Sensors & Transducers, Vol. 169, Issue 4, April 2014, pp. 235-240

Sensors & Transducers

Feature Optimize and Classification of EEG Signals: Application to Lie Detection Using KPCA and ELM

[1] GAO Junfeng, [1] QIU Jianhui, [1] ZHANG Wenjia, [2] YANG Yong
[1] Key Laboratory of cognitive science (South-Central University for Nationalities), State Ethnic Affairs Commission, Wuhan, 430074, China
[2] School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, 330077, China
[1] Tel.: +86 13163332686
E-mail: junfengmst@163.com

Received: 30 December 2013 Accepted: 7 March 2014 Published: 30 April 2014

Abstract: EEG signals had been widely used to detect liars recent years. To overcome the shortcomings of current signals processing, kernel principal component analysis (KPCA) and extreme learning machine (ELM) was combined to detect liars. We recorded the EEG signals at Pz from 30 randomly divided guilty and innocent subjects. Each five Probe responses were averaged within subject and then extracted wavelet features. KPCA was employed to select feature subset with deduced dimensions based on initial wavelet features, which was fed into ELM. To date, there is no perfect solution for the number of its hidden nodes (*NHN*). We used grid searching algorithm to select simultaneously the optimal values of the dimension of feature subset and *NHN* based on cross-validation method. The best classification mode was decided with the optimal searching values. Experimental results show that for EEG signals from the experiment of lie detection, KPCA_ELM has higher classification accuracy with faster training speed than other widely-used classification modes, which is especially suitable for online EEG signals processing system. *Copyright © 2014 IFSA Publishing, S. L.*

Keywords: EEG signal processing, Lie detection; Extreme learning machine; Kernel principal component analysis; P300.

1. Introductions

Mental realness detection, called as lie detection, is one of the hot issues in neuroscience filed. With development of neuroscience, Event Related Potential (ERP) has been conducted on lie detection. P300 (P3) [1], an endogenous ERP component, is in response to meaningful stimuli. Because of the lower signal-to-noise ratio (SNR) of ERP under the single response, it is very difficult to conduct lie detection when the simple features of P300 waveform, such as amplitude and latency, were collected. Thus, this study attempted to improve the robustness and accuracy of lie detection by extracting wavelet feature from EEG acquisition. Feature selection is an important aspect of machine learning problems, and usually carries out to further optimize and reduce original feature space dimension, to remove redundant features, to improve classification accuracy. In this paper, we applied KPCA, which is widely used feature selection method, to optimize and reduce original feature space dimension, and select the best feature subset. Subspace dimension can be got from sample training. Hence, Kernel Principal Component Analysis (KPCA), a widely-used feature selection method, was applied to further select optimal features from the initial feature sets in order to further increase the accuracy of lie detection.

Lie detection method based on machine learning, relative to BAD (Bootstrapped Amplitude Difference) [2] and BCD (Bootstrapped Correlation Difference) [3], can extract various features and use the latest machine learning algorithms. Experimental results show that the accuracy of the machine learning methods have been improved to some extent [4-5]. Gao et al. used Support Vector Machine (SVM) in the lie detection and showed that SVM outperformed fisher discrimination analysis (FDA) and back-propagation neural network (BPNN) by obtaining the highest classification accuracy of 91.8 % [6].

SVM is a new machine learning method and is proposed by Vapnik and other scholars on the basis of statistical learning theory. SVM is based on VC dimension theory and structural risk minimization principle. Its basic idea is to map a linear inseparable data into a high-dimensional space by kernel function. By solving linearly constrained quadratic programming in the space, SVM finds the optimal separating surface with minimization structural risk.

When developing a detection system for EEG task, we should take into account the time of training the classification models. Hence, taking this issue into account, SVM and BPNN may be unsuitable due to their high computational cost. Extreme Learning Machine (ELM) [7-9] randomly specifies the input weights and biases, and then analytically calculates the output weights with the smallest norm. Hence, ELM tends to provide good generalization power at extremely faster training speed. During the past several years, ELM has drawn considerable attention in many fields related to pattern recognition. Traditional ELM randomly specifies the number of biases, but improper biases number impacts classification results.

Base on the above consideration, a novel model based on KPCA algorithm and ELM is proposed. It can be drawn the conclusion from experimental results that this new method can obtain better prediction accuracy and less prediction time which other methods can not achieve.

2. Experimental Protocol and EEG Acquisition

2.1. Experimental Protocol

Three stimuli protocol was used in this study. Six different jewels and their pictures were served as stimuli during detection. A safe containing one or two jewels was given to each participant. The guilty was instructed to steal only one object from the safe, and it was served as probe (P) stimulus. The other one in the safe was Target (T) stimulus; the remaining four pictures were irrelevant (I) stimuli.

The innocent only memorized the object in the safe that was T stimulus. Then, P stimulus was selected randomly from the remaining five pictures. After that, the remaining four images were I stimuli. After above preparation, each stimulus was randomly displayed at a screen, which remained 0.6 s with 30 iterations for one session. Each session was about 5 minutes with 2 minutes' resting time. The inter-stimulus interval was 1.6 s. Each subject performed 5 sessions. One push button was given to each subject who pressed "Yes" and "No" button facing with familiar and unknown items, respectively. The guilty group pressed "Yes" and "No" button when facing with T and I stimuli, respectively. With a P stimulus, they were asked to press the "No" button, trying to hide the stolen act. The innocent made honest responses to all the stimuli.

2.2. Acquisition for EEG Signals

The experiment was performed on 30 healthy participants (mean age: 21). Pz electrode from an International 10-20 system was used. The vertical EOG (VEOG) signal and the horizontal EOG (HEOG) signals were also recorded. EEG and EOG signals were filtered online with a band pass filter of 0.1-30 Hz, and digitized at 500 Hz using Neuroscan SynAmps. All electrodes were referenced to the double earlobes.

3. Methods

3.1. Feature Extraction Method Based on Wavelet Transform

ERP signal is typical non-stationary signals, which contains several different frequency components. Wavelet transform uses a longer time window to analyze low-frequency signal and uses a shorter time window to analyze high-frequency signal, so wavelet transform also can be called multiscale analysis. Because of these characteristics make WT particularly suitable for the analysis of ERP signals. Firstly by applying wavelet transform we selected 300 ms time-domain waveform before stimulus and the 500ms time-domain after stimulus to decompose 7 Wavelet coefficient sets corresponding to 0.3-3.9, 3.9-7.8, 7.8-15.6, 15.6-31.2, 31.2-62.5, 62.5-120 Hz and 125-250 Hz. Since the frequency range of P300 has been proven mainly in δ band, 22 wavelet coefficients of wavelet features corresponding to 0.3-3.9 Hz were last selected as the wavelet feature, and they, as the initial features, will be further processed by KPCA to select the most optimal feature subsets.

3.2. Kernel Principal Component Analysis

KPCA [10-11] uses the kernel method to generalize linear PCA into nonlinear. The method of

KPCA is to firstly map given data vectors x_t into a high-dimensional feature space $\Phi(x_t)$ and then to calculate the linear PCA in $\Phi(x_t)$. The linear PCA in $\Phi(x_t)$ corresponds to a nonlinear PCA in x_t. By mapping x_t into $\Phi(x_t)$ whose dimension is assumed to be larger than the number of training samples E, KPCA solves the eigenvalue problem:

$$\lambda_i u_i = C u_i, i = 1,..., E, \tag{1}$$

where $C = \frac{1}{E}\sum_{t=1}^{E}\Phi(x_t)\Phi(x_t)^T$ is the sample covariance matrix of $\Phi(x_t)$. λ_i is one of the non-zero eigenvalues of C. u_i is the corresponding eigenvector. Above equation can be transformed to the eigenvalue problem:

$$\tilde{\lambda}_i \partial_i = K\partial_i, \quad i=1, \dots, E, \tag{2}$$

where K is the $E\times E$ kernel matrix, and the value of every element of K is

$$K = \Phi\Phi^T, \tag{3}$$

where $\tilde{\lambda}_i$ is one of the eigenvalues of K and $\tilde{\lambda}_i = N\lambda_i$. ∂_i is the corresponding eigenvector of K;

$$u_i = \sum_{j=1}^{E}\partial_i(j)\Phi(x_j) \; j=1, \dots, E, \tag{4}$$

The centered kernel matrix can be expressed as

$$K_c = (\mathrm{I}-\frac{1}{E} j_E j_E{}')K(\mathrm{I}-\frac{1}{\mathrm{E}} j_E j_E{}') \; j=1, \dots, E, \tag{5}$$

In this paper, we used RBF kernel function $k(i,j) = \exp(-\frac{\| x_i - x_j \|^2}{q})$, and q is a parameter that is related to the variance of the data. Based on our experience and simple trying, q = 0.2 was finally decided and used.

3.3. Extreme Learning Machine

Given N different training samples (x_i, t_i) where $x_i=[x_{i1}, x_{i2},\dots, x_{in}]^T \in R^n$ and $t_i=[t_{i1}, t_{i2},\dots, t_{im}]^T \in R^m$. We train a Single hidden layer feed-forward neural network (SLFN) with S hidden nodes and activation function $g(x)$. This network can be expressed as:

$$\sum_{i=1}^{S}\beta_i g_i(x_j) = \sum_{i=1}^{S}\beta_i g(a_i \times x_j + b_i) = t_j, \quad j=1, \dots, N, \tag{6}$$

where $a_i=[a_{i1}, a_{i2},\dots, a_{in}]^T$ denotes the weight vector connecting i th hidden node and n input nodes; b_i is bias of i^{th} hidden node; $\beta_i=[\beta_{i1}, \beta_{i2},\dots, \beta_{im}]^T$ denotes the weigh vector connecting ith hidden node and the m output nodes.

Above equation can be rewritten in a matrix as

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T}, \tag{7}$$

where

$$\mathbf{H} = \begin{bmatrix} g(\boldsymbol{a}_1\cdot\boldsymbol{x}_1+b_1) & \cdots & g(\boldsymbol{a}_S\cdot\boldsymbol{x}_1+b_S) \\ \vdots & \cdots & \vdots \\ g(\boldsymbol{a}_1\cdot\boldsymbol{x}_N+b_1) & \cdots & g(\boldsymbol{a}_S\cdot\boldsymbol{x}_N+b_S) \end{bmatrix}_{N\times S}$$
$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1^T \\ \vdots \\ \boldsymbol{\beta}_S^T \end{bmatrix}_{S\times m} \quad \mathbf{T} = \begin{bmatrix} \boldsymbol{t}_1^T \\ \vdots \\ \boldsymbol{t}_N^T \end{bmatrix}_{N\times m}, \tag{8}$$

H is called the hidden layer output matrix with the ith column being the output of the ith hidden node (Huang et al., 2000). To learn N instances for a SLFN, one traditionally finds the solution set **W**, including $\boldsymbol{a}_i$, $\boldsymbol{\beta}_i$ and b_i, to minimize following cost function:

$$E(\mathbf{W}) = \sum_{j=1}^{N}(\sum_{i=1}^{S}\boldsymbol{\beta}_i g(\boldsymbol{a}_i\cdot\boldsymbol{x}_j+b_i)-\boldsymbol{t}_j)^2, \tag{9}$$

Given an arbitrarily small value ε>0, Huang et al. proved that if the input weights and biases of hidden nodes are assigned randomly and the activation function in the SLFN is infinitely differentiable, the SLFN can approximate the N training data with ε error, i.e., $\|\mathbf{H}\boldsymbol{\beta}\text{-}\boldsymbol{T}\| \leq \varepsilon$ In this case, matrix **H** has been randomly fixed. Hence, the training procedure of SLFN is equal to seeking a least-squares (LS) solution of the linear system:

$$\left\|\mathbf{H}\hat{\boldsymbol{\beta}} - \mathbf{T}\right\| = \min_{\beta}\left\|\mathbf{H}\boldsymbol{\beta} - \mathbf{T}\right\|, \tag{10}$$

where $\hat{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{T}$ is the LS solution of above seeking problem with the smallest norm, and $\mathbf{H}^{\dagger}$ is the Moore-Penrose generalized inverse of **H**. As Huang et al. [7] pointed out that for SLFNs, the smaller their output weights are, the better generalization ability they have. Thus, ELM algorithm can obtain not only the smallest training error but also the best generalization ability.

3.4. The Proposed Method

The new method which combines ELM algorithm with KPCA can not directly decide the number of feature subset and the number of hidden nodes *(NHN)* of ELM during feature selection process, we use wrapped l [12] feature selection method to train above two parameters, and the training procedure is shown in Fig. 1.

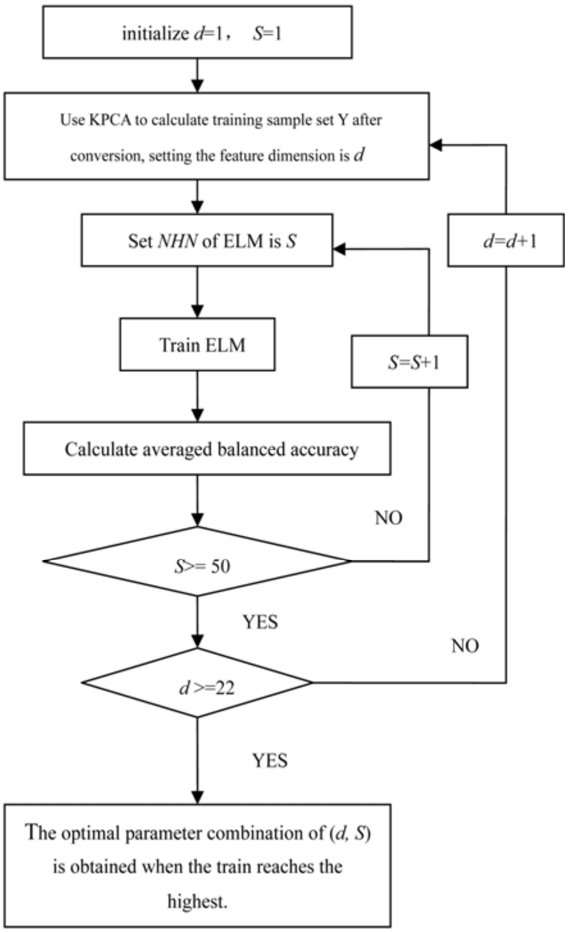


Fig. 1. The training process of the KPCA_ELM for EEG signals.

The proposed method is divided into following steps:

1. Using KPCA algorithm to calculate training samples X of the original feature space, and select the new training sample collection Y from d-dimensional space according to the order of the eigenvalues during the calculation, initialize d=1.

2. The feature vector was fed into an ELM classifier with S hidden nodes by setting Y, initialize S=1. A grid searching program was used to select the most optimal parameter combination *(S, d)*, which was performed on the training data. A 10-fold CV was performed, resulting in 10 pairs of training sets and testing sets. Hence, an averaged balanced accuracy, namely *BA_train*, across the 10 pairs of training sets can be obtained by averaging sensitivity accuracy TR_{sen} and specificity accuracy TR_{spe}, i.e., $BA=(TR_{sen}+TR_{spe})/2$.

3. Update S by S+1. Repeat *step 2-3* until S equals *50*.

4. Update d by d+1. Repeat *step 1-4* until d equals *22*.

5. Comparing all the *BA_train* obtained in step 2, the optimal parameter combination（d, S）is finally obtained when the *BA_train* reached highest, at the same time, the TR_{sen} ±*SD* and TR_{spe} ±*SD* are correspond as an Optimal training accuracy of the network.

By above procedure, the optimal parameters d can be decided after training 10 different test samples, and we can get the optimal testing accuracy TE_{sen} ±*SD* and TE_{spe} ±*SD*.

To evaluate the performance of the proposed method objectively, KPCA_BPNN and KPCA_SVM, were also performed. Like the above steps, each model was trained to obtain the optimal classifier parameters when the training accuracy *BA_train* reached highest.

For the BPNN, learning rate η and control precision ε were set to be 0.002 and 0.001, respectively. The Levenberg-Marquardt algorithm was used; For SVM, the penalty parameter A and radial width σ for radial basis function (RBF) (kernel function $K(x,y)=e^{-1/2*(\|x-y\|/\sigma)^2}$, Burges, 1998) were tuned with the following grid: $A = [2^{-2}, \ldots, 2^{10}]$, $\sigma= [2^{-3}, \ldots, 2^{10}]$ (step size = 2^1).

4. Results

4.1. Preprocessing Results

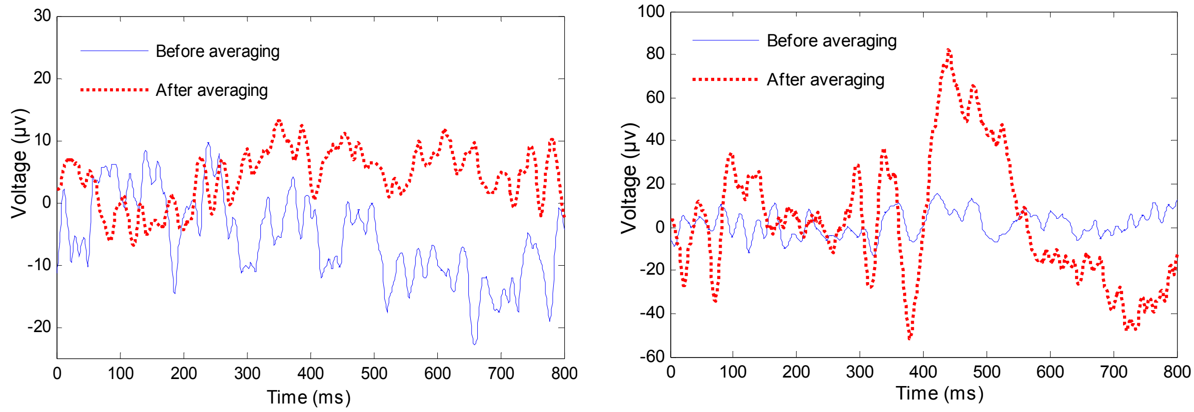
First, a temporally average filter is applied to each five response waveforms for each subject to alleviate impact of mental state fluctuations of subjects and related artifacts. We randomly selected a liar and honest people, and compared waveform before average and after average, the results are shown in Fig. 2. From the Fig. 2, we can see that the peak at 350 ms is more obvious after stimulation in liar comparing to honest people, which is show that time-domain average is good for the latter feature extraction and feature classification.

4.2. Results

The experimental results are shown in Table 1. First, we compared the classification results before and after using KPCA algorithm. From the testing accuracy we can see that the accuracy of training have been improved to some extent after using KPCA algorithm, such as the sensitivity accuracy and specificity accuracy of KPCA_SVM were 95.60 % and 96.34 %, but SVM were 94.22 % and 92.14 %, thus KPCA algorithm is favorable for lie detection.

From this table, it can be concluded that the training accuracy of KPCA_ELM is highest (97.78 % and 86.83 %, respectively) in the 6 classification models, and the value of parameter d has been reduced to 7 from the initial number of features after using KPCA algorithm.

Secondly, observing the training and testing time in this table, we can see that ELM cost the minimum training time compared with those in the other models.



(a) (b)

Fig. 2. The contrast of waveform before and after averaging. Fig. 2a and 2b are the results of liar and honest subjects, respectively.

Table 1. The classification of the various classification models.

Models	***S***	***d***	**Training and testing accuracy (%)**				**Time**	
			$TR_{sen} \pm SD$	$TR_{spe} \pm SD$	$TE_{sen} \pm SD$	$TE_{spe} \pm SD$	**Training (hour)**	**Test (second)**
BPNN	31	22	90.21±2.23	82.36±4.55	89.56±1.38	88.27±2.59	1.09	0.89
SVM	54	22	94.22±0.88	92.14±0.98	94.84±2.38	92.61±3.05	17.73	14.02
ELM	40	22	95.89±0.71	95.71±0.76	92.70±0.51	95.30±0.72	0.03	0.008
KPCA_BPNN	20	9	93.53±1.28	92.60±1.85	93.58±2.31	93.53±2.34	88.05	1.22
KPCA_SVM	43	9	95.60±0.27	96.34±0.46	95.38±0.24	94.85±0.29	648.82	21.2
KPCA_ELM	21	7	97.78±0.36	96.83±0.39	97.44±0.58	96.83±0.45	0.33	0.01

5. Discussions and Conclusion

In this paper, EEG signals were applied to detect liars and we proposed a novel method KPCA_ELM to enhance the classification accuracy. This method used KPCA to select the most optimal feature subsets. Meanwhile, the performance of the proposed method was evaluated objectively. The other two classification algorithm, BPNN and SVM, were also performed in this paper.

The experimental results show that the training accuracy and the testing accuracy are higher than other machine learning algorithms by using ELM algorithm. This method speeded up the testing speed by reducing *NHN*. These results had shown that feature selection was very important for the EEG signal processing.

ELM had some good properties compared with other machine learning algorithms. First, there was almost no classifier parameters needed to be turned in its learning procedure compared with BPNN and SVM, which greatly reduced the computational cost and training complexity. Second, the learning speed is also slow when a large number of training data need to learn and hence too slow to meet the requirements of real-time applications. ELM could overcome this shortcoming effectively. Third, artificial neural networks such as BPNN are proved to fall into a local minimum. In contrast, ELM has a surprising learning speed because the output weights are determined randomly instead of by an iterative strategy. It is worthwhile to note that the ELM classification is more suitable than the other classifiers to combine with feature selection, because ELM classifier only have few parameters to affect the classification results, thus it is better to use Wrapped synchronization training mechanisms.

The experimental results also suggest that there's a connection between *NHN* and feature points of input samples. Furthermore, training process of KPCA_ELM can run in parallel to improve training efficiency. At last, there are many other feature selection algorithm can be used to combine with ELM, including F-score, mutual information (MI), genetic algorithm (GA). These methods are our future research work.

Acknowledgements

The work was supported by National Nature Science Foundation of China (No. 81271659), and by Academic Team of South Central University for Nationalities: Biomedical Signals Processing (No. XTZ09002).

References

[1]. J. Polich, P300 in clinical applications: Meaning, method, and measurement, *American Journal of EEG Technology*, Vol. 31, No. 3, 1991, pp. 201-231.

[2]. J. P. Rosenfeld, B. Cantwell and V. T. Nasman, A modified event-related potential-based guilty knowledge test, *International Journal of Neuroscience*, Vol. 42, No. 1, 1988, pp. 157-161.

[3]. L. A. Farwell and E. Donchin, The truth will out: interrogative polygraphy ("lie detection") with event-related potentials, *Psychophysiology*, Vol. 28, No. 5, 1991, pp. 531-547.

[4]. C. Davatzikos, K. Ruparel and Y. Fan, Classifying spatial patterns of brain activity with machine learning methods: application to lie detection, *NeuroImage*, Vol. 28, No. 3, 2005, pp. 663-668.

[5]. Junfeng Gao, Xiangguo Yan, Jiancheng Sun, Denoised P300 and machine learning-based concealed information test method, *Computer Methods and Programs in Biomedicine*, Vol. 104, No. 3, 2011, pp. 410-417.

[6]. Junfeng Gao, Liang. Lu, Yong. Yang, A Novel Concealed Information Test Method Based on Independent Component Analysis and Support Vector Machine, *Clinical EEG and Neuroscience*, Vol. 43, No. 1, 2012, pp. 54-63.

[7]. G. B. Huang, Q. Y. Zhu, C. K. Siew, Extreme learning machine: theory and applications, *Neurocomputing*, Vol. 70, No. 1, 2006, pp. 489-501.

[8]. G. B. Huang, L. Chen, Enhanced random search based incremental extreme learning machine, *Neurocomputing*, Vol. 71, No. 16, 2008, pp. 3460-3468.

[9]. Y. Miche, P. Bas and C. Jutten, A methodology for building regression models using extreme learning machine: OP-ELM, in *Proceedings of the 16th European Symposium on Artificial Neural Networks*, Bruges, Belgium, 2008, pp. 1-6.

[10]. A. R. Teixeira, A. M. Tomé and K. Stadlthanner, KPCA denoising and the pre-image problem revisited, *Digital Signal Processing*, Vol. 18, No. 4, 2008, pp. 568-580.

[11]. Liu Jianping, Chong Zhang, and Chongxun Zheng. EEG-based estimation of mental fatigue by using KPCA–HMM and complexity parameters. *Biomedical Signal Processing and Control*, Vol. 5, No. 2, 2010, pp. 124-130.

[12]. L. Talavera, An evaluation of filter and wrapper methods for feature selection in categorical clustering, in *Proceeding of the 6th International Symposium on Intelligent Data Analysis. Madrid: Springer*, 2005, pp. 440-451.