Deep Belief Networks for EEG-Based Concealed Information Test

Qi Liu[1(✉)], Xiao-Guang Zhao[1], Zeng-Guang Hou[1],
and Hong-Guang Liu[2]

[1] The State Key Laboratory of Management and Control for Complex Systems,
Institute of Automation, University of Chinese Academy of Sciences,
Beijing 100190, China
popofayl7@gmail.com
[2] Institute of Crime, Chinese People's Public Security University,
Beijing 100038, China

Abstract. This paper introduces a deep learning approach to the feature extraction of P300 cognitive component existing in electroencephalogram signals collected in an autobiographical paradigm test. A thorough belief mechanism is used for the extraction of deep characteristics rather than raw feature vectors to train the classifier. It is shown that the classification accuracy is satisfactory by learning deep from the experimental data. Experiments have validated the usefulness of the algorithm. The hidden information has been obtained accurately with a single electroencephalogram channel. Moreover, performances of support vector machine with different feature extraction methods are compared.

Keywords: Electroencephalogram · Concealed information test · Deep feature extraction · Deep belief networks

1 Introduction

In recent years, EEG-based concealed information test has drawn considerable attention in the field of criminal investigation. Many effective methods have been used for EEG signal analysis in Concealed Information Test (CIT) [1]. Compared to traditional methods based on physiological responses which are easily affected by emotions and stress, cognitive behavior based polygraph is considered more reliable and scientific that can reduce the risk in false positive errors [2]. In addition, EEG is more convenient, more harmless and more economical than other brain activity monitoring methods such as PET, MEG and fMRI [3].

Due to the complexity and particularity of actual criminal investigation tasks and poor ratio of signal intensity to noise intensity (SNR), increasing performance of recognition of raw EEG signals remains a live problem. In which, methods based on machine learning algorithms have achieved the most effective results. Numerous feature extraction approaches have been adopted in machine learning algorithms such as time or periodicity methods [4], model parameter methods [5], as well as methods on the basis of wavelet decomposition [6], etc. [7]. However, the distinguishability of a

certain feature is uncertain in different tasks, which may lead to a failure of recognition. Therefore, feature extraction methods which are capable of feature self-learning are necessary to be studied in this field. Recently, deep learning strategy has made great progress and the related algorithms have also been adopted in various fields such as EEG signal processing [8]. It can be viewed as a computational intelligent method since its similar mechanism to human brain. To improve the generalization performance of EEG feature, deep belief networks (DBN) is adopted to learn features automatically.

In this paper, we use the CIT technique and focus primarily on the feature extraction process of different brain waves evoked by relevant stimulus and control stimulus. DBN was applied to self-learn features of EEG signals. Then support vector machine (SVM) was implemented as the classifier. The classification performance is satisfactory and the runtime is acceptable.

2 Methods

2.1 Data Description

Data in this paper were from an autobiographical paradigm test [9]. There were 11 volunteering subjects in total participated who were all males at the age of between 22 and 35. They are all used to using the right hand and their vision are all normal or corrected to normal range. They have no idea what the test is based on and just know how to carry out the test. All the subjects were required to offer five numbers which all contained 4 digits and one of the numbers was the year of birth. The experimenter was not informed by the subjects of the birth date number until when the experiment ended. In the experiment, subject 11 took part in 3 runs while other subjects were involved in 2 runs. Due to wrong target stimulus counting (as was shown below), subject 1, 3, and 7 saw one of their runs vetoed. Finally, the study applied a total of 20 runs in the experiment. To achieve whole stimuli, in each run, the subject was exposed to each number with random for thirty times. Each number was revealed for one second and there was a two-second blank in the screen between the numbers. The experiment required the subjects to count how many times the number of the birth year was revealed instead of responding to the items. The subjects did not know that the entire target stimulus were displayed with 30 repetitions. EEG signals sampled at 256 Hz digitally were recorded at the Fz, Cz and Pz electrode positions of the 10–20 international electrode placement system (Fig. 1). The electrodes referred to linked mastoids. For the purpose of blink artifact detection, the experiment also recorded vertical EOG signals.

2.2 Methods

For the complexity and weak anti-interference capability of EEG, it is not easy to recognize effective data from raw signals. Figure 2 shows the raw waveforms. It is observed that the potential offset value of each sample belonging to the same category is quite different and there is no obvious distinction between samples belonging to separate categories.
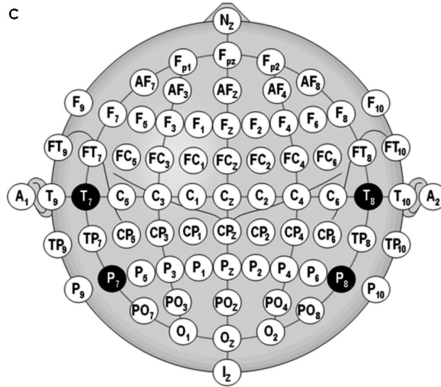
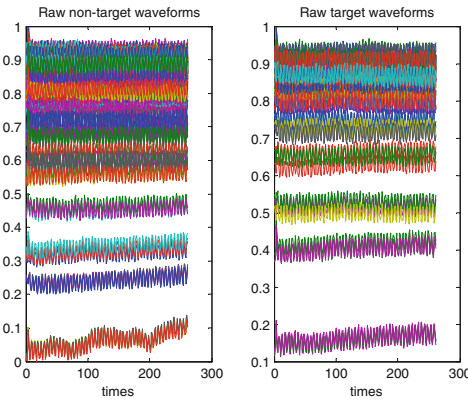**Fig. 1.** 10–20 system of electrode placement

Fig. 2. Raw EEG waveforms

Figure 3 shows that the signal process mainly includes data collection, pre-processing, feature exaction and pattern classification.

(1) **Pre-processing**

This process consists of electrodes selection, segmentation of signals, superposition and filtering. For low SNR of EEG signals, the stimulations are repeated to remove unnecessary signals and enhance useful signals. Because the P300 frequency is primarily allocated in area with low frequency, the experiment designed a 6-order band pass Chebyshev Type I filter with cut-off frequencies 0.5 and 35 Hz to penetrate each epoch. Moreover, the data information matrix is designed into a range from 0 to 1 according to Eq. (1).

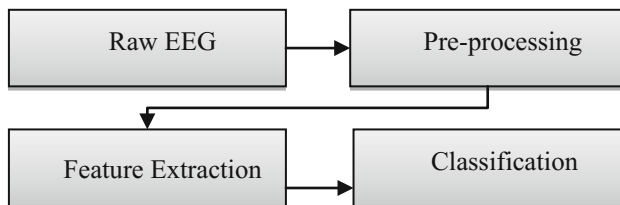$$\mathbf{x}_{norm} = \mathbf{x} - \mathbf{x}_{min}/\mathbf{x}_{max} - \mathbf{x}_{min} \tag{1}$$
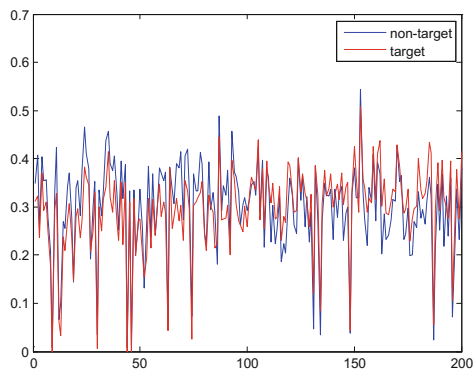
Fig. 3. The flowchart of signal processing

Fig. 4. Comparison of mean values of the two categories

(2) Deep Feature Extraction

To begin with, k-means method is adopted to represent features preliminary as described in [11]. Using subject 1 as an example, some differences between the two categories can be seen in Fig. 4 after the initial feature extraction. However, the difference is still too small to distinguish samples. Further feature extraction is implemented as following.

DBN could be considered to be a stack of RBMs (Restricted Boltz-man Machines), which are motivated from the idea of equilibrium from the statistical physics literature [12]:

$$E(\mathbf{v}, \mathbf{h}; \boldsymbol{\theta}) = - \sum_j a_j v_j - \sum_i b_i h_i - \sum_{i,j} v_j h_i w_{ij} \qquad (2)$$

Where w_{ij} is the symmetric interaction term between distinct unit v_j and covered unit h_i, a_i as well as b_j are both the bias term. $\boldsymbol{\theta} = \{\mathbf{w}, \mathbf{a}, \mathbf{b}\}$ is the model parameter need to be learned.

Equation (2) could be optimized in a tricky way by contrastive divergence that has been usually applied to border on the expectation by a sample deriving from a certain amount of Gibbs sampling iterations [13].

When defined on a probability space, the joint distribution over $\mathbf{v}$ and $\mathbf{h}$ is:

$$P(\mathbf{v}, \mathbf{h}) = \frac{1}{z} e^{-E(\mathbf{v}, \mathbf{h})} \tag{3}$$

where z is a standardized factor. Then

$$P(\mathbf{v}) = \sum_{\mathbf{h}} P(\mathbf{v}, \mathbf{h}) = \frac{e^{-F(\mathbf{v})}}{z} \tag{4}$$

in which

$$F(\mathbf{v}) = -\log \sum_{\mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})} \tag{5}$$

Model (2) can be simplified by using binary input variables. The conditional probabilities can be formulated as:

$$\begin{aligned} P(h_i = 1|\mathbf{v}) &= sigm(b_i + w_i \mathbf{v}) \\ P(v_j = 1|\mathbf{h}) &= sigm\left(a_j + w_j'\mathbf{h}\right) \end{aligned} \tag{6}$$

Then

$$F(\mathbf{v}) = -\mathbf{a}'\mathbf{v} - \sum_i \log\left(1 + e^{(c_i + w_i \mathbf{v})}\right) \tag{7}$$

$$-\frac{\partial \log P(\mathbf{v})}{\partial \theta} = \frac{\partial F(\mathbf{v})}{\partial \theta} - \sum_{\tilde{\mathbf{v}}} P(\tilde{\mathbf{v}}) \frac{\partial F(\tilde{\mathbf{v}})}{\partial \theta} \tag{8}$$

To make RBM stability, the energy of system should be the minimum. By the above formulas, $P(\mathbf{v})$ should be maximized. The partial derivative of loss function $-P(\mathbf{v})$ is calculated as:

$$\begin{aligned} -\frac{\partial \log P(\mathbf{v})}{\partial w_{ij}} &= E_{\mathbf{v}}\left[P(h_i|\mathbf{v}) \cdot v_j\right] \\ &\quad - v_j^{(i)} \cdot sigm\left(w_i \cdot v^{(i)} + c_i\right) \\ -\frac{\partial \log P(\mathbf{v})}{\partial c_i} &= E_{\mathbf{v}}[P(h_i|\mathbf{v})] - sigm\left(w_i \cdot v^{(i)}\right) \\ -\frac{\partial \log P(\mathbf{v})}{\partial b_j} &= E_{\mathbf{v}}\left[P(v_j|\mathbf{h})\right] - v_j^{(i)} \end{aligned} \tag{9}$$

Thus, the parameter θ corresponding to maximum $P(\mathbf{v})$ is obtained. DBN could then be trained with the greedy layer-wise method [12]. Each RBM is trained greedily and unsupervised [14]. The posterior distribution of the first RBM is used as the input

distribution of the second RBM. Then the weights are fine-tuned by back propagation (BP) neural network. Figure 5 shows the architecture of DBN model and Fig. 6 displays the comparison of mean values between two categories. The difference is significant after feature learning by DBN.
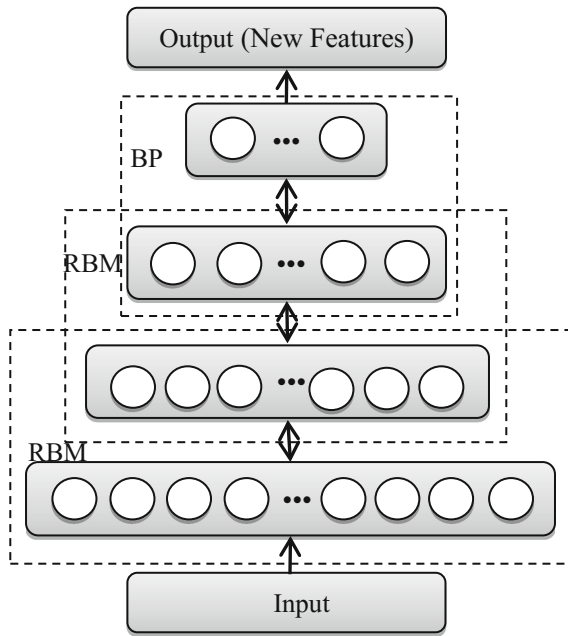

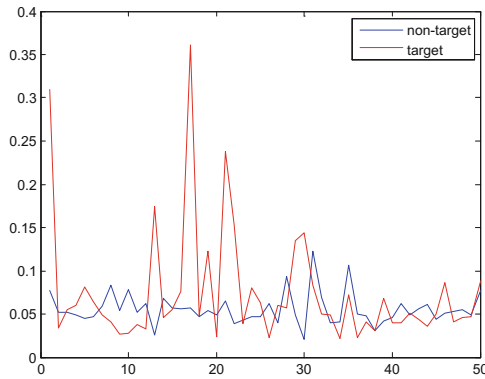**Fig. 5.** The architecture of DBN model

Fig. 6. Comparison of mean values of the two categories

(3) **Classification**

DBN model is viewed as a feature extraction system in this paper. Outputs of the last model were used as the new input feature vectors with labels of samples to train the SVM classifier.

3   Experiments

Responses to the birth year of the subject are expected to contain the P300 component, which is a late positive component, which is considered as the most typical and common event-related potential (ERP) closely related to human cognitive process. For the time-locked conception between the stimulus and the response [15], the value of the signal during 0–700 ms was set after stimulus onset. The experiment randomly assigned the weights with an initial value and the turning parameters were set as: learning rate = 0.07, momentum = 0.95. For the first RBM, the visible unit is set at 200 and the hidden unit is 100. For the second RBM, the number of the visible units is 100 and that of the hidden units is 50. The fifty-dimensional feature vector is input to libsvm.

To ensure the accuracy of training as well as testing data, a 10-fold cross-validation method was employed. According to this technique, the dataset was divided into ten subsets [16]. To improve the dependability, the 10-fold cross-validation procedure was performed with ten repetitions. And in each time, only one subset was used as the testing dataset and the other 9 ones were collected to constitute the training dataset. Particularly, data from test fold is not be involved in the optimization procedure. All final data were calculated by averaging the ten results.

4   Results and Discussion

This section made a test of performance of the DBN-SVM classification algorithm on the basis of the dataset presented in Sect. 2.1. Table 1 and Fig. 7 reveal the results. Specifically, Table 1 displays the recognition accuracy and runtime over all eleven subjects. Figure 7 compares performances of classifiers adopted different effective feature extraction methods for SVM classifier. All the experiments are repeated ten times, and the average results are reported.

From the effects of perspective, a high average accuracy is obtained. In addition, as shown in Fig. 7, compared with other features used methods, the performance of our approach is significantly better.

Moreover, it is worth noticing that it does not require pre-processing operations including artifact removal or bootstrapping which takes much time and allows the approach possible to be applied to actual tasks.

However, the complex application environment and unpredictable interference will definitely put forward higher requirements considering the practical applications in crime information identification tasks. As for future works, it would be interesting to investigate a way to overall fine-tune the weights of DBN model with regard to SVM learning rule [13].

Table 1. Performances of the algorithm over all subjects

Subject	Amount of samples	Accuracy (%)
S1	150	95.5
S2	300	98.9
S3	300	97.6
S4	150	96.7
S5	150	97.5
S6	300	98.0
S7	300	97.0
S8	150	96.3
S9	300	97.6
S10	300	96.2
S11	450	98.9
Average		97.3
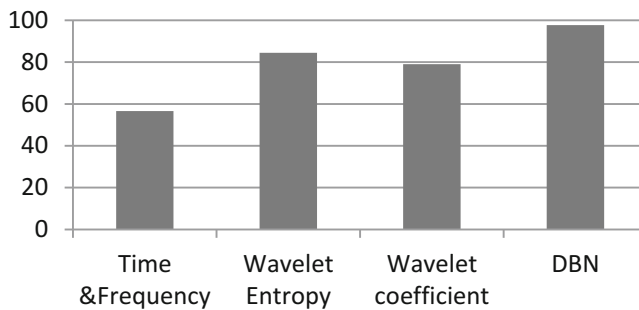
Fig. 7. Comparison of classification performances over different feature extraction methods

5 Conclusion

In this paper, deep learning strategy is applied for signal processing in concealed information test based on EEG. The introduction of DBN aims to better express characteristics of different signals. We choose SVM as the classifier which can avoid over-fitting effectively. According to the results, the method has been highly recognized. The study in this paper suggests that it is valuable to do further development on deep learning or other computational intelligence strategies applied in CIT based on EEG as well as provide reliable supports to actual future explorations.

References

1. Abootalebi, V., Moradi, M.H., Khalilzadeh, M.A.: A comparison of methods for ERP assessment in a P300-based GKT. Int. J. Psychophysiol. **62**(2), 309–320 (2006)
2. Zhao, M., Zhang, C., Zhao, C.: New approach for concealed information identification based on ERP assessment. J. Med. Syst. **36**(4), 2401–2409 (2011)
3. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfortscheller, G., Vaughan, T.M.: Brain-computer interfaces for communication and control. Clin. Neurophysiol. **113**(6), 767–791 (2002)
4. Guo, X.J., Wu, X.P., Zhang, D.J.: Motor imagery EEG detection by empirical mode decomposition. In: International Joint Conference on Neural Networks, pp. 2619–2622 (2008)
5. Zhao, M.Y., Zhou, M.T., Zhu, Q.X.: Feature extraction and parameters selection of classification model on Brain-computer interface. In: IEEE 7th International Symposium on Bioinformatics and Bioengineering, pp. 1249–1253 (2007)
6. Sherwood, J., Derakhshani, R.: On classifiability of wavelet features for EEG-based Brain-computer interfaces. In: International Joint Conference on Neural Networks, pp. 2895–2902 (2009)
7. Subha, D.P., Joseph, P.K., Acharya, U.R., Lim, C.M.: EEG signal analysis: a survey. J. Med. Syst. **34**(2), 195–212 (2010)
8. Jirayucharoensak, S., Pan-Ngum, S., Israsena, P.: EEG-based emotion recognition using deep learning network with principal component based covariate shift adaptation. Sci. World J. **2014**, 1–10 (2014)
9. Abootalebi, V., Moradi, M.H., Khalilzadeh, M.A.: Detection of the cognitive components of brain potentials using wavelet coefficients. Iran. J. Biomed. Eng. **1**(1), 25–46 (2004)
10. Wang, D., Miao, D., Blohm, G.: A new method for EEG-based concealed information test. IEEE Trans. Inf. Forensics Secur. **8**(3), 520–527 (2013)
11. Coates, A., Ng, A.Y.: Learning feature representations with k-means. In: Montavon, G., Orr, G.B., Müller, K.-R. (eds.) Neural Networks: Tricks of the Trade. LNCS, vol. 7700, pp. 561–580. Springer, Heidelberg (2012). doi:10.1007/978-3-642-35289-8_30
12. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural network. Science **313**(5786), 504–507 (2006)
13. Hinton, G.E.: Training products of experts by minimizing contrastive divergence. Neural Comput. **14**(8), 1771–1800 (2002)
14. Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H.: Greedy layer-wise training of deep networks. Adv. Neural. Inf. Process. Syst. **19**, 146–153 (2007)
15. Quiroga, R.Q.: Quantitative analysis of EEG signals: time-frequency methods and chaos theory. Institute of Physiology-Medical University Lubeck and Institute of Signal Processing-Medical University Lubeck (1998)
16. Ripley, B.D.: Pattern Recognition and Neural Networks. Cambridge University Press, Cambridge (1996)