FEATURE EXTRACTION WITH MULTISCALE AUTOREGRESSION OF MULTICHANNEL TIME SERIES FOR P300 SPELLER BCI

Lin He, Zhenghui Gu, Yuanqing Li, Zhuliang Yu

College of Automation Science and Engineering, South China University of Technology, Guangzhou, China, 510640

ABSTRACT

P300 is one of the most studied components of event related potentials which reflects the responses of brain to events in the external environment. In this paper, we present a new method that utilizes multiresolution autoregression of multichannel time series (MAMTS) for feature extraction of P300 wave. First, it adopts multiresolution autoregression on dyadic tree to depict the characteristic of electroencephalogram (EEG) signal. Then the corresponding autoregression noise of multichannel time series is extracted as the feature. The experiment results verified the effectiveness of this new feature for P300 speller brain compute interface (BCI).

Index Terms—BCI, P300 speller, feature extraction, multiresolution autoregression, multichannel time series

1. INTRODUCTION

A brain–computer interface (BCI) is such a device that provides a direct communication channel from the user’s brain to the external world by reading the electrical signatures of brain’s activity [1][2]. BCI allows a brain to control a computer directly, without relying on normal neuromuscular pathways. The P300 event related potentials (ERP), first described by Sutton [3], is evoked in multichannel electroencephalogram (EEG) by rare task-relevant events. It is often recorded as a significant positive peak that occurs 300ms after an infrequent or significant stimulus. In 1988, Farwell and Donchin first introduced P300 potential into BCI [4] and it has been proven to be a reliable response for controlling a BCI [5]-[7].

There has been several machine learning and pattern classification algorithms about P300 based BCI [7][8]. Regarding feature extraction in P300 speller BCI, people usually concatenate data in different channels of the preprocessed multichannel EEG signal into a feature vector [7][8]. Developing new method of feature extraction from the preprocessed EEG may provide potential to improve the spelling accuracy of the P300 speller.

The area of multiresolution autoregression on dyadic tree (MADT) has flourished in the last decade [9][10]. It can uncover the relationship among observations in different scales effectively. The related successful applications include stochastic processes and image segmentation[9][10]. In multichannel time series analysis such as EEG analysis and communications, different signals demonstrate different characteristics of correlation among frequency bands. Therefore, it is valuable to study method of MADT suitable for multichannel time series recognition or classification, which can be utilized to analyze EEG signal in BCIs.

In this paper, we present a novel multiresolution autoregression of multichannel time series (MAMTS) based method for EEG feature extraction in P300 speller application, which is inspired by MADT. Experiment results show that our method can extract useful discriminant information for classification of EEG time series and demonstrate the capability to enhance the spelling accuracy of P300 speller.

2. P300 SPELLER

In our lab, we use NeuroScan SynsAmps2 as EEG signal amplifier for the P300 speller BCI system. It can record 64 channels of EEG data at most from electrodes placed according to the extended 10-20 system. Fig 1 shows the graphic user interface of this P300 speller. The upper half of the screen is the editing area, and the lower half part of the screen is a virtual keyboard. There are totally 40 buttons, including all English alphabets, numbers 1 to 10 and 4 other characters. All buttons successively and randomly flash at a rate of 250 Hz. With respect to each button flashing that contains a desired character, the EEG responses of the user evoked by these infrequent stimuli are different from those evoked by the stimuli that do not contain the desired character. The former usually involves a P300 wave. For each character epoch, the interface is designed as follows: the matrix is displayed for a 70ms period, and during this period...
period each character has the same intensity (i.e. the matrix was blank). Subsequently, each button in the keyboard randomly is intensified for 100 ms. After the intensification of a button, the keyboard returns blank for 30 ms. Button flashing is randomized in block of 40 buttons. The set of 40 intensifications is repeated 10 times for each character epoch, i.e., there are 10 trials in a character epoch. Thus, there are totally 400 flashing in each character epoch. Each character epoch is followed by a 180ms blank period. This period inform the user that the input of the current character is completed and he should focus attention on the next character.

Fig 1 GUI of our P300 speller

3. MAMTS BASED FEATURE EXTRACTION

For a certain segment of EEG data in P300 speller, let hypotheses \( H_0 \) and \( H_1 \) denote one class without P300 signal and the other class with P300 signal, respectively. We have

\[
H_0: X \sim p(x_i(y_{-i}), \ldots, x_{N_C}(y_{-i})), \ldots,
\]

\[
H_1: X \sim p(x_i(y_{-i}), \ldots, x_{N_C}(y_{-i})), \ldots,
\]

where \( X = (x_i(y_{-i}), \ldots, x_{N_C}(y_{-i})), \ldots, x_{N_C}(y_{-i}) \) denotes multiscale observation of the EEG data in scale 0, ..., L-1, \( N_C \) denotes the number of channels, \( x_i(j) \) denotes the data of \( i \)th channel in scale \( l \).

For \( H_0 \) hypothesis, in terms of product rule of joint probability density function (PDF), as well as Markov property of \( P \)-order Markov chain being characteristics of observation of multiscale multichannel time series, the multiscale likelihood of class one can be approximately formulated as

\[
p(X(y^{L-1}), X(y^{L-2}), \ldots, X(y^0)) \mid H_0
\]

\[
= p(X(y^{L-1}) \mid H_0)p(X(y^{L-2}) \mid X(y^{L-1}), H_0) \ldots
\]

\[
p(X(y^{L-P-2}) \mid X(y^{L-P-1}), \ldots, X(y^{L-P}), H_0) \ldots
\]

\[
p(X(y^0) \mid X(y^1), \ldots, X(y^{L-2}), H_0)
\]

\[
(2)
\]

where \( X(y') = [x_i(y'), x_i(y'), \ldots, x_{N_C}(y')] \) represents \( N_C \) channel EEG data mapped onto the scale \( i \). MADT [10] can reveal the relationship among the data in different channels and scales. If assuming that multiscale observation of the \( i \)th channel \( (x_i(y^{L-1}), \ldots, x_i(y^0)) \) and that of the \( j \)th channel \( (x_j(y^{L-1}), \ldots, x_j(y^0)) \) are independent, we formulate the relationship among data in different scales as follows

\[
X(y') = \sum_{j=0}^{P} A_{i(j)} \cdots 0 \quad a_{i0} + v_i
\]

\[
(3)
\]

where \( a_{i0} \) is the bias vector, \( v = (v_1, v_2, \ldots, v_B)^T \) is noise vector. Without loss of generality, we assume

\[
A_{i(j)} = a_{i0} I_{N_C}
\]

\[
(4)
\]

We can calculate the coefficient matrices in (4) via minimizing square sum of residuals from all channels and scales (SSRCS) defined as

\[
J_{SSRC} = (X(y') - \sum_{j=0}^{P} A_{i(j)} \cdots 0 a_{i0} + v_i)^T X(y') \sum_{j=0}^{P} A_{i(j)} \cdots 0 a_{i0} + v_i
\]

\[
(5)
\]

Obviously, minimizing \( J_{SSRC} \) is equivalent to minimizing the sum of square sums of regression residuals among different scales in each channel (SSRSEC). Thus, we obtain

\[
(6)
\]

where \( J_{SSRC,N_C} \) denotes the SSRSEC corresponding to \( iN_C \)th channel.

According to (4), the formulation of multiscale autoregression of \( k \)th channel comprised in (6) is

\[
x_i(y') = \sum_{j=0}^{P} a_{i(j)} x_i(y^{j-1}) + a_{i0} + v_i
\]

which depicts the relationship of data in different scale corresponding to \( k \)th channel. Autoregression coefficients in (7) can be solved through minimizing square sum of a single channel (SSSC).

\[
J_{SSSC} = \sum_{k=0}^{N_C}(x_{i,j}(y') - \sum_{j=0}^{P} a_{i(j)} x_{i,j}(y^{j-1}))^2
\]

\[
(8)
\]
Thus, coefficient matrices of multiscale regression in (5) can be obtained through (6) and (7). Due to the space limitation, the following equation is given while the proof is omitted
\[ p_x(v_i|H_0) = p(x(y^{(v_i)}),x(y^{(v_i+1)}),H_0) \]

This equation shows the statistical equivalent of autoregression noise and autoregression coefficients calculated from data in scale \( i \) to \( i + P \).

Other factors decomposed in (2) follow similar deduction. With respect to class with P300, there is a similar procedure of deduction as class without P300. Different multiscale autoregression noises corresponding to different categories of P300 signals have different statistical characteristics. Thus, multiscale autoregression noises can be utilized as the feature vector to discriminate the EEG segment without P300 potential and that with P300. We call this feature MAMTS based feature.

4. EXPERIMENTAL RESULTS

Data utilized in this section is from the P300 speller experiment of one subject in our lab. We use 30 channels in EEG data collection, including channel Fp1, Fp2, F7, F3, Fz, F4, F8, T7, T3, T4, T8, CP7, CP3, CPZ, CP4, TP7, P7, P3, Pz, P4, P8, O1, O2, Oz, O2. During the experiment, a 4-by-10 matrix that include 40 different characters are presented to the subject on the computer screen. Each button flashes successively in a random order. One of these 40 flashes in a run contains a desired symbol, where P300 potential is generated. For each character, 10 runs of 40 intensifications are carried out. Thus, when the user focuses on a character, there is 10 \( \times \) 40=400 button flashings in which 10 flashings happen on the button containing the character, while the other 390 flashings happen on the other 39 buttons not corresponding to desired characters.

Character series presented to the subject when the system collect EEG signal is “N7615ER1.WMS0VTPJUDB KAY1H4,DSO2NY7KB.AMF” which contains 47 characters totally. In fact, different character series with same length makes no difference on spelling accuracy [9]. The data from all the channels are used in our experiment. The preprocessing has 3 steps, including lowpass filtering, trial averaging and downsampling. Frequency band of lowpass filtering is 0 to 20Hz. Downsampling rate is 3. In an epoch, the 10 trials corresponding to one character are averaged.

As P300 peak usually occurs in the adjacent of 300ms after the visual stimuli happens, we extract features of signal segment between 48 ms to 556ms from the start of the intensification, i.e., 128 data points, also for the convenience of the following data decomposition. For the conventional P300 feature extraction, each feature vector of analyzed data contains 128 \( \times \) 30/3 = 1280 features. If there are \( N_v \) training character epochs and \( N_t \) test character epochs, the corresponding number of training sample is \( 40 \times N_v \) and number of test sample is \( 40 \times N_t \). For the feature constructed in this paper, we merely calculate autoregression noise vector between scale 0 and scale 1 to present preliminary results due to space limitation. Dimension of the corresponding feature vector is the same as that of the conventional feature mentioned above.

We adopt SVM to obtain the scores of samples. Let class with P300 potential be denoted as 1 and class without P300 be denoted as -1. We apply

\[
\text{score}_{ij}^\text{P300 present in ith trial} = \max_{j \neq -1,0} (\text{score}_{ij})
\]

\[
\text{score}_{ij}^\text{P300 absent in ith trial} = \max_{j \neq -1,0} (\text{score}_{ij})
\]

as the P300 detection rule. In our experiment, SVM utilized is SVM with linear kernel and its parameter is determined.

In our experiment, the whole dataset is divided into two parts. One is for training, the other is for test. For the purpose of comparison, we conducted four kinds of partitions, with the training epoch containing character No. 1 to 15, 1 to 20, 1 to 25 and 1 to 28, respectively. Correspondingly, the test set contains character No. 16 to 47, 21 to 47, 26 to 47 and 29 to 47, respectively. Fig. 2(a) and 2(b) show the SVM scores when conventional feature and our feature are utilized for 27 test character epoch, respectively. Fig. 3(a) and 3(b) shows the curve of
score corresponding to 17th character. From these two figures, it can be seen that scores of two features are different. That indicates that we extract different discriminant information from different features.

Table 1 demonstrates the spell accuracy corresponding to conventional feature and MAMTS feature. In addition, spell accuracy is also shown when the sum of the values of MAMTS feature and conventional feature is applied. We name this feature as integrated feature whose corresponding spell accuracy, to a certain extent, can verify if our MAMTS feature extract new discriminant information from EEG data compared with the conventional feature.

<table>
<thead>
<tr>
<th>Number of training epochs</th>
<th>Number of test epochs</th>
<th>MAMTS feature</th>
<th>Integrated Feature</th>
<th>Conventional feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>32</td>
<td>0.7813</td>
<td>0.9375</td>
<td>0.8125</td>
</tr>
<tr>
<td>20</td>
<td>27</td>
<td>0.8519</td>
<td>0.9259</td>
<td>0.8519</td>
</tr>
<tr>
<td>25</td>
<td>22</td>
<td>0.9091</td>
<td>0.9091</td>
<td>0.8636</td>
</tr>
<tr>
<td>28</td>
<td>19</td>
<td>1</td>
<td>0.9474</td>
<td>0.8421</td>
</tr>
</tbody>
</table>

It can be seen that MAMTS feature constructed in this paper can lead to higher spell accuracy than that of the conventional feature at moderate size of the training set, while integrated feature always leads to higher spell accuracy than the conventional feature at all illustrated training data size. And it is shown that our MAMTS feature really reveals the effective and, relative to conventional feature, new discriminant information in P300 signals.

5. CONCLUSIONS

In this paper, we construct new MAMTS feature for P300 speller BCI. In our experiment, it is shown that this feature leads to higher accuracy compared with the conventional feature when the training set is not small. If we integrate the MAMTS feature with the conventional feature, the corresponding spell accuracy is always higher than that of the conventional feature even if the training set is small. Therefore, the effectiveness of our MAMTS feature for P300 speller is verified.

REFERENCES


