Network Boosting for BCI Applications

Shijun Wang, Zhonglin Lin, and Changshui Zhang

Department of Automation, Tsinghua University, Beijing 100084, China
{wsj02, linzl02}@mails.tsinghua.edu.cn, zcs@mail.tsinghua.edu.cn

Abstract. Network Boosting is an ensemble learning method which combines learners together based on a network and can learn the target hypothesis asymptotically. We apply the approach to analyze data from the P300 speller paradigm. The result on the Data set II of BCI (Brain-computer interface) competition III shows that Network Boosting achieves higher classification accuracy than logistic regression, SVM, Bagging and AdaBoost.

BCI (Brain-computer interface) is a direct cybernetic link between a mind and a computer which does not depend on the brain's normal output pathways of peripheral nerves and muscles [1]. Most BCIs make use of mental tasks that lead to distinguishable *electroencephalogram* (EEG) signals of two or more classes. P300 potentials provide a means of detecting user's intentions concerning the choice of objects within a visual field. Farwell and Donchin [2] first introduced P300 potentials into BCI, and proposed P300 speller paradigm [3]. In P300 speller paradigm the key task is to detect presence or absence of the P300 component in noisy EEG signals accurately and fast.

Network Boosting (NB) [4] is a new ensemble learning method which combines classifiers on the basis of a network. Theoretic analysis based on the game theory shows that the algorithm can learn the target hypothesis asymptotically. NB is more suitable than other ensemble learning methods for noisy data and distributed application. In this paper We utilize NB for classifying *electroencephalogram* (EEG)-signals to detect absence or presence of the P300 component in EEG event related potentials.

The basic idea of NB is that through the cooperation between classifiers, we expect the learned classifier ensemble has high accuracy as well as high resistance to the noise. The idea comes from our recent research [5] on complex network [6]. In order to facilitate the cooperation of classifiers, a network topology is introduced which serves as a communication structure between them.

Fig. 1 shows the NB algorithm. Assume there are K nodes in the network and the training round is T. In the learning phase, given training set $Z = \langle (x_1, y_1), (x_2, y_2), ..., (x_l, y_l) \rangle$, each classifier on the classifier network is provided with the same training instances and maintains a weight record $w_{k,t}(i)$ for $k = 1, ..., K$, $t = 1, ..., T$, $i = 1, ..., l$ of the instances respectively. Then the classifier in the classifier network is built by the training set sampled from the training data according to the weights record. After that, the weights of the instances of

A. Hoffmann, H. Motoda, and T. Scheffer (Eds.): DS 2005, LNAI 3735, pp. 386–388, 2005.

every node are updated according to the classification results of the node and its predecessors. The classifier network is trained T rounds in such way.

We compare 5 algorithms on data set II of BCI Competition III [7]. In P300 speller paradigm, the user faces a 6×6 matrix of letters. The user's task is to focus attention on characters in a word that was prescribed by the investigator (i.e., one letter at a time). All rows and columns of this matrix are successively and randomly intensified at a rate of 5.7Hz. Two out of 12 intensifications of rows or columns contains the desired letter (i.e., one particular row and one particular column). The responses evoked by these infrequent stimuli (i.e., the 2 out of 12 stimuli that did contain the desired letter) are different from those evoked by the stimuli that do not contain the desired letter.

Algorithm Network Boosting

Input: Examples $Z = \langle (x_1, y_1), (x_2, y_2), ..., (x_l, y_l) \rangle$
Directed Network N
Training rounds T
Sampling parameter ρ
Weight update parameter β

Initialize: $w_{k,1}(x_i) = 1$ for all sample $i = 1, ..., l$ and node $k = 1, ..., K$

Do for: 1. Generate a replicate training set $T_{k,t}$ of size $l\rho$, by weighted sub-sampling with replacement from training set Z for $k = 1, 2, ..., K$.
2. Train the classifier (node) C_k in the classifier network with respect to the weighted training set $T_{k,t}$ and obtain hypothesis $h_{k,t} : x \mapsto \{-1, +1\}$ for $k = 1, ..., K$.
3. Update the weight of instance i of node k :

$$w_{k,t+1}(i) = w_{k,t}(i)\,\beta^{I\left(h_{k,t}(x_i)=y_i\right)+\sum_n I\left(h_{n,t}(x_i)=y_i\right)} / Z_{k,t}, \qquad (1)$$

where node n is predecessor of node i. I is indication function and $Z_{k,t}$ is a normalization constant, such that $\sum_{i=1}^{l} w_{k,t+1}(x_i) = 1$.

Output: Final hypothesis by majority voting using the learned hypotheses $h_{k,t} : x \mapsto \{-1, +1\}$ for $k = 1, ..., K$ and $t = 1, ..., T$.

Fig. 1. Algorithm Network Boosting

The data set comes from two subjects' experiments. For evaluation, we only use the labeled training set. In preprocessing, we find that epoch 11, 62 and 63 of subject A have much larger amplitude than others and we treat them as outliers and discard them. So there are 82 epochs from subject A and 85 epochs from subject B. Then the preprocessing are performed as following: All data are band-pass filtered between 0.5-15Hz; The No. 34, 11, 51, 62, 9, 13, 49, 53, 56 and 60 channels are selected; Signals (lasting 900ms from stimulus) from above channels are concatenated, and then down-sampled to 1/8.

All the algorithms are used to classify single-trial signal. If a signal is judged to have P300 potential, the corresponding code's score is incremented. After 15 classifications for each code, the two codes which gain the highest score gives the target character. We divide epochs as 4 folds, taking 3 as training set and the remaining one test, for subject A and B respectively. After 4 repetitions, we predict all characters. Table.1 gives the error rates by algorithms: Logistic Regression (LR), Support Vector Maching (SVM), Bagging, AdaBoost and NB. Logistic Regression as base classifier is used in all the ensemble learning methods (SVM, Logistic Regression, Bagging and AdaBoost are implemented using WEKA [8].) For Bagging and AdaBoost, 100 base classifiers were used. For NB, $NB(100, 10, 1/3, 0.7)$ and directed random network with connection probability 0.03 (for each directed link) is used.

Table 1. Comparisons of 5 methods on Data set II of BCI competition III

Name	Logistic Regression	SVM	Bagging	AdaBoost	Network Boosting
Subject A (82)	10.98%	10.98%	14.63%	9.76%	6.10%
Subject B (85)	7.06%	17.65%	11.76%	10.59%	5.88%

The comparison results show that NB achieves higher accuracy than others and is more robust than other ensemble learning methods. In the present work, all the data from different channel are combined together as the training data of classifiers. What if we apply one classifier for one channel in the NB algorithm and combine the final results together? It may be a way of future research.

References

[1] Wolpaw, J., et al.: Brain-computer interfaces for communication and control. Clin. Neurophysiol., vol 113. (2002) 767-791

[2] Farwell, L.A., and Donchin, E.: Talking off the top of your head: toward a mental prothesis utilizing event-related brain potentials. Electroenceph. Clin. Neurophysiol, vol. 70. (1988) 510-523

[3] Donchin, E., et al.: The mental prosthesis: assessing the speed of a P300-based brain-computer interface. IEEE Trans. Rehabi. Eng., vol. 8. (2000) 174-179

[4] Wang, S.J., Zhang, C.S.: Network game and boosting. In The 16th European Conference on Machine Learning, (2005)

[5] Wang, S.J., Zhang, C.S: Weighted competition scale-free network. Phys. Rev. E **70**, 066127 (2004)

[6] Albert, R. and Barabási, A.: Statistical mechanics of complex networks. Reviews of Modern Physics, 74 (1). (2002) 47-97

[7] http://ida.first.fraunhofer.de/projects/bci/competition_iii/

[8] Witten, I.H. and Frank, E.: Data Mining: Practical machine learning tools with Java implementations, Morgan Kaufmann, San Francisco, (2000)