A BCI System Based on Wavelet Decomposition and Support Vector Machine with a Dual-Class Voting Mechanism

Guang Yang, Kenji Nakayama and Akihiro Hirano
Graduate School of Natural Science and Technology, Kanazawa University
Kakuma-machi, Kanazawa, Ishikawa 920-1192, Japan
E-mail: buaagy@gmail.com

***Abstract*—An original dual-class voting mechanism was put forward as the final decision method for brain computer interface, which is based on wavelet decomposition (WD) to extract the brainwave features from EEG signals, and support vector machine (SVM) to classify five mental tasks. Moreover, several preprocessing methods were applied. Segmentation along the time axis for increasing the correct classification rate, and nonlinear as well as linear normalization for emphasizing the important information in small magnitude and optimizing data distribution. Further, an especial grouping method was proposed to realize optimizing parameters automatically. Approximately, about 90%~95% of correct classification rate is obtained based on the proposed method.**

Keywords-dual-class voting mechanism; brain computer interface (BCI); wavelet decomposition (WD); support vector machine (SVM); grouping method

I. INTRODUCTION

Nowadays, some kinds of interactive systems between handicapped persons and external devices have been proposed and developed. Among them, brain computer interface (BCI) [1] has been attractive recently. When a person imagines some mental task, the brainwave signals are measured and analysed to identify the mental task. Furthermore, the external devices, such as computers or machines are controlled. The BCIs are intended to assist, or repair human cognitive or sensory-motor functions [2]. Currently, as shown in Fig.1, the realization process of a BCI system is consist of brainwave measurement, feature extraction, classification and machine operation.

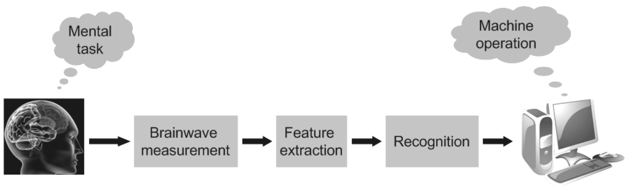

Fig.1. Realization process of a BCI system.

At present, with respect to the design and constructing for a BCI system, there're two main concerned problems: how to select and extract the useful brainwave features, and how to recognize and classify the mental tasks effectively. In previous research, many feature extraction methods and classification algorithms for BCI systems have been employed. For instance, the feature extraction methods mainly include time-frequency analysis, fast Fourier transform (FFT) and autoregressive model, etc. On the other hand, the commonly used classification algorithms include linear classifiers, neural networks [3], nonlinear Bayesian classifiers, and so on.

In this paper, we extract the energy information of the EEG data by using wavelet decomposition as the initial brainwave features. On the side, SVM is applied for mental task classification in view of its outstanding generalization capability. Besides, regard to the final decision method for classification, an novel dual-class voting mechanism is proposed, and several preprocessing methods are also applied effectually [4].

Simulations were carried out by using the EEG data sets, which are available from the web site of EEG Pattern Analysis Lab, Colorado state university [5]. In this work, an especial method is put forward to realize grouping and modifying parameters. Detailedly, among the data sets, one set is used for testing, one set is used for choosing mother wavelet and optimizing the parameters of SVM, and then, the rest data sets are used for training. The selections of data sets for training, modifying parameters and testing are changed. Finally, the correct classification rate is averaged among the possible selections. The simulation results have proved the effectiveness of proposed method.

II. PREPROCESSINGS AND FEATURE EXTRACTION

A. Brainwaves and Mental Tasks

In this paper, the brainwaves, which are available from the web site of Colorado state university are used and have been applied in many studies [6]. And the following five mental tasks are applied in this work.

· *Baseline (B)*
· *Multiplication (M)*
· *Letter-composing (L)*
· *Rotation of 3-D object (R)*
· *Counting numbers (C)*

Further, as shown in Fig.2, there're seven channels to simultaneously measure brainwaves, including C3, C4, P3, P4, O1, O2 and EOG. EOG is used for measuring the movement of eyeballs. Moreover, the brainwaves are measured for 10sec and sampled by 250Hz for each mental task and channel. Therefore, 10sec*250Hz=2500 samples are obtained for one channel of a mental task. One data set includes five mental tasks with seven channels.

B. Segmentation along Time Axis

In order to increase the correct classification rate and make the BCI responce fast, the brainwaves measured

Biography: Guang Yang(1986-), Male, Born in China, Master, Kanazawa University, Research areas: BCI, pattern recognition and signal processing.

during 10sec are divided into the segments with 0.2, 0.5, 2.0 and 5.0sec length (response time). The segmentation is shifted by the half of segment length. For instance, the segment with 2.0sec length can be obtained every 1.0sec, which is shown in Fig.3. In addition, it is noteworthy that, the segments with 0.2 and 0.5sec length (voting segments) are specifically used to form a dual-class voting system for classify the segments with 2.0, 5.0 and 10.0sec length (target segments).

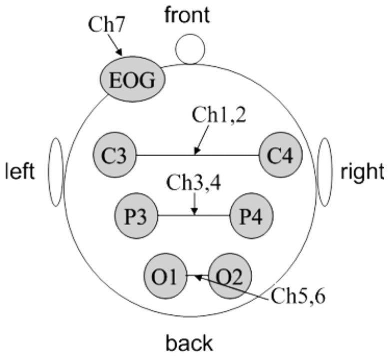

Fig.2. Location of seven channel electrodes.

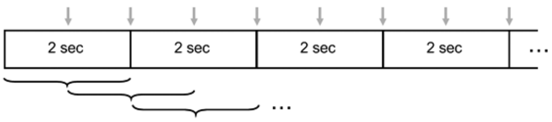

Fig.3. Segmentation of brainwaves along time axis.

C. Energy Information of Wavelet Decomposition

In this research work, the feature extraction was realized by adopting wavelet packdecomposition(WD). Normally, wavelet decomposition produces a family of hierarchically organized decompositions. The selection of a suitable level for the hierarchy depends on the signal and experience. Meanwhile, by adopting the wavelet toolbox in Matlab, the output decomposition structure usually contains the wavelet decomposition vector *C* and the bookkeeping vector *L*. As shown in Fig.4, the structure is organized as in this level-3 decomposition example.

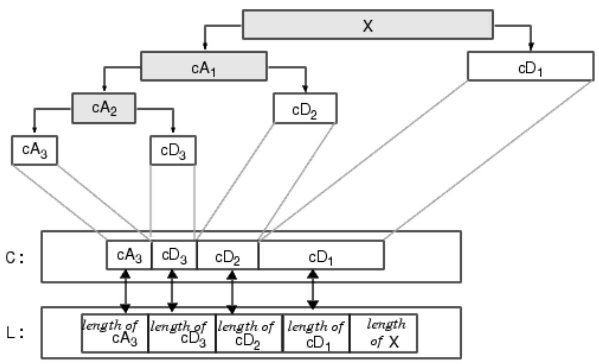

Fig.4. Level-3 decomposition structure of WD.

Moreover, in this work, the Daubechies wavelets [7] are used, which are a family of orthogonal wavelets defining a discrete wavelet transform (DWT) and characterized by a maximal number of vanishing moments for some given support. Daubechies wavelets D2-D20 (even index numbers only) are commonly used. In general, the index number refers to the number N of coefficients. Each wavelet has a number of vanishing moments equal to half the number of coefficients. For instance, D6 has three vanishing moments, D8 has four, etc. The vanishing moment limits the wavelet's ability to represent polynomial behaviour or information in a signal.

In this work, based on Daubechies wavelets D6, D8 and D10, the brainwave signals were decomposed up to 2~6 levels, which depended on different subjects and segment lengths. Further, through a searching process, the optimal mother wavelet and number of decomposition level are chosen.

Finally, all of the levels were selected, and then, energy information, in detail, the integration of square amplitude of WD in each level, are used as initial features. For instance, for the segment with 5.0sec length which decomposed up to 3 levels, there're 4 subspaces. Further, for this segment, the energy information formed by all the levels of j-th channel can be shown as $I_j=\{E_1,E_2,E_3,E_4\}$. Finally, the energy information in each channel is calculated, then the feature vector composed by all the 7 channels can be shown as $F=\{I_1,I_2,I_3,I_4,I_5,I_6,I_7\}$.

D. Nonlinear and Linear Normalization

The energy information of WD is widely distributed. However, small samples also contain some important information for classifying the mental tasks. For this reason, the nonlinear normalization as shown in Fig.5 and Eq.(1) is employed in this paper. The small samples are expanded and emphasized. Further, a linear normalization as shown in Eq.(2), is applied after nonlinear normalization in this work to optimize data distribution.

$$f(x)=\frac{\log(x-\min+1)}{\log(\max-\min+1)} \qquad (1)$$

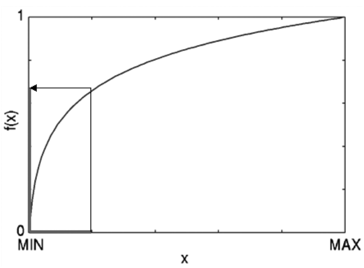

Fig.5. Nonlinear normalization. Horizontal axis is input and vertical axis is output.

$$f(x)=\frac{x-\min}{\max-\min} \qquad (2)$$

III. SUPPORT VECTOR MACHINE FOR MENTAL TASK CLASSIFICATION

Support vector machine (SVM) is an supervised learning method to analyze data and recognize patterns, commonly used for classification and regression analysis. Formally, SVM constructs a discriminant hyperplane that maximizes the margins to identify classes, as shown in Fig.6. Compared with other classifiers, SVM has a good generalization property and is insensitive to overtraining.

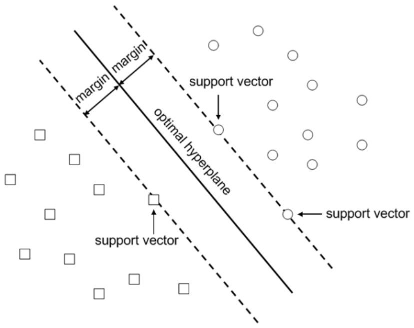

Fig.6. SVM searches a hyperplane that maximizes the margins.

Specifically, given a training set of instance-label pairs (xi, yi), i=1, . . . , k where xi $\in R^n$ and y $\in \{1,-1\}^k$, the SVM requires the solution of the following optimization problem:

$$\min_{w,b,\varepsilon} \frac{1}{2} w^T w + C\sum_{i=1}^{l} \varepsilon_i \tag{3}$$

$$\text{subject to} \quad \begin{aligned} y_i(w^T \phi(x_i) + b) &\geq 1 - \varepsilon_i, \\ \varepsilon_i &\geq 0. \end{aligned} \tag{4}$$

Here, training vectors xi are mapped into a higher dimensional space by the function ϕ . Then SVM finds a linear separating hyperplane with the maximal margin in this higher dimensional space. Furthermore, a SVM is possible to create nonlinear decision boundaries by using a kernel function $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ [8], with just a low increase of the classifier's complexity. The kernel used in this paper is the radial basis function (RBF) kernel:

$$K(x, y) = \exp(\frac{-\|x - y\|^2}{2\sigma^2}) \tag{5}$$

The RBF kernel can map samples nonlinearly, so it, can handle the case when the relation between class labels and attributes is nonlinear. The corresponding SVM is known as RBF SVM.

In this paper, concerning the input of SVM, the EEG signals from seven channels of one mental task are simultaneously applied. An example of the input of SVM is shown in Fig.7. In this figure, the left hand side shows the input before the nonlinear normalization, and the right is that after normalization. Since the decomposition level of WD is 3 here, there're 4 subspaces in total. Therefore, the feature vector length is 4*7=28, as shown in the horizontal axis.

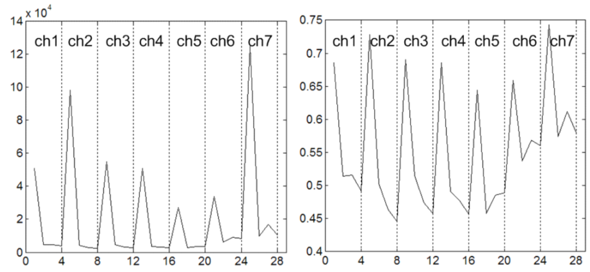

Fig.7. Input of SVM including 7-channels. Left: Before normalization, Right: After normalization.

Regard to the training target of SVM, in this work, we use a single number to represent a categorical attribute, and due to there're five mental tasks altogether, the training target vector for output is set as (1,2,3,4,5).

In addition, RBF SVM has two parameters, namely, the regularization parameter C and the RBF width σ , that normally need to be checked and defined by hand, or by the method of cross validation. In this paper, instead of the conventional approaches, a grouping method is put forward specially for searching the best parameters automatically in each independent trial. In other words, for different trial and training data sets, the parameters C and σ are also altered.

IV. GROUPING, DECISION AND EVALUATION METHODS

A. *Grouping Method*

For different subjects, the EEG data with 10sec length for five mental tasks were measured 10 or 5 times. Therefore, 10 or 5 data sets are available. In this paper, an especial grouping method is introduced for searching the best parameters automatically. Specifically, among the data sets, one set is used for testing, one set is used for modifying, which means choosing the mother wavelet, modifying the decomposition level of WPD, as well as optimizing the parameters C and σ of SVM. Then, the rest data sets are used for training. In addition, the search ranges of C and σ are both from 0.5 to 5, and the search step value is 0.5.

Further, 5 data sets are selected as the testing data set respectively. Thus, 5 independent trials were carried out in this work, and classification result is evaluated based on the average over 5 trials. For subject 2 with 5 data sets, the schematic diagram of grouping method is shown in Fig.8.

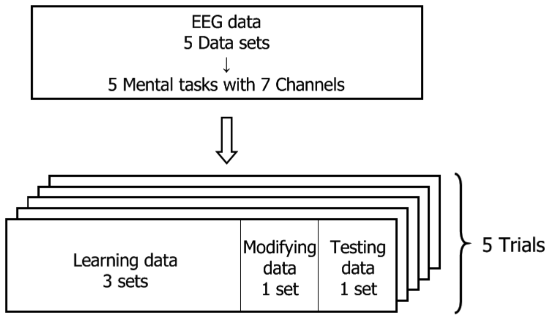

Fig.8. Schematic diagram of grouping method for subject 2

B. *Final Decision Method*

In this paper, regard to the final decision method for classification, an novel dual-class voting mechanism is introduced. In detail, for the segments with 2.0, 5.0 and 10.0sec length (target segments), the segments with 0.2 and 0.5sec length (voting segments) are used to vote in the two rounds of voting. In detail, in the first round of voting, the classification results of segments with 0.2sec length are applied for voting with a strong rejection criterion, which can provide a low misclassification rate. Then, the rejected target segments are forwarded to the second round of voting, and the classification results of segments with 0.5sec length are used with a well-balanced rejection criterion.

In addition, for the segment with N length, when using a segment with M length for voting, the total number of votes V is shown in Eq.(6). Furthermore, a schematic diagram to specifically represent the voting mechanism is shown in Fig.9.

$$V = N / M \times 2 - 1 \tag{6}$$

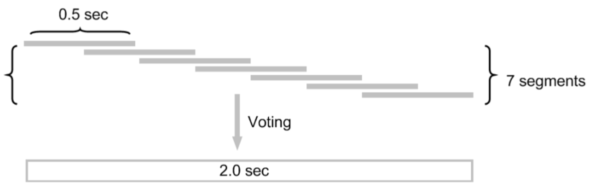

Fig.9. Specific diagram of the voting mechanism.

Furthermore, through setting different vote thresholds (VTs), the classification results can be modified. If the maximum of votes among 5 mental tasks is greater than the VT, the mental task with most votes is decided as the final classification result; otherwise, the classification result of SVM is rejected, that means any mental task cannot be estimated. In this paper, we set the VTs in the first round of voting as 10, 25 and 50 (the total numbers of votes are 19, 49 and 99), and three different VTs for each segment in the second round.

In order to contrast, for all of the segments with 2.0, 5.0 and 10.0sec length, the classification results have also been calculated without the proposed voting mechanism.

C. Evaluation Method

The evaluation of classification results is based on the probabilities of correct classification (Pc) and error classification (Pe), as well as the correct classification rate (Rc) [9].

$$P_c = \frac{N_c}{N_t} \tag{7}$$

$$P_e = \frac{N_e}{N_t} \tag{8}$$

$$N_t = N_c + N_e + N_r \tag{9}$$

$$R_c = \frac{P_c}{P_c + P_e} . \tag{10}$$

In the above equations, Nc is the number of correct classifications, Ne is the number of error classifications, and Nr is the number of rejections.

V. RESULTS OF EXPERIMENT

The simulation was executed by using MATLAB. In order to confirm the commonality, the EEG data of subject 1 and 2, which can be obtained from the web site of Department of computer science, Colorado state university, were used. According to the proposed methods (based on the dual-class voting mechanism or not), the simulation results, which estimated by Pc, Pe and Rc of testing data sets are shown in Table I~II. The probabilies were averaged over 5 independent trials.

TABLE I. Pc AND Pe OF TSETING DATA FOR SUBJECTS 1 AND 2 WITHOUT VOTING MECHANISM

	Subject 1		Subject 2	
Length	Pc	Pe	Pc	Pe
2.0s	0.853	0.147	0.800	0.200
5.0s	0.787	0.213	0.827	0.173
10.0s	0.840	0.160	0.680	0.320

TABLE II. Pc AND Rc OF TSETING DATA FOR SUBJECTS 1 AND 2 WITH DUAL-CLASS VOTING MECHANISM

	Subject 1			Subject 2		
Length	VT	Pc	Rc	VT	Pc	Rc
2.0s (19/7)	3	0.933	0.933	3	0.867	0.871
	4	0.924	0.941	4	0.853	0.901
	5	0.893	0.957	5	0.813	0.915
5.0s (49/19)	6	0.960	0.960	6	0.880	0.880
	8	0.960	0.960	8	0.853	0.889
	10	0.947	0.986	10	0.840	0.913
10.0s (99/39)	11	1.000	1.000	11	0.880	0.880
	14	1.000	1.000	14	0.880	0.880
	17	1.000	1.000	17	0.840	0.913

The number under "Length" is total number of votes in two rounds of voting.

As can be seen, for different subjects and segment lengths, the simulation performance is obviously changed. As extending the segment length (response time) from 2.0 to 10.0sec, the simulation performance can be improved.

The probability of correct classification (Pc) based on the method without the voting mechanism is 78.7~85.3% for subject 1, and 68.0~82.7% for subject 2. In contrast, through adopting the proposed method by using the dual-class voting mechanism, Pc has been raised to 93.3~100.0% for subject 1, and 86.7~88.0% for subject 2; in addition, the correct classification rate (Rc) for subject 1 is 95.7~100.0% , and 91.3~91.5% for subject 2. For different application purposes, we can modify the vote thresholds in the two rounds of voting.

In addition, along with changes of different subjects, segment lengths and independent trials, the Daubechies mother wavelet, decomposition level of WD, and two parameters of SVM are also altered automatically. For the segments with 2.0 and 5.0sec length of subject 1, 2 in 5 independent trials, these changes are shown in Table III.

TABLE III. CHANGES OF PARAMETERS FOR DIFFERENT SUBJECTS, SEGMENT LENGTHS AND TRIALS

SUB	Length	Trial	DW	Level	C	σ
1	2.0s	1	D6	2	2.0	1.0
		2	D8	2	1.5	5.0
		3	D6	5	1.0	2.5
		4	D10	2	5.0	2.5
		5	D10	4	2.5	0.5
	5.0s	1	D10	2	4.5	0.5
		2	D10	2	3.0	0.5
		3	D6	4	4.0	1.5
		4	D8	2	4.5	2.5
		5	D10	5	5.0	0.5
2	2.0s	1	D10	5	3.5	3.5
		2	D6	2	2.5	4.5
		3	D10	3	4.5	1.5
		4	D6	3	3.5	1.0
		5	D8	2	5.0	1.5
	5.0s	1	D10	2	5.0	1.0
		2	D8	2	5.0	2.0
		3	D6	3	3.0	1.0
		4	D8	2	4.5	2.5
		5	D10	2	4.0	1.0

SUB=Subject, DW=Daubechies wavelet

VI. CONCLUSION

A novel dual-class voting mechanism and grouping method have been put forward for a EEG-based BCI system, which is based on wavelet decomposition (WD) and support vector machine (SVM) to solve the BCI problem. Besides, preprocessing methods have also been applied effectively.

Simulations were executed by adopting the EEG signals from two subjects during five mental tasks. Compared to the conventional methods, the overall correct classification rates(Rc) of subject 1, 2 were increased to 95.7~100.0% and 91.3~91.5% based on the proposed method in this paper.

REFERENCES

[1] G. Dornhege, J. d. r. Millán, T. Hinterberger, D. J. McFanland and K. R. Müller: Toward Brain-Computer Interfacing, The MIT Press, 2007.

[2] G. Pfurtscheller and C. Neuper: Motor imagery and direct brain-computer communication, Proc. of the IEEE, vol. 89, no. 7, pp. 1123.1134, 2001.

[3] K. Nakayama, H. Horita and A. Hirano: Neural network based BCI by using orthogonal components of multi-channel brain waves and generalization, Lecture Notes in Computer Science, Vol. 5164, pp. 879-888, 2008.

[4] K. Nakayama and K. Inagaki: A brain computer interface based on neural network with efficient pre-processing, Proc. IEEE, ISPACS2006, Yonago, Japan, pp.673-676, Dec. 2006

[5] EEG Pattern Analysis: http://www.cs.colostate.edu/eeg/

[6] C. Anderson and Z. Sijercic: Classification of EEG signals from four subjects during five mental tasks, EANN'96, ed. by Bulsari, A.B., Kallio, S., and Tsaptsinos, D., Systems Engineering Association, PL 34, FIN-20111 Turku 11, Finland, pp. 407-414, 1996.

[7] Bang-hua Y., Guo-zheng Y., Rong-guo Y., and Ting W.: Feature extraction for EEG-based brain-computer interfaces by wavelet packet best basis decomposition. Journal of Neural Engineering, 3, 251-256, 2006.

[8] B. Sch¨olkopf and A. J. Smola: Learning with Kernels. MIT Press, 2002.

[9] K. Nakayama, Y. Kaneda and A. Hirano: A brain computer interface based on FFT and multilayer neural network-feature extraction and generalization, Proc. IEEE, ISPACS, pp. 101-104, 2007.