EEG Hidden Information Mining Using Hierarchical Feature Extraction and Classification

Deng Wang, Duoqian Miao, Chen Xie, Hongyun Zhang
Department of Computer Science and Technology, Tongji University
Key Laboratory of Embedded System & Service Computing, Ministry of Education of China, Tongji University
Shanghai, 201804, China
E-mail: w_deng208@hotmail.com; miaoduoqian@163.com

***Abstract*—This paper presents a hierarchical feature extraction and classification method for electroencephalogram (EEG) hidden information mining. It consists of supervised learning for fewer features, hierarchical knowledge base (HKB) construction and classification test. First, the discriminative rules and the corresponding background conditions are extracted by using autoregressive method in combination with the nonparametric weighted feature extraction (NWFE) and k-nearest neighbor. Second, through ranking the discriminative rules according to validation test correct rate, a hierarchical knowledge base HKB is constructed. Third, given an EEG sequence, it chooses one or several discriminative rules from the HKB using the up-bottom search strategy and calculates classification accuracy. The experiments are carried out upon real electroencephalogram (EEG) recordings from five subjects and the results show the better performance of our method.**

Index Terms—Electroencephalogram, feature extraction, EEG classification, discriminative rule, hierarchical knowledge base.

I. INTRODUCTION

The electroencephalogram (EEG) signal is believed to contain considerable information with regard to mental activity. By recording and analyzing EEG signals from scalp, more brain's mental activities patterns can be mined. One of important research fields bases on security testing such as guilty knowledge tests, deception detection, lie detection or concealed information tests [1-6]. The applications may involve criminal justice and counterterrorism. For example, P300-based EEG signal analysis can determine the presence or absence of specific details such as a crime or group affiliation stored in investigator's brain and this method has been suggested as an alternative approach for conventional polygraph [7].

Usually, application of pattern recognition approach on EEG signal includes following steps: signal acquisition, preprocessing, feature extraction, feature selection and classification. In this study, an important step is extraction of relevant features of EEG signals which containing necessary information of EEG signals. Recently, much attention has been paid to feature extraction methods for EEG. Polat [8] used discrete fourier transform (DFT) and Subasi [9] used discrete wavelet transform (DWT) as the extraction methods. Morphological features, frequency features and wavelet features were defined and evaluated for lie detection in [10]. Autoregressive (AR) model [11], band power [12] and fractal dimension [13] to extract features from recorded signals were also considered and applied on EEG signals. The problem is that the analysis of EEG signals from channels often results in high-dimensional data vectors including all spatiotemporal information. With large increases in dimensionality, processing time will increase significantly. In addition, the number of samples available for training is relatively small compared to their dimensionality. Thus, an open question is how to map the high dimensional feature space into a lower dimensional space while maintaining signal information and get an accurate classification results in a short time? This is especially useful for an EEG-based communication system (brain-computer interface, BCI) designs as it can reduce the complexity of the classification problem and increase the information transfer rate in BCI applications. Inspired by human problem solving, hierarchical features extraction and classification make human decision-making adapt to the changeable real world. Therefore, in our EEG signal analysis, the open question can be represented by how to form a hierarchical feature space and what level of it is sufficient to predict accurately the related information from high-dimensional EEG signals.

In this study, we proposed a hierarchical feature extraction and classification method for EEG hidden information mining including three stages: supervised learning, hierarchical knowledge base construction and classification test. The autoregressive (AR) method in combination with the nonparametric weighted feature extraction (NWFE) [14-15] method were employed for feature extraction. In order to evaluate the validate of the method, similar to other recognition methods, this paper included a paradigm for conducting the test and recording brain signals for analyzing the records to detect the target from non-targets. The scenario in the test was a mock attention (birthday) which was similar to paradigms used in other studies on lie detection. Finally, k-nearest neighbor (k-NN) classifier was implemented on the data. Experimental results showed the performance include classification accuracy and runtime of our method. Under ten-fold cross validation, the average classification accuracy can achieve up to 89.1%.

II. SYSTEM ARCHITECTURE

In pattern classification applications, EEG-based BCI technology has huge applicable potentials since the corresponding recording method for EEG signals is relatively

Proc. 9th IEEE Int. Conf. on Cognitive Informatics (ICCI'10)
F. Sun, Y. Wang, J. Lu, B. Zhang, W. Kinsner & L.A. Zadeh (Eds.)
978-1-4244-8040-1/10/$26.00 ©2010 IEEE

convenient, inexpensive, harmless and possesses a high temporal resolution [16]. However, the demerit of recording EEG signal such as low signal-noise rate and low spatial resolution makes the improving of classification accuracy become a challenging aim for researchers. On the other hand, the information transfer rates is another key factor for evaluating the performance of a BCI system which means the processing time should be decreased. Generally, many methods report a good classification accuracy while also need to long time for training in order to get a test optimization. Inspiration from human cognition, fast or real-time decision-making with environment is more useful instead of exhaustive search for getting global optimization. From [17], we know that the local information is important, and from analysis of human cognition, hierarchical features extraction and classification [18-19] also investigated in our study.

A. Supervised learning

Learning set is fed into for extracting discriminative rules. It is well known that feature extraction is a crucial step affecting both performance and computation time of classification systems in pattern recognition applications. Nonparametric weighted feature extraction (NWFE) [14] is one of the useful tools for solving this problem. The main ideas of NWFE are putting different weights on every sample to compute the “weighted means” and compute the distance between samples and their weighted means as their “closeness” to boundary, and then define nonparametric between-class and within-class scatter matrices which put large weights on the samples close to the boundary and de-emphasize those samples far from the boundary. In order to reduce the space of features, autoregressive (AR) coefficients are computed by Burg’s method using points of the original EEG signals.

An autoregressive model with time-varying coefficients p is defined by

$$Y_{i,c}(t)=\sum_{i=1}^{p}a_{i,c}Y_{i,c}(t-i), \tag{1}$$

where $a_{i,c}$ is the i^{th} coefficient of the AR model for channel c. $i=1,\cdots p$.

The nonparametric within-class scatter matrix and between-class scatter matrix for L classes is defined as same as follows:

$$S_w^{NW}=\sum_{i=1}^{L}P_i\sum_{l=1}^{N_i}\frac{\lambda_l^{(i,j)}}{N_i}(x_l^{(i)}-M_i(x_l^{(i)}))(x_l^{(i)}-M_i(x_l^{(i)}))^T, \tag{2}$$

$$S_b^{NW}=\sum_{i=1}^{L}P_i\sum_{\substack{j=1\\ j\neq i}}^{L}\sum_{l=1}^{N_i}\frac{\lambda_l^{(i,j)}}{N_i}(x_l^{(i)}-M_j(x_l^{(i)}))(x_l^{(i)}-M_j(x_l^{(i)}))^T \tag{3}$$

where P_i denotes the prior probability of class i, N_i is training sample size of class i, and $x_l^{(i)}$ refers to the l^{th} sample from class i. The $\lambda_l^{(i,j)}$ denotes the scatter matrix weight which is the function of $x_l^{(i)}$ and $M_j(x_l^{(i)})$.

$$\lambda_l^{(i,j)}=\frac{dist(x_l^{(i)},M_j(x_l^{(i)}))^{-1}}{\sum_{l=1}^{N_i}dist(x_l^{(i)},M_j(x_l^{(i)}))^{-1}}, \tag{4}$$

where dist(a, b) denotes the Euclidean distance from a to b. If the distance between $x_l^{(i)}$ and $M_j(x_l^{(i)})$ is small, then its weight will be close to 1; otherwise, $\lambda_l^{(i,j)}$ will be close to 0. The sum of $\lambda_l^{(i,j)}$ for class i is 1. $M_j(x_l^{(i)})$ denotes the weighted mean of $x_l^{(i)}$ in class j and defined as

$$M_j(x_l^{(i)})=\sum_{k=1}^{N_j}w_{lk}^{(i,j)}x_k^{(j)}, \tag{5}$$

where

$$w_{lk}^{(i,j)}=\frac{dist(x_l^{(i)},x_k^{(j)})^{-1}}{\sum_{k=1}^{N_j}dist(x_l^{(i)},x_k^{(j)})^{-1}} \tag{6}$$

is the weight of a sample $x_l^{(i)}$ in class i corresponding to $x_k^{(j)}$ in class j. If the distance between $x_l^{(i)}$ and $x_k^{(j)}$ is small, then its weight $w_{lk}^{(i,j)}$ will be close to 1; otherwise, $w_{lk}^{(i,j)}$ will be close to 0. The sum of the $w_{lk}^{(i,j)}$ is 1.

The optimal features f are the f eigenvectors with largest f eigenvalues of the following matrix:

$$(S_w^{NW})^{-1}S_b^{NW}. \tag{7}$$

Two advantages of using the nonparametric scatter matrices mentioned in [14] are: First, they are generally of full rank. Second, the nonparametric nature of scatter matrices reduces the effects of outliers and works well even for non-normal datasets.

B. Hierarchical knowledge base (HKB) construction

HKB = <F, K, R>, which is a 3-tuple where F is the feature set; K is the number of neighbors with the same classification. It corresponds background references called K-domain, and parameter R between 0 and 1 represents the ability of correct classifying under F and K.

The construction algorithm is as follows.

1) Using equation (1) to extract AR coefficients for all learning samples.

2) Divide learning set LS into M subsets.

3) Select one of subsets to validation test set VTS, and the others to validation learning set VLS.

4) Validation feature set F = ∅.

5) Select a column from VLS and add it to F.

6) Divide VLS to Target class and Non-target class.

7) Calculate S_w^{NW} and S_b^{NW} for each class using equation (2) and equation (3), respectively.

8) Calculate the optimal features f using equation (7).

9) Calculate class probability R using KNN classifier (using the Euclidean distance) and record the corresponding best K value. Thus, a piece of discriminative rule DRi = <R, F, K> is obtained.

10) Repeat 5-9 until all feature columns from VLS been added to F.

11) Repeat 3-10 for all subsets.

12) Sort all DRi to HKB = <R, F, K> by R in descending order.

13) Return HKB.

Note that the goal of this algorithm is not to learn a functional relation, but a set of discriminative rules, then through ranking the discriminative rules according to class probability R, a hierarchical knowledge base HKB was constructed. The different size of feature set F and the different K value with the corresponding class probability R make ways to scale of feature number for predicting the classification results.

C. Classification test

In order to evaluate the learning algorithm, a testing algorithm can be performed according to the following steps.

1) Load the HKB.

2) Select a testing sample from test set TS.

3) Test feature set F' $=\varnothing$.

4) Select a discriminative rule DRi from HKB using the up-bottom search strategy and add it to F'.

5) Calculate class probability R' by using test feature set F' and KNN classifier.

6) Record the best R' and the corresponding class label of the sample.

7) Repeat 4-6 for all DRi.

8) Repeat 2-7 for all testing sample.

9) Calculate classification accuracy (Acc)

10) Return Acc.

To sum up, the main idea of this algorithm is to assign a label to a sample using a set of previously assessed discriminative rules on the best feature level.

III. EXPERIMENTS AND RESULTS

A. Materials

The data [20] used in this paper were recorded during a simple oddball test (an autobiographical paradigm). Each subject was asked to say 5 numbers, and one of them was his birth year (such as: 1980, 1981, 1983, 1986 and 1984). The subject didn't say his birth date till the end of experiment. The numbers were displayed to the subject randomly and with 30 repetitions (Totally 5*30 numbers displayed to the subject in each experiment). Each number displayed 1 second and between the numbers, the screen is empty for 2 second. The signals were recorded at frontal (Fz), central (Cz), and parietal (Pz) electrode positions from 10-20 international system.

All sites were referenced to the linked mastoids. Vertical EOG was also recorded for blink artifact detection. EEG signals were digitally sampled with 256 Hz frequency. To provide an idea of the nature of the signals distribution, Figure 1 gives an illustration of the EEG recorded including target and non-target stimulus on channel Pz from one of the subjects.

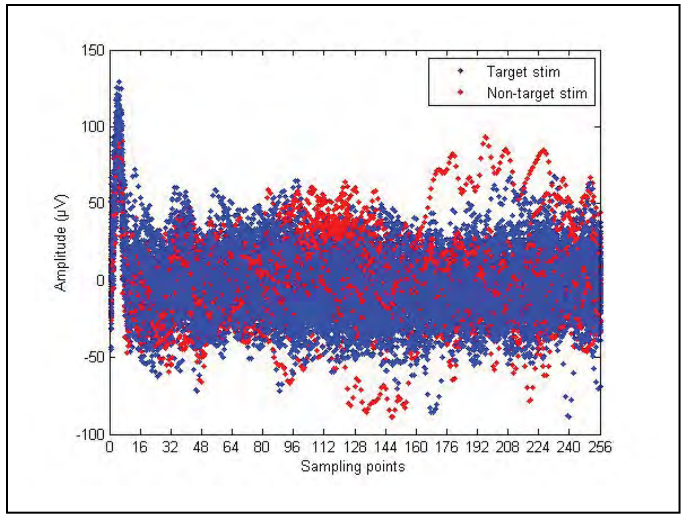

Figure 1. An illustration of the EEG signals recorded on channel Pz from one of the subjects before feature extraction. The time horizon is one second and thus includes 256 sampling points. The vertical axis is the voltage value in µV. Blue points represent target stimulus, i.e. subject's birth year. Blue points are not the target stimulus.

B. Pre-processing

Before constructing a HKB and the cross validation, several preprocessing operations were applied to the data. First, using MATLAB software offline, all data were digitally filtered in the 0.3-30 Hz range. This is the frequency range which is used typically in P300-based concealed information detection studies [21]. Second, normalization methods were applied to EEG signals for reducing large measurement noise or high variability. In our experiments, four different normalization methods were tested, respectively.

1) Min-max normalization 1 (minmax[0+1])

$$D' = \frac{D - \min(D)}{\max(D) - \min(D)}(U - L) + L, \tag{8}$$

where D' is the normalized data matrix, D is the natural data matrix and U and L are the upper and lower normalization bound. The equation maps the data matrix into a bound between 0 and 1.

2) Min-max normalization 2 (minmax[-1+1])

Above equation maps the data matrix into a bound between -1 and 1.

3) Nonliear normalization (log)

$$D' = \frac{\log(D - \min(D) + 1)}{\log(\max(D) - \min(D) + 1)} . \tag{9}$$

4) Zero-mean normalization (z-score)

$$D' = (D - \overline{D}) / \sigma , \tag{10}$$

where $\overline{D}$ is the mean of the natural data matrix D and σ is the standard deviation of the same data matrix.

Third, the continuous EEG signals were separated into several epochs. Each epoch is a one-second long. Epochs containing eye blinks in EOG channel above 400μv were discarded.

C. Experimental results

All the experiments of this section were done over five subjects. In order to evaluate the performance of our proposed method, k-fold cross-validation method was employed. In the present work, we let k = 10, that is to say, the data from each subject was divided into ten subsets randomly, one of the subsets was used as the test set and the other nine subsets were put together to form a learning sets. Then this procedure was repeated ten times and the average classification accuracy was computed. According to several experimental test runs, the order in autoregressive method was chosen at six.

In Table I, we report the average accuracy and runtimes[1] for all of possible combinations of channels across five subjects. It shows that there is no apparent difference in runtimes. However, the channel Pz performs the best classification accuracy and the accuracy does not improve with the number of channels.

TABLE I. AVERAGE ACCURACY OBTAINED AND TOTAL RUNTIMES (IN MINUTES) SPENT ACROSS FIVE SUBJECTS BY TESTED CHANNELS

Channel	Accuracy	Runtimes
Fz	0.830	21.298
Pz	0.838	21.832
Cz	0.836	21.281
Fz, Pz	0.825	21.354
Fz, Cz	0.817	21.221
Pz, Cz	0.820	21.141
Fz, Pz, Cz	0.818	21.153

In Table II, the average classification accuracy obtained by using min-max normalization (minmax[0+1], minmax[-1+1]), nonliear normalization (log) and zero-mean normalization (z-score), respectively. Based on the results, we can see that the minmax[-1+1] and log methods are competitive in accuracy than other normalization methods in the present study.

[1] These runtimes were obtained using an Intel Pentium Dual CPU E2140 @ 1.60GHz, 1GB RAM.

Finally, we select channel Pz as the target channel, minmax[-1+1] as the normalization method for classification. We use 10-fold cross-validation and the results shows in table III. Subject 1 gives the best classification accuracy 93.6% and the whole average classification accuracy can achieve up to 89.1% which can be applied in BCI system in future. In addition, the total runtime is also acceptable in real-time applications.

TABLE II. AVERAGE ACCURACY RESULTS USING FIVE DIFFERENT NORMALIZATION METHODS

Subject	Normalization			
	Minmax [0+1]	***Minmax [-1+1]***	***log***	***z-score***
s1	0.797	0.801	0.806	0.803
s2	0.810	0.833	0.811	0.815
s3	0.835	0.851	0.855	0.852
s4	0.869	0.858	0.864	0.835
s5	0.787	0.834	0.826	0.784
Average	**0.819**	**0.835**	**0.832**	**0.818**

TABLE III. FINAL CLASSIFICATION RESULTS

Subject	No. of Samples	Accuracy
s1	144	0.936
s2	278	0.822
s3	245	0.850
s4	115	0.918
s5	145	0.929
Average		**0.891**

IV. CONCLUSION

In this study, we developed a simple and feasible hierarchical feature extraction and classification method. Using autoregressive method to acquire fewer feature space of EEG signals and in combination with the nonparametric weighted feature extraction, a hierarchical knowledge base was constructed. Finally, the proposed method was successfully applied to EEG signals for the hidden information mining. The classification accuracy and runtime can be acceptable in real-time applications.

ACKNOWLEDGMENT

The authors gratefully acknowledge Dr. Vahid Abootalebi and the research center of intelligent signal processing (RCISP), Iran, for providing the data. The authors are supported from the National Natural Science Foundation of China (No. 60970061, No. 60775036, No. 60475019) and the Research Fund for the Doctoral Program of Higher Education (No. 20060247039).

References

[1] L.A. Farwell, E. Donchin, “The Truth Will Out: Interrogative Polygraphy ("Lie Detection") With Event-Related Brain Potentials”, Psychophysiology, vol. 28, pp. 531-547, 1991.

[2] L.A. Farwell, S.S. Smith, “Using brain MERMER testing to detect concealed knowledge despite efforts to conceal”, Journal of Forensic Sciences, vol. 46, pp. 1-9, 2001.

[3] A.R. Miller, J.P. Rosenfeld, M. Soskins, M. Jhee, “P300 amplitude and topography in an autobiographical oddball paradigm involving simulated amnesia”, Journal of Psychophysiology, vol. 16, pp. 1-11, 2002.

[4] J.P. Rosenfeld, B. Cantwell, V.T. Nasman, V. Wojdac, S. Ivanov, L. Mazzeri, “A modified, event-related potential-based guilty knowledge test”, International Journal of Neuroscience, vol. 24, pp. 157-161, 1988.

[5] J.P. Rosenfeld, “Event-related potentials in the detection of deception, malingering, and false memories”, In M. Kleiner (Ed.), Handbook of Polygraph Testing. Academic Press, New York, 2002.

[6] J.P. Rosenfeld, M. Soskins, G. Bosh, A. Ryan, “Simple effective countermeasures to P300-based tests of detection of concealed information”, Psychophysiology, vol. 41, pp. 205-219, 2004.

[7] V. Abootalebi, M.H. Moradi, M.A. Khalilzadeh, “A comparison of methods for ERP assessment in a P300-based GKT”, International Journal of Psychophysiology, vo. 62, pp. 309–320, 2006.

[8] K. Polat, S. Günes, “A novel data reduction method: Distance based data reduction and its application to classification of epileptiform EEG signals”, Applied Mathematics and Computation, vol. 200, pp. 10-27, 2008.

[9] A. Subasi, “EEG signal classification using wavelet feature extraction and a mixture of expert model”, Expert Systems with Applications, vol.32, pp. 1084-1093, 2007.

[10] V. Abootalebi, M.H. Moradi, M.A. Khalilzadeh, “A new approach for EEG feature extraction in P300-based lie detection”, Computer Methods and Programs in Biomedicine, vol. 94, pp. 48-57, 2009.

[11] C.W. Anderson, E.A. Stolz, S. Shamsunder, “Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks”, IEEE Transactions on Biomedical Engineering, vol. 45, pp. 277-286, 1998.

[12] W. Klimesch, M. Doppelmayr, H. Russegger, T. Pachinger, J. Schwaiger, “Induced alpha band power changes in the human EEG and attention”, Neuroscience Letters, vol. 244, pp. 73-76, 1998.

[13] A. Accardo, M. Affinito, M. Carrozzi, F. Bouquet, “Use of the fractal dimension for the analysis of electroencephalographic time series”, Biological Cybernetics, vol. 77, pp. 339-350, 1997.

[14] B.C. Kuo, D.A. Landgrebe, “Nonparametric weighted feature extraction for classification”, IEEE Transactions on Geoscience and Remote Sensing, vol. 42, pp. 1096-1105, 2004.

[15] C.T. Lin, K.L. Lin, L.W. Ko, S.F. Liang, B.C. Kuo, I.F. Chung, “Nonparametric single-trial EEG feature extraction and classification of driver’s cognitive responses”, EURASIP Journal on Advances in Signal Processing, 2008.

[16] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, “Brain-computer interfaces for communication and control”, Clinical Neurophysiology, vol. 113, pp. 767-791, 2002.

[17] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, San Diego, 1990.

[18] M. Ishikawa, K. Doya, H. Miyamoto, T. Yamakawa, “Hierarchical feature extraction for compact representation and classification of datasets”, 14th International Conference on Neural Information Processing, ICONIP 2007, pp. 556-565.

[19] Chulhee Lee. Feature extraction and classification algorithms for high-dimensional data. Purdue University West Lafayette, IN, USA. 1993.

[20] V. Abootalebi, M.H. Moradi, and M.A. Khalilzadeh, “Detection of the cognitive components of brain potentials using wavelet coefficients”, Iranian Journal of Biomedical Engineering, vol. 1, pp. 25-46, 2004 (Article in Persian).

[21] J.P. Rosenfeld, M. Soskins, G. Bosh, A. Ryan, “Simple effective countermeasures to P300-based tests of detection of concealed information”, Psychophysiology, vol. 41, pp. 205-219, 2004.