31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

MEG Based Classification of Wrist Movement

Nasim Montazeri, Mohammad Bagher Shamsollahi and Sepideh Hajipour
Biomedical Signal and Image Processing Laboratory (BiSIPL), School of Electrical Engineering,
Sharif University of Technology, Tehran, Iran

Abstract— **Neural activity is very important source for data mining and can be used as a control signal for brain–computer interfaces (BCIs). Particularly, Magnetic signals of neurons are enriched with information about the movement of different part of the body such as wrist movement. In this paper, we use MEG (Magneto encephalography) signals of two subjects recorded during wrist movement task in four directions. Data were prepared for BCI competition 2008 for multiclass classification. Our approach for this classification problem consists of PCA as a noise reduction method, ULDA for feature reduction and various linear classifiers such as Bayesian, KNN and SVM. Final results (58%-62% for subject 1 and 36%-40% for subject 2) prove that the suggested method shows better performance compared with other methods.**

Keywords: MEG, BCI, SVM classifier, LDA.

INTRODUCTION

A brain–computer interface (BCI) can be used to control applications based on signals measured, invasively or noninvasively, from the human and animal brain. BCIs can thus, e.g. help severely motor-disabled persons to obtain some motor control and ability to communicate [1]. A prominent approach to BCIs uses a supervised training phase where brain signals or movement imagery are recorded and a classifier is trained to separate different classes, e.g. left, right, up or downward wrist movement [2]. Success requires the effective interaction of two adaptive controllers: the user's brain, which produces brain activity that encodes intent, and the BCI system, which translates that activity into device control commands [3]. Indeed, BCIs have been primarily conceived as a potential new therapy to restore motor control in severely disabled patients [4].

BCI Competitions are organized in order to foster the development of improved BCI technology by providing an unbiased validation of a variety of data analysis techniques. In each competition a variety of data sets was made available in a documented format via internet. Each data set is a record of brain signals from BCI experiments of leading laboratories in BCI technology that split into two parts: one part of labeled data ('training set') and another part of unlabeled data ('test set'). Researchers throughout the world can present their methods to the training data and submit the output of their innovated algorithms for the test data. The true labels of the test data were kept secret until deadline, were used to evaluate the submissions. This procedure guarantees that the assessment of performance is not biased by over fitting the selection of methods and the choice of their parameters to the data. For the first time, in BCI competition 2008, MEG (Magneto encephalography) signal recorded during wrist movement task has been presented. In this paper, we explain a method for classification of MEG wrist movement signals. In the following sections, the characteristics of data set and different stages of classification procedure are explained. Finally, we present experimental results demonstrating the capabilities of our approach and comparing its effectiveness with the winner of BCI 2008.

I. DATA

We have used the data sets 3 on the BCI competition 2008 web site. The data set contains directionally modulated low-frequency MEG activity that was recorded while two right handed subjects performed wrist movements in four different directions. The task was to move a joystick from a center position toward one of four targets located radially at 90° intervals (four-class center-out paradigm) using exclusively the right hand and wrist. Movement amplitude was 4.5 cm.

In each trial, the target was self-chosen by the subject. Targets were arranged in the form of a rhombus in the horizontal plane with corners pointing left, right, away from and toward the subject's body. For each subject, there are only 10 channels of MEG signals (filtered to 0.5-100Hz, 400Hz sampling rate), which are mostly chosen from left side of the brain, and 40 trails for each class as the training set. In one trail, visual trigger signals were presented on a screen to start a trial or to indicate possible time violations and the subject was asked to move a joystick to a self desired direction. Signals consist of samples from 0.4sec before to 0.6sec after the movement. For classification, all the channels and time samples are used [13].

II. METHODOLOGY

There are different stages in all procedures for classification such as feature extraction and training stage. However, most of the time, extracted features require special selection or reduction in dimension to maximize the discrimination between classes. First, two feature groups were chosen during feature selection. All the selected features are gathered in a unique matrix called feature matrix. Remaining procedure is performed on the this matrix. Because of high dimensionality, it is necessary to reduce the dimension. Besides the Dimension reduction, 10% of principle components are thrown out to eliminate

noise effects. Our method can be simply explained in a chart (Fig. 1):

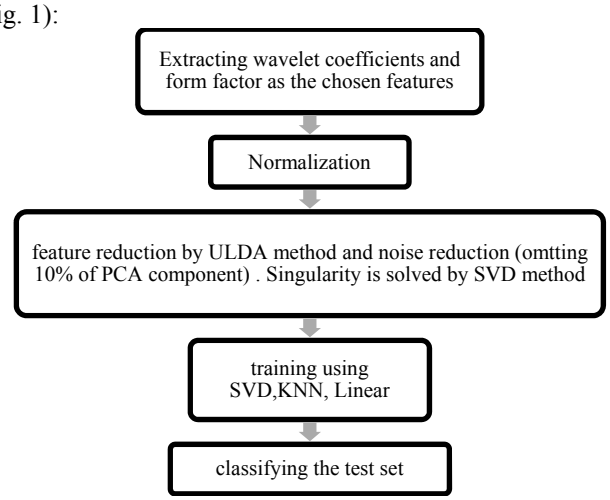

Fig. 1. Chart of the proposed method

In the following subsections, different parts of this diagram are stated.

A. Feature Selection

In MEG signal processing, researchers mostly used frequency features such as power spectra of famous bands like Gamma [1], [8]. However we were eager to find new features which are more prominent. Many features, listed in a table below, are extracted among various types.

TABLE I
LIST OF FEATURES

type	Features
time	mean, variance, form factor
frequency	Power spectra in δ, α, β, θ, γ bands, DCT, DST,
time-frequency	wavelet
entropy	shannon, renyi
parametric	AR coefficients: 4,8,12,16

For selecting the most significant features, a classifier-based feature selection algorithm has been used. The training set is divided to two different set: 1) train-train (2/3 of training set), 2) train-test (the remaining). After feature extraction, feature matrix is normalized in order to have different features in a same range. Features from different types are in different ranges so normalization is really essential to transfer the most information to classifier. Simply, linear SVM classifier was trained by the train-train data and then used for classifying the train-test data. Features with best classification accuracy were selected. According to this kind of selection two groups of features show more performance in classification:

1- Wavelet coefficients (db8): Each signal is divided in to two parts wavelet decomposition in: 1) Approximation consisting of a part of a signal with low frequency. 2) Details consisting of the other part with high frequency. As a feature in this work, approximation part is used because neural activates of movements occurred in low frequencies significantly.
2- Form factor: Form factor can be defined as the equation below:

Using all of these two feature groups, can enhance the classification accuracy.

B. Feature Reduction

In this stage, although features with most accuracy are chosen but still the accuracy is too low to justify about the performance of the algorithm. Looking at the feature space, we can conclude that samples of different classes are located randomly (Fig.2) and also the quantity of dimension of features are too high. Consequently, a supervised technique should be added to extend the separability between classes and also reduce the dimension of features. A commonly used feature reduction technique is LDA [6]. LDA searches for a linear transformation (A) to transfer the features of each sample (x_i) to a new space where the data points (y_i) of different classes are far from each other while requiring data points of the same class to be close to each other.

$$y_i = Ax_i \quad (1)$$

LDA considers maximizing the following criterion:

$$A = \max_A \frac{\mathrm{tr}(A^T S_b A)}{\mathrm{tr}(A^T S_t A)} \quad (2)$$

where, S_b is the "between classes scatter matrix". S_w can be defined as the "within classes scatter matrix". The summation of these two matrices gives us S_t as the total scattering matrix. The definitions of the scatter matrices are:

$$S_b = \sum_{k=1}^{c} m_k (\mu^{(k)} - \mu)(\mu^{(k)} - \mu)^T \quad (3)$$

$$S_w = \sum_{k=1}^{c} (\sum_{i=1}^{m_k} (x_i^{(k)} - \mu^{(k)})(x_i^{(k)} - \mu^{(k)})^T \quad (4)$$

$$S_t = S_w + S_b = \sum_{i=1}^{m} (x_i - \mu)(x_i - \mu)^T \quad (5)$$

where, c is the number of classes, m_k is the number of samples in k-th class, μ is the total sample mean vector, $\mu^{(k)}$ is the average vector of the k-th class, $x_i^{(k)}$ is the i-th sample in the k-th class and tr() denotes matrix trace. The optimization problem in Eq. (2) is equivalent to find the eigenvectors of following generalized eigen-problem associated with maximum eigen values:

$$S_b a = \lambda S_t a \quad (6)$$

where, a is the columns of A. Since the rank of Sb is bounded by c−1, there are at most c−1 eigenvectors corresponding to non-zero eigen values [10]. To get a stable solution of the above generalized eigen vector, S_t is required to be nonsingular which is clearly not true when the number of features are larger than the number of samples. In the past few decades, various approaches have been proposed to solve this problem. For instance, some additional preprocessing steps (*e.g.*, PCA, SVD) are required to guarantee the non-singularity of scatter matrices. Some popular methods for the first stage include Principle Component Analysis (PCA) [7] or (SVD) [9]. In this work, a method based on SVD is used to avoid singularity. Our approach in this stage can be found in [11], [12].

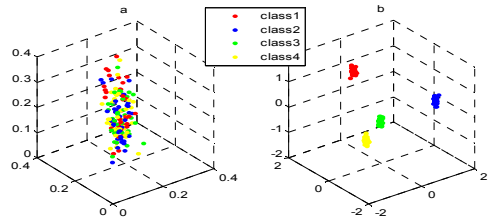

Fig. 2. Separability of the feature matrix before (a) and after (b) feature reduction by LDA. In (a), we have just shown the three first dimensions. After using LDA, dimension is reduced to c-1 (here as there are 4 class, dimension reduced to 3).

Let:

$$\bar{x}_i = x_i - \mu\,, \tag{7}$$

$$\bar{X}^k = \left[\bar{x}_1{}^{(k)}, \cdots, \bar{x}_{m_k}{}^{(k)}\right], \tag{8}$$

$$\bar{X} = \left[\bar{X}^{(1)}, \cdots, \bar{X}^{(c)}\right], \tag{9}$$

consequently, S_b and S_t can be expressed as:

$$S_t = \bar{X}\bar{X}^T, \qquad S_b = \bar{X}W\bar{X}^T, \tag{10}$$

$$W = \begin{bmatrix} W^{(1)} & 0 & \cdots & 0 \\ 0 & W^{(2)} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & W^{(c)} \end{bmatrix}$$

where $W^{(k)}$ is a $m_k \times m_k$ matrix with all the elements equal to $1/m_k$. Substituting Eq.(10) in Eq.(6), we can obtained a new form for the criterion:

$$\bar{X}W\bar{X}^T a = \lambda \bar{X}\bar{X}^T a \tag{11}$$

Suppose that we have the SVD of $\bar{X} = U\Sigma V^T$, so we can obtain:

$$U\Sigma V^T W V \Sigma U^T a = \lambda U \Sigma V^T V \Sigma U^T a$$

$$V^T W V b = \lambda b, \quad b = \Sigma U^T a \tag{12}$$

According to Eq. (12), b is the eigen vector of $V^T W V$. As a conclusion, finding the solution of the optimization can be summarized in following words:

1- Finding the SVD of $\bar{X} = U\Sigma V^T$.
2- Finding b as the eigen vectors of $V^T W V$.
3- Choosing one of the solutions for a like $a = U\Sigma^{-1} b$.

Some considerations should be noted for each part. First of all, in software implementation, the zero eigen values and their eigen vectors in SVD should be omitted. In this work, 90% of eigen values of Σ, are used to reduce the noise effect. This is exactly the same as using 90% of PCA component of $\bar{X}$ when all the zero eigen values are omitted. In calculation of matrix b, zero eigen vectors are also betaken. This kind of transformation is expected to separate classes linearly in new space (Fig.2).

C. Classifiers

Classification consists of estimating a qualitative variable *"the class label"* using a set of other variables. Numerous algorithms have been proposed in order to achieve data classification and to improve its efficiency. We used three different classifiers (SVM, KNN, and Bayesian based on Euclidean distance). SVM is used in many papers as an efficient classifier which results in one of the most accurate classification. Bayesian, according to its definition gives us a sense about the classification error so it can be used as an error criterion. A variation of KNN density estimation technique results in a suboptimal, yet popular in practice, nonlinear classifier [14]. So By these three classifiers, we can evaluate the efficiency of the feature space according to nonlinearity and optimality.

III. RESULTS & DISCUSSION

To validate the presented algorithm, two kinds of validation are used: 1) Hold-out Cross-Validation and 2) Leave-One-Out (LOO). In Hold-out the train set is divided into two part: train-train set (2/3 of train set) and train-test set (the rest of the train set). Then classifier is trained by train-train set and labels the train-test set. Choosing different samples to form the train-train would change the accuracy of labeling the train-test set, so we can repeat the procedure for instance 10 times in order to assess the results statistically and report the mean of accuracy as the result of validation. In LOO, the algorithm repeats as much as the number of samples of train set. Only one of the samples is kept out of the train-train set in each repetition. The results of validations are reported in Table II and III. For preparing the result of them, program was run for 10 times because shuffling was done on providing the train set. Table IV lists the result of presented algorithm with different classifiers

TABLE II
VALIDATION OF METHOD BY HOLD-OUT IN PERCENTAGE OF ACCURACY

subject	Classifier used in method	mean	mode	Variance(%)
1	SVM	42.88	38.46	6.48
	KNN	40.57	40.38	6.04
	Bayesian	43.46	40.38	5.22
2	SVM	30.38	32.69	3.60
	KNN	30.77	30.77	6.15
	Bayesian	33.65	32.69	3.87

TABLE III
VALIDATION OF METHOD BY LOO IN PERCENTAGE OF ACCURACY

subject	Classifier used in method	accuracy
1	SVM	46.88
	KNN	45
	Bayesian	47.5
2	SVM	33.75
	KNN	31.25
	Bayesian	31.87

applied to the final test set. These results are obtained by the whole training data (40 trail per class and 10 channels) in our hand and we can generally conclude that among the chosen classifiers, Bayesian shows the best performance of

all. The results obtained by Bayesian classifier are so better than the best result in BCI competition 2008.

A spectacular element of our method is using LDA to change the space to a linear one. Challenging to fit a hyper plane to mixed classes has more error. Table V and Fig.3 compare them and show the better performance of our method.

TABLE IV
PERCENTAGE OF ACCURACY FOR TEST DATA

subject	Classifier used in method	accuracy
1	SVM	58.11
	KNN	59.46
	Bayesian	62.16
2	SVM	38.36
	KNN	36.99
	Bayesian	39.73

TABLE V
COMPARISON OF PRESENTED METHOD WITH THE WINNER OF BCI COMPETITION 2008

subject	Best result of proposed method	Result of the winner
1	62.16	59.5
2	39.73	34.3
mean	50.94	46.9

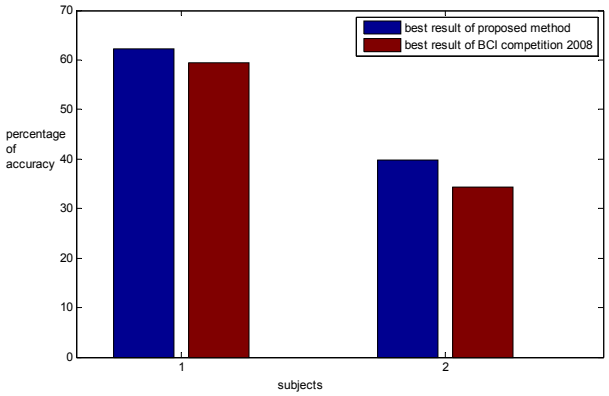

Fig. 3. Percentage of accuracy of presented algorithm and of the winner of BCI competition 2008.

REFERENCES

[1] L. Kauhanen, T. Nykopp, M. Sams: Classification of single MEG trials related to left and right index finger movements. Clinical Neurophysiology 117 (2006).

[2] Julie Blumberg, J¨orn Rickert, Stephan Waldert, Andreas Schulze Bonhage, Ad Aertsen, Carsten Mehring: Adaptive Classification for Brain Computer Interfaces. Conference of the IEEE EMBS, (2007)

[3] Benjamin Blankertz, Klaus-Robert Müller, Dean Krusienski, Gerwin Schalk, Jonathan R. Wolpaw, Alois Schlögl, Gert Pfurtscheller, José del R. Millán, Michael Schröder, Niels Birbaumer: The BCI Competition III: Validating Alternative Approaches to Actual BCI Problems. IEEE transactions on neural systems and rehabilitation engineering, vol. XX, no. Y, 2006

[4] Mikhail A. Lebedev and Miguel A.L. Nicolelis: Brain–machine interfaces: past, present and future. Elsevier TRENDS in Neurosciences Vol.29 No.9

[5] David Cohen, Eric Halgren: Magneto encephalography (Neuromagnetism). Preprint from Encyclopedia of Neuroscience (3rd Edition 2003, Elsevier)

[6] Cohen, Arnon, Biomedical and Automatic Recognition, Volume II, Compressin and Automatic Recognition, CRC Press, Second Printing, 1988.

[7] P. N. Belhumeur, J. P. Hepanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7):711–720, 1997.

[8] Stephan Waldert, Hubert Preissl, Evariste Demandt, Christoph Braun, Niels Birbaumer, Ad Aertsen, and Carsten Mehring: Hand Movement Direction Decoded from MEG and EEG. The Journal of Neuroscience, January23,2008,28(4):1000-1008; doi:10.1523/JNEUROSCI.5171-07.2008

[9] Peg Howland and Haesun Park: Generalizing Discriminant Analysis Using the Generalized Singular Value Decomposition. IEEE transactions on pattern analysis and machine intelligence, vol. 26, no. 8, August 2004

[10] K. Fukunaga. Introduction to Statistical Pattern Recognition. Academic Press, 2nd edition, 1990.

[11] Jieping Ye: Characterization of a Family of Algorithms for Generalized Discriminant Analysis on Under sampled Problems. Journal of Machine Learning Research 6 (2005) 483–502

[12] Deng Cai, Xiaofei He, Jiawei Han, "SRDA: An Efficient Algorithm for Large Scale Discriminant Analysis", IEEE Transactions on Knowledge and Data Engineering, 2007.

[13] BCI Competition IV, data sets 3, Available: http://ida.first.fraunhofer.de/projects/bci/competition_iv/

[14] Sergios Theodoridis and Konstantinos Houtroumbas: Pattern Recognition, second edition, Academic Press