APSIPA ASC 2011 Xi'an

Self-training Algorithm for Channel Selection in P300-Based BCI Speller

Jinyi Long and Zhenghui Gu and Yuanqing Li* and Tianyou Yu
College of Automation Science and Engineering, South China University of Technology, 510640
* corresponding author, E-mail: auyqli@scut.edu.cn

***Abstract*—In this paper, we address the important problem of channel selection for a P300-based brain computer interface (BCI) speller system in the situation of insufficient training data with labels. An iterative semi-supervised support vector machine (SVM) is proposed for time segment selection as well as classification, in which both labeled training data and unlabeled test data are utilized. The performance of our algorithm has been evaluated through the analysis of a P300 dataset provided by BCI Competition 2005. The results show that our algorithm for channel selection and classification achieves satisfactory performance, meanwhile it can significantly reduce the training effort of the system[1].**

I. Introduction

A brain computer interface (BCI) is a direct pathway between a brain and an external device for the purpose of communication and control, particularly for the paralyzed people who suffer severe neuromuscular disorders, through exploiting the brain signals such as non-invasive electroencephalogram (EEG) or invasive neural spikes [1], [2]. Typically, P300 ERP is an evoked potential of the brain to some specific external stimulus including auditory, visual, or somatosensory stimuli in a stream of frequent stimuli [3]. P300-based BCI has been implemented to help disabled to communicate with computers through virtual keyboard [4], [5], and the whole system is called a P300 speller.

Although P300 can be detected at distributed sites of scalp, it has a dominant parietal topography. Hence, most of the existing P300 based BCI research has focused on the EEG signals from a few standard P300 scalp locations (e.g., Fz, Cz, Pz) [6]. However, as a matter of fact, there exists significant spatial discrepancy of P300 on scalp among individuals. Therefore the fixed sets of standard P300 channels cannot meet the needs of building BCIs with high performance for all the subjects. From the data analysis reports of BCI Competition 2005 [7], personalized automatic channel selection plays an important role in the overall performance of the P300 BCI speller. As a typical example, in [8], recursive channel elimination based on discriminative score has been used for channel selection. By eliminating four least important channels in each iteration, the classification performance of a large validation data set is adopted as a fitness index to select the significant subset of channels. Channel selection and classification regarding a P300 BCI dataset from BCI Competition 2003 has also been implemented by a genetic algorithm in [9]. When sufficient training data are available, the aforementioned channel selection and classification approaches can achieve outstanding performance.

The state-of-art method, stepwise linear discriminant analysis (SWLDA) has been widely used in recent work to select the features for P300 speller [5]. However, this method is a supervised method without using the information of the test dataset, which may be not stable over time especially when the training dataset is small. For the purpose of improving the performance of P300 speller, we consider to use a semi-supervised approach in this paper for channel selection. Along with the feature selection, classification is also performed.

In this paper, we extend the algorithm in [10] for channel selection in P300 speller where labeled data are insufficient. Our proposed algorithms make use of both training data with labels and test data without labels. The statistical distance of two classes is measured by Fisher ratio, the computation of which also involves both labeled and unlabeled data. Our data analysis results from a P300 dataset provided by BCI Competition 2005 validate the effectiveness of the proposed algorithm and the benefit brought by using unlabeled data.

II. Self-training algorithms for channel selection

In P300-based BCIs, the detection of P300 in a segment of EEG signal can be well described by a two class problem. The signal containing P300 is labeled by 1, and the signal without P300 is labeled by -1. In this section, we first define a Fisher ratio based on SVM score to measure the statistical distance of feature vector ensembles from the two classes. Then we present the details of self-training semi-supervised SVM algorithms for channel selection.

A. Fisher Ratio Based on SVM Score

In this paper, we use SVM as a classifier. Given the N_c epochs of training data set $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_{N_c}, y_{N_c})\}$, where $\mathbf{x}_i \in \mathcal{R}^m$ is an m-dimensional feature vector and $y_i \in \{-1, 1\}$ is the label indicating the class that $\mathbf{x}_i$ belongs to. A standard SVM for two-class problem can be defined as [11]

[1]This work was supported by National Natural Science Foundation of China under grants 60825306 and 60802068, Natural Science Foundation of Guangdong Province, China under grant 9251064101000012, Science and Technology Programme Foundation of Guangdong Province, China under grant 2009B080701053, National 973 Program of China under grant 2007CB311001), and Fundamental Research Funds for the Central Universities, SCUT under grants 2009ZZ0055 and 2009ZZ0059.

$$\min \frac{1}{2}\|\mathbf{w}\| + C\sum_{i=1}^{N_c}\xi_i$$
$$s.t.\ y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0, i = 1\dots N_c, \tag{1}$$

where $\mathbf{w}$ denotes the weight vector of the classifier and ξ_i denotes the ith slack variable; $\|\cdot\|$ indicates L2-norm operation; and the parameter $C > 0$ controls the tradeoff between the slack variable penalty and the margin. The training of the SVM classifier finds a suitable weight vector $\mathbf{w}$ and new data point $\mathbf{x}$ is classified according to the sign of $d(\mathbf{x})$ given by

$$d(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b, \tag{2}$$

where $d(\mathbf{x})$ is designated SVM score. SVM score is proportional to the distance between the decision boundary and the data point $\mathbf{x}$. In this paper, we define a Fisher ratio based on SVM score for feature selection as well as classification of P300-based EEG signal.

Generally, Fisher ratio describes the discriminability of data points from two classes. It is defined as the ratio of the interclass difference to the intraclass spread [12] and has been successfully used as an index for feature selection of a motor imagery BCI [13]. Herein, Fisher ratio based on SVM score is defined as follows to measure the statistical distance of two classes of feature vectors. One class refers to P300, the other one refers to background.

$$FR = \frac{(mean(d_i, i \in Cl_1) - mean(d_i, i \in Cl_2))^2}{(std(d_i, i \in Cl_1))^2 + (std(d_i, i \in Cl_2))^2}, \tag{3}$$

where d_i denotes SVM score of the ith data point; Cl_1 and Cl_2 denote the two classes of epoches with labels being $+1$ and -1 respectively; $mean(\cdot)$ and $std(\cdot)$ represent mean and standard deviation operations respectively.

In the case of insufficient training data with labels, however, the SVM model is generally not reliable, and therefore the resultant Fisher ratio calculated from SVM score is subject to bias. We try to solve this problem through semi-supervised learning where unlabeled data points are also utilized together with labeled data.

B. Self-Training Algorithm for channel Selection

Assume the availability of an insufficient training data set with labels and a large test data set without labels. Based on these data, we present a self-training SVM algorithm for channel selection and signal classification for P300-based BCI, where both the Fisher ratio and the SVM classifier are iteratively updated until the algorithm converges.

The two data sets under consideration include a training data set D_c containing N_c epochs of EEG signal matrix $\overline{\mathbf{X}}_i \in \mathcal{R}^{L\times T}, i = 1,\dots,N_c$ with labels $y_i \in \{+1,-1\}$, $(i = 1,\dots,N_c)$, and a test data set D_t containing N_t epochs of downsampled EEG signal $\overline{\mathbf{X}}_i \in \mathcal{R}^{L\times T}, i = N_c+1,\dots,N_c+N_t$ without labels, where L denotes the number of EEG channels and T denotes the number of samples in time domain.

Firstly, all channels are ranked according to their individual Fisher ratio in descending order. Secondly, the ranked channels are grouped into a number of subsets followed by computing Fisher ratio for each subset. Typically, L subsets can be obtained, with the first subset containing the top ranked channel, the second subset containing the top two ranked channels, and so on. Finally, the subset of channels with the highest Fisher Ratio is supposed to bear the most significant discriminability, and is chosen for classification. The self-trained channel selection and classification algorithm is summarized below for P300-based BCI.

Algorithm 1: Channel Selection

Define:
[1] Feature vector construction function: $FV(\overline{\mathbf{X}}_j, Q_i)$ consisting the operations of squeezing data point $\overline{\mathbf{X}}_j$ by deleting the rows not in the subset of channels Q_i and then vectorizing the resultant matrix. $FV(\overline{\mathbf{X}}_j, l)$ picks out the lth row of $\overline{\mathbf{X}}_j$.
[2] Iteration stopping criterion: the normalized difference between labels predicted in two successive iterations being less than a predefined threshold δ_1.
Input: the training set $D_c = \{\overline{\mathbf{X}}_1, \overline{\mathbf{X}}_2, \dots, \overline{\mathbf{X}}_{N_c}\}$ and their corresponding labels $\{y_1, y_2, \dots, y_{N_c}\}$ the test set $D_t = \{\overline{\mathbf{X}}_{N_c+1}, \overline{\mathbf{X}}_{N_c+2}, \dots, \overline{\mathbf{X}}_{N_c+N_t}\}$
threshold δ_2 for stopping the iterations
$iter = 1$
$\mathbf{x}_{l,j} = FV(\overline{\mathbf{X}}_j, l)$, for $l = 1$ to L, $j = 1\ to\ N_c + N_t$
Repeat
  For $l = 1\ to\ L$
    If $iter == 1$
      $\{\mathbf{w}, b\}$ =SVMtrain$\{(\mathbf{x}_{l,j}, y_j)|j = 1,2,\dots,N_c\}$ by solving Eq.(1)
    Else
    $\{\mathbf{w}, b\}$ =SVMtrain$\{(\mathbf{x}_{l,j}, y_j)|j = 1,2,\dots,N_c+N_t\}$ by solving Eq.(1), where y_j, $j = N_c+1$, $\dots, N_c+N_t$ are the labels predicted in the previous iteration.
    End
    For $j = N_c + 1\ to\ N_c + N_t$
      $y_{l,j}$ =SVMclass$(\mathbf{x}_{l,j}, \mathbf{w}, b)$ according to Eq.(2)
    End
    with $\{(\mathbf{x}_{l,j}, y_j)|j = 1,\dots,N_c\}$ and $\{(\mathbf{x}_{l,j}, y_{l,j})|j = N_c+1,\dots,N_c+N_t\}$,
    obtain $FR(l)$ by Eq.(3)
  End
  rank $FR(1),\dots,FR(L)$ in a descending order, and the corresponding channel sequence is denoted as a vector Q
  For $i = 1$ to L
    $Q_i = \{Q(1), Q(2),\dots,Q(i)\}$, which defines an i-channel subset $\mathbf{x}_{Q_i,j} = FV(\overline{\mathbf{X}}_j, Q_i)$, for $j = 1$ to $N_c + N_t$
    If $iter == 1$
      $\{\mathbf{w}, b\}$ =SVMtrain$\{(\mathbf{x}_{Q_i,j}, y_j)|j = 1,2,\dots,N_c\}$ by solving Eq.(1)

Algorithm 1: Channel Selection

Else

$\{\mathbf{w}, b\}$ =SVMtrain$\{(\mathbf{x}_{Q_i,j}, y_j)|j = 1, 2, \ldots, N_c + N_t\}$ by solving Eq.(1)

End

For $j = N_c + 1 \ to\ N_c + N_t$

$y_{Q_i,j}$ =SVMclass$(\mathbf{x}_{Q_i,j}, \mathbf{w}, b)$ according to Eq.(2)

End

calculate $FR(Q_i)$ by Eq.(3)

End

$Q^{(s)} = \arg\max_{Q_i}\{FR(Q_1), \ldots, FR(Q_L)\}$

corresponding predicted labels

$y_j = y_{Q^{(s)},j}, j = N_c + 1, \ldots, N_c + N_t$

$iter = iter + 1$

Until stopping criterion satisfied

Output: the subset of channels $Q^{(s)}$ and the corresponding labels $y_{Q^{(s)},j}$, $j = N_c + 1, \ldots N_c + N_t$

III. Data analysis and results

In this section, we illustrate the application of Algorithm 1 on the data set II of a P300 speller from BCI Competition III [7]. The data is briefly described as follows. Each subject was presented with a 6×6 matrix of characters shown in Fig. 1, and was asked to pay attention to one character in each run. His/her 64-channel EEG signal was sampled at 240 Hz. The data set was recorded from two different subjects (A and B). The sequence of 12 row-column intensifications was repeated 15 times (named "repeats" in this paper) for the spelling of each character. For each subject, the data of totally 185 character spellings were provided by the organizer. We adopt similar pre-processing techniques as in [8] for the convenience of comparison of different algorithms. For each channel, the signals between 200-500 ms posterior to the beginning of an intensification have been extracted and processed with a bandpass filter of 0.1 to 10 Hz. The extracted signal has been decimated by a rate of 10. The data point resulting from a post-stimulus signal is of dimension 7×64, representing 7 temporal samples and 64 channels respectively.

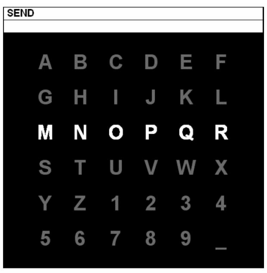

Fig. 1. User interface of the P300 speller used in BCI Competition III

In the following data analysis with Algorithm 1, we simply use the first 5 consecutive characters provided by the Competition as the initial training set to simulate a small training set scenario. The next 20 characters were used as the test data set without labels for retraining in Algorithm 1. Regarding the independent test set, we use the 100-character test set provided by the Competition so that the results are comparable to the other methods. We perform Algorithm 1 for channel selection. The threshold δ_1 for stopping the iterations of Algorithm 1 is set as 0. Furthermore, we find that the algorithm converges often within 3 iterations. Regarding the number of repeats being 15, results show that the optimal number of channels being 35 for subject A. The selected number of channels for subject B is 30. At the same time, we perform prediction of labels for the test data set. Performance has been evaluated according to the percentage of correctly predicted characters in the test datasets and in the independent test sets. For the number of repeats being 3, 4, 5, 10, or 15, the accuracies of the prediction averaged over the two subjects obtained by our Algorithm 1 are shown in Table I.

TABLE I

Accuracy of character prediction in percentage. (The results of both our Algorithm 1 and the standard SVM are based on the training set with 5 characters, while Rakoto's results [8] were based on 85 characters as training data. "Test set" indicates the 20 unlabelled characters used in retraining in Algorithm 1. "Ind. test set" indicates the independent test set with 100 unlabeled characters, which is the same as the test data set in the Competition.)

	test set (20)	ind. test set (100)
our Algorithm 1 with time segment 200-500 ms (5 training characters)	**90**	88.5
semi-supervised SVM without channel selection (5 training characters)	87.5	85.5
standard SVM (5 training characters)	77.5	80
Rakoto's method (85 training characters)	not applicable	**96.5**

For comparison, we applied a standard SVM without self-training to the same data sets as that used in the above evaluation of Algorithm 1. From Table I, the prediction accuracy of the standard SVM is much lower than the proposed algorithm with respect to both the 20-character test set and the independent test set.

We further compare the results of our Algorithm 1 with the best performance achieved by Rakotomamonjy and Guigue (whose method [8] will be denoted by Rakoto's method in the following). Notice that these two methods were applied to different size of the training data set, but the same 100-character independent test set. As mentioned above, the results of our Algorithm 1 are based on the training set with 5 characters. Since the recursive channel elimination approach in Rakoto's method requires sufficient labelled data, it is not applicable to the case of small training set. Therefore, we cite the results in [8] which were based on 85 characters as training data. From Table I, it is found that, at the number of repeats from 3 to 15, our proposed algorithm achieves comparable performance to Rakoto's method although the sizes of their respective training sets are of significant disparity. The outstanding performance of Algorithm 1 can be explained by the iterative update to the model with the test data set and the predicted labels. Meanwhile, the results also confirm the efficiency of the semi-supervised channel selection that utilizes augmented training set instead of the small initial training set for reliable feature selection to improve the classification. As a consequence, this paradigm demonstrates that our algorithm can potentially reduce the training process of BCI speller while not affecting the accuracy.

IV. Conclusions

This paper focuses attention on improving the performance of P300 speller when training data is insufficient. In this case, traditional model selection methods, e.g., cross-validation, usually do not work. Herein, we presented a self-trained SVM algorithm for EEG channel selection, where Fisher ratio calculated by SVM scores was used as an index. The SVM classifier was retrained with both labeled training data and unlabeled test data to improve its performance of prediction, at the same time to achieve better channel selection. The data analysis results of the off-line example demonstrate the effectiveness of our algorithm.

References

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain–computer interfaces for communication and control," *Clin. Neurophysiol.*, vol. 113, no. 6, pp. 767–791, 2002.

[2] G. Dornhege, *Toward brain-computer interfacing*. MIT Press, 2007.

[3] B. Röder, F. Rösler, E. Hennighausen, and F. Näcker, "Event-related potentials during auditory and somatosensory discrimination in sighted and blind human subjects," *Cognitive Brain Research*, vol. 4, no. 2, pp. 77–93, 1996.

[4] L. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalography and clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988.

[5] E. Donchin, K. M. Spencer, and R. Wijesinghe, "The mental prosthesis: assessing the speed of a P300-basedbrain-computer interface," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 174–179, 2000.

[6] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced P300 speller performance," *Journal of neuroscience methods*, vol. 167, no. 1, pp. 15–21, 2008.

[7] B. Blankertz, K. R. Müller, D. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlögl, G. Pfurtscheller, J. R. Millán, M. Schröder, and N. Birbaumer, "BCI Competition III," *Fraunhofer FIRST. IDA, http://ida. first. fraunhofer. de/projects/bci/competition_iii*, 2005.

[8] A. Rakotomamonjy and V. Guigue, "BCI competition III: Dataset II-ensemble of SVMs for BCI P300 speller," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 3, pp. 1147–1154, 2008.

[9] L. Citi, R. Poli, C. Cinel, and F. Sepulveda, "Feature selection and classification in brain computer interfaces by a genetic algorithm," in *Proceedings of the Genetic and Evolutionary Computation Conference*. Citeseer, 2004.

[10] Y. Li and C. Guan, "Joint feature re-extraction and classification using an iterative semi-supervised support vector machine algorithm," *Mach. Learn.*, vol. 71, no. 1, pp. 33–53, 2008.

[11] V. Vapnik, *The nature of statistical learning theory*. Springer Verlag, 2000.

[12] C. Bishop, *Neural networks for pattern recognition*. Oxford University Press, USA, 1995.

[13] T. N. Lal, M. Schroder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Scholkopf, "Support vector channel selection in BCI," *IEEE Trans. on Biomed. Eng.*, vol. 51, no. 6, pp. 1003–1010, 2004.