = [C_{b1}, C_{b2}] \begin{bmatrix} S_b \\ S_{c2} \end{bmatrix} \tag{2}$$

where S_a and S_b are the events-related signals, S_{c1} and S_{c2} are the background signals, C_{ai} and C_{bi} are the combination coefficients. Assume the S_{c1} and S_{c2} are the same background signals, then CSP can be used to extract the events-related signals S_a and S_b.

As we see, CSP is only available for labeled two categories problem. However, the vigilance analysis is a multiple categories problem and in which the labeled information is inadequate. So the strategy is that, firstly we coarsely divide the EEG signals into three categories clear-headed, sleeping and others. Then we use the labeled clear-headed and sleeping EEG signals as the two conditions EEG signals 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 process EEG signals. Denote the whole artifact rejected EEG signals as X, the projected signals as Y, then

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

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 the two terminal states of vigilance and the EEG changing during vigilance states 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 PSD changing of EEG signals [4][6][13]. So firstly, we use discrete short time Fourier transform to extract the PSD of each CSP projected EEG signals Y and take the PSD bellow $50Hz$ as the feature information with frequency resolution $1Hz$.

$$V_{PSD} = STFT\{Y\} \tag{4}$$

where $STFT$ denotes short time Fourier transform, and V_{PSD} is the PSD matrix with dimension $50 \times M$ by N (number of time window).

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

$$V_R = P_R V_{PSD} \tag{5}$$

where P_R is the matrix of principal spatial patterns with dimension m by $50 \times M$, and V_R is the dimension reduced feature matrix with dimension m by N.

After that, we use a mutual information based feature selection method [20][21]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 ith feature of EEG signals as x_i and denote the vigilance states as c. As we only have two types of labels clear-headed and sleeping. We just use these labeled information to select the feature subset. The reason for doing like this is just like what is for selecting the projection matrix mentioned above. Denote the mutual information between x_i and c as $I(x_i; c)$. Then the relevance between feature subset S_k and class c is defined as follow,

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

and the redundancy among the feature subset is 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,

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

So, the selected feature subset should maximize Φ.

In practice, we choose the feature subset S_{MI} by an incremental search methods as described in []. And adjust feature subset S_{MI} according the following clustering results. Finally, we get a feature matrix V_{FS} with dimension m_{FS} by N, where m_{FS} is the elements number 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) [23]. After GFC, it use vigilance states transform property to adjust the clustering results.

Firstly, let's briefly introduce the 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. The main idea of GFC is that for any pairwise data relations graph, there exist a latent bipartite graph according to which the data was generated and the pairwise data relations graph was formed. See in Figure 2(a),2(b). Where 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.

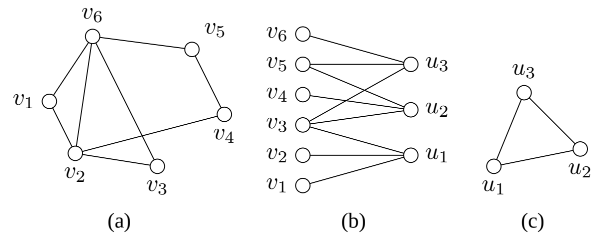


Fig. 2. (a) The original graph representing data relations; (b) The bipartite graph representing data to cluster relations; (c) The induced graph representing clusters relations

The algorithm of GFC is described as follow. Denote W as the matrix of pairwise data relations with dimension N by N, and B as 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 of among V_i can be formed 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 a optimal estimation of B, the divergence between W and $\widetilde{W}$ must be minimized. To make the problem easy to solve, replace $B\Lambda^{-1}$ by H. Then the objective function is formed,

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

where $l(\cdot,\cdot)$ is a divergence operator. Defines $l(X, Y) = \sum_{i,j}[X_{ij}log(X_{ij}/Y_{ij}) - X_{ij} + Yij]$, 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 seem as the probability that the data belong to the clusters. In Figure 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.

$$p(u_i, v_j) \propto B_{ij} \tag{15}$$

$$p(v_i, v_j) \propto W_{ij}^c \tag{16}$$

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

$$p(v_j) \propto \lambda_j \tag{18}$$

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

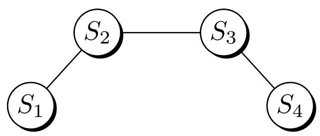


Fig. 3. Each node denotes a vigilance state, and each edge denotes there existing direct transform between these two states

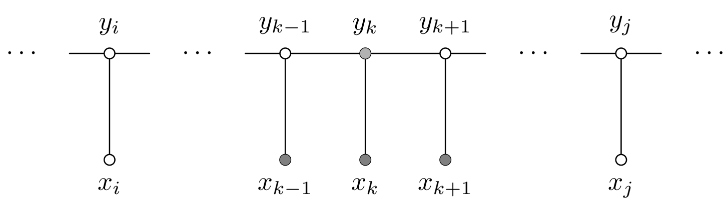


Fig. 4. x_i denotes observation, y_i denotes clustering label of x_i. The final clustering label y_k is determined by x_{k-1}, x_k and x_{k+1}

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 which is shown in Figure 4. The detailed algorithm is described as follow. Denote P_v as the conditional probability of data to clusters, P_u as the conditional probability of clusters to clusters, $\{x_i\}$ as the observation sequence, $\{y_i\}$ as the label sequence, and $\{s_i\}$ as the cluster set. Firstly we define the relevance of x_{i-1}, x_i and x_{i+1} to $y_i = s_i$.

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

$$R(s_i, x_{i-1}) = \sum_{j,k} P_u(s_i|s_j) P_v(s_j|x_{i-1}) \tag{20}$$

$$R(s_i, x_{i+1}) = \sum_{j,k} P_u(s_i|s_j) P_v(s_j|x_{i+1}) \tag{21}$$

where P_v and P_u can be calculated by Equations (15) (16) (17) (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 their relations with y_{i-1} or y_{i+1}. Then we define a criterion function as,

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

where α_i is the coefficient which reflect the contribution of each x_i to the target label y_i. This function considers neighbors' contribution. And the coefficient α_i can be adjusted to get reasonable grouping of data.

The XGFC can be described as follow. Firstly, use GFC to clustering the EEG data for get the probabilistic information. Then use conditional probability model to adjust clustering results.

III. Experimental Setup

A total of 16 healthy volunteers (ages from 19 to 25) took part in our study. Each subject performed at least four turns of experiments. 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 arrange based on extended 10/20 system. Figure 5 shows the electrodes distribution. 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 volume of the music was tuned such that the subject would not be disturbed when the subject was sleeping. If the subject heard the music which shown that he or she was awake, the subject just opened his or her eyes. If not, the subject just kept on sleeping and it meant that he or she fallen asleep. We used a DV camera to record the subject's activities.

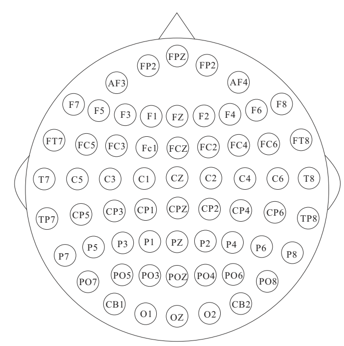


Fig. 5. Electrodes distribution of extended 10/20 system

After each experiment, we used the subjects feedback combined with the video disk to label the clear-headed EEG and sleeping EEG. The EEG around the period when the music is played is discarded. Only when both sides estimation of vigilance states were the same, the EEG was labeled.

IV. Clustering Results

After getting the EEG data, we use K-Mean, Normalized-Cut [24], GFC and XGFC to cluster the EEG data in different situations, then compare and analyze the different clustering results. During clustering the EEG data, we make a decision on the current vigilance state of the subject every 4 seconds. Figure 6 shows the waveforms of original EEG data. The sharp peaks in the figure are the EOG signals.

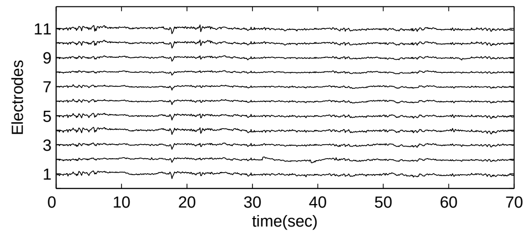


Fig. 6. The original EEG data

A. Results of Directly Cluster the Original EEG Data

We use K-Mean and Normalized-Cut to directly cluster the PSD of the original artifact rejected EEG data into two states: wake and sleep. This EEG data is recorded in one day from subject A. The results is shown in Figure 7(a)7(b). Overlaps in the figure mean there exist some middle states. 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, and according to the subject's feedback some of them are obvious wrong.

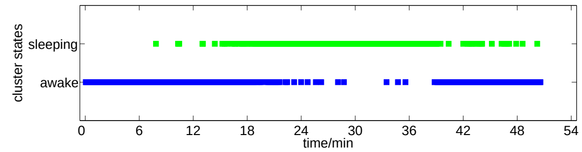
B. Clustering Results after Vigilance Related Feature Selection

Here, we use the vigilance related features to cluster the EEG data. Figure 7(c)7(d) shows results of clustering the data after vigilance related feature selection. We firstly use K-Mean and Normalized-Cut as the clustering algorithms to cluster the same EEG data in Figure 7(a). Compared to directly clustering the original EEG data, we can see that many overlaps disappeared. And the results is more close to subject's feedback and the observations from the recorded video. So the feature selection process can effectively improve the performance of clustering algorithm.

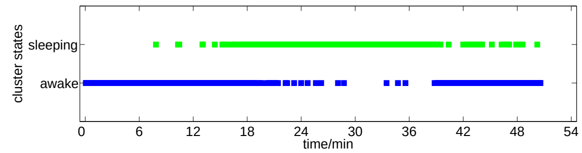
Nextly, we cluster the vigilance related features of EEG data using GFC and XGFC. Figure 7(e)7(f) shows the results. From this figure we can see that GFC get a similar result as K-Mean or Normalized-Cut, while XGFC get even better results which is almost consistent with subject's feedback and video observing results.

Although XGFC can greatly improve the grouping of EEG data, carefully observation on figure 7(f) reveals a clear overlap around the time 20 minutes. This may be due to multiple vigilance states when falling asleep. As a result, we cluster the data around that particular time into four vigilance states. Figure 8(a) shows the result. There 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 $3Hz$ which is shown in Figure 8(b). From the figure we can see the average energy from state 1 to state 4 is gradual increasing. This phenomenon is consistent with physiological result which reflects our clustering result is reasonable and meaningful.

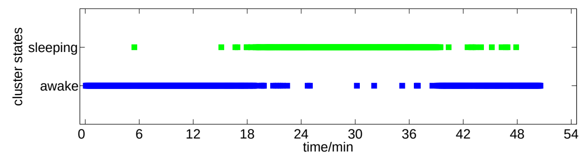
Besides these, the feature patterns calculated from one subject in one day combined with XGFC are also applicable



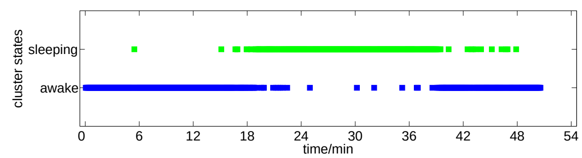
(a)



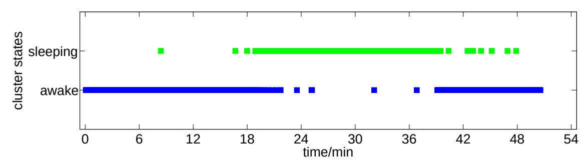
(b)



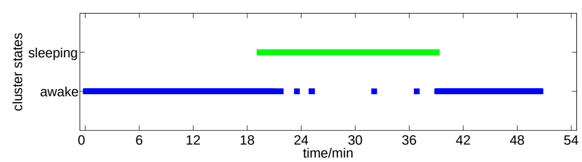
(c)



(d)



(e)



(f)

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.

to cluster the EEG data from the same subject in other days or even from different subjects. Figure 9 shows the results which is close to the subject's feedback and the video observing results.

V. Conclusions

In this paper, we proposed a semi-supervised clustering method for vigilance analysis based on EEG signals. Firstly, we used the inadequate labeled information to guide the vigilance related feature selection indirectly supervised the similarity computing. Then considering the vigilance states transform property, we proposed XGFC model for EEG data

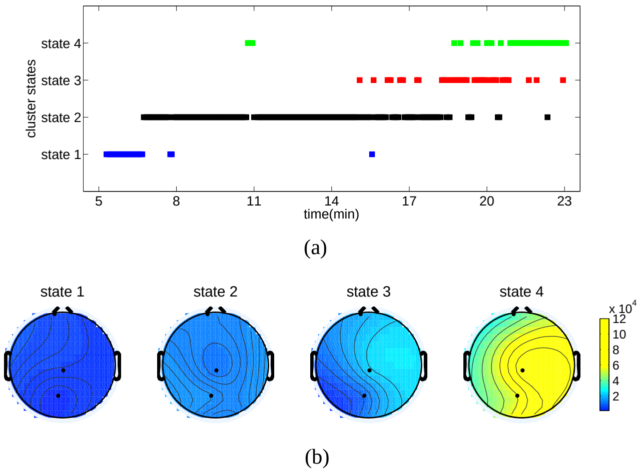


Fig. 8. (a) Cluster the data from wake to sleep using XGFC; The states shown in this figure are from state 1 to state 4 in a bottom up order. (b) Distributions of different states EEG energy with bandwidth around 3Hz on the scalp.

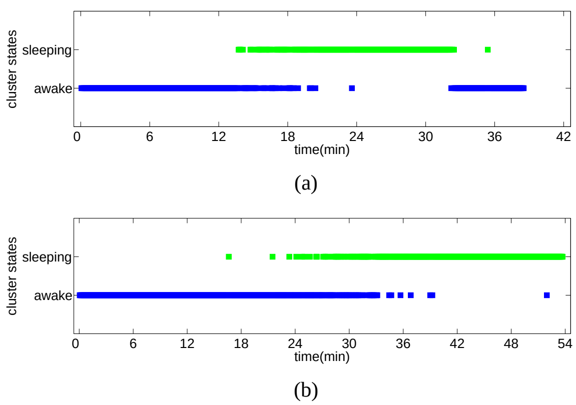


Fig. 9. (a) Clustering result of EEG data from the same subject but in another day. (b) Clustering result of EEG data from another person.

clustering. From the 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 data. So, although labeled 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, use the clustering results to guide the vigilance labeling and vigilance estimation.

Acknowledgment

The authors would like to thank the volunteers for taking part in our vigilance experiments. The authors also would like to thank Zhongyi-Zhao, Zhifei-Ye, Mu-Li, Jiawei-Fu and Zuofan-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] C. Cajochen, J. M. Zeitzer, C. A. Czeisler and D. J. Dijk, "Dose-response Relationship for Light Intensity and Ocular and Electroencephalographic Correlates of Human Alertness," *Behavioural Brain Research*, vol. 115, pp. 75–83, 2000.

[4] T. P. Jung and S. Makeig, "Estimating Level of Alertness from EEG," *Engineering in Medicine and Biology Society*, vol. 2, pp. 1103–1104, 1994.

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

[6] 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.

[7] K. Hyoki, M. Shigeta, N. Tsuno, Y. Kawamuro, T. Kinoshita, "Quantitative Electro-oculography and Electroencephalography as Indices of Alertness," *Electroencephalography and clinical Neurophysiology*, vol. 106, pp. 213–219, 1998.

[8] 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.

[9] 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.

[10] B. J. Wilson and T. D. Bracewell, "Alertness Monitor using Neural Networks for EEG Analysis," Neural Networks for Signal Processing, vol. 2, pp. 814–820, 2000.

[11] 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.

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

[13] 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," Engineering in Medicine and Biology Society, pp. 5738–5741, 2005.

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

[15] D. Cohn, R. Caruana, A. McCallum, "Semi-supervised Clustering with User Feedback," *Proc. Technical Report TR2003-1892*, Cornell University, 2003.

[16] S. Basu, M. Bilenko, 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.

[17] M. Bilenko, S. Basu, 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.

[18] 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.

[19] Y. Wanga, P. Bergc, 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.

[20] A. Kraskov, H. Stogbauer and P. Grassberger, "Estimating mutual information," *Physical Review*, E 69, 066138, 2004.

[21] 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.

[22] 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.

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

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