Fuzzy support vector machine for classification of EEG signals using wavelet-based features

Qi Xu*, Hui Zhou, Yongji Wang, Jian Huang

Key Laboratory of Image Processing and Intelligent Control, Department of Control Science and Technology, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan 430074, China

A R T I C L E   I N F O

Article history:
Received 20 April 2008
Received in revised form 10 April 2009
Accepted 28 April 2009

Keywords:
Brain–computer interface
Discrete wavelet transform
Fuzzy support vector machine
Motor imagery
Electroencephalogram

A B S T R A C T

Translation of electroencephalographic (EEG) recordings into control signals for brain–computer interface (BCI) systems needs to be based on a robust classification of the various types of information. EEG-based BCI features are often noisy and likely to contain outliers. This contribution describes the application of a fuzzy support vector machine (FSVM) with a radial basis function kernel for classifying motor imagery tasks, while the statistical features over the set of the wavelet coefficients were extracted to characterize the time–frequency distribution of EEG signals. In the proposed FSVM classifier, a low fraction of support vectors was used as a criterion for choosing the kernel parameter and the trade-off parameter, together with the membership parameter based solely on training data. FSVM and support vector machine (SVM) classifiers outperformed the winner of the BCI Competition 2005 and other similar studies on the same Graz dataset, in terms of the competition criterion of the mutual information (MI), while the FSVM classifier yielded a better performance than the SVM approach. FSVM and SVM classifiers perform much better than the winner of the BCI Competition 2005 on the same Graz dataset for the subject O3 according to the competition criterion of the maximal MI steepness, while the FSVM classifier outperforms the SVM method. The proposed FSVM model has potential in reducing the effects of noise or outliers in the online classification of EEG signals in BCIs.

© 2009 IPEM. Published by Elsevier Ltd. All rights reserved.

1. Introduction

A brain–computer interface (BCI) system is intended to help severely disabled people to communicate with computers or control electronic devices through their thoughts. To achieve this goal, BCIs should be designed to translate brain signals into control commands. Various electroencephalographic (EEG) signals have been used as inputs to a BCI. The commonest BCI systems are based on the analysis of spontaneous EEG signals. There is increasing evidence indicating that some mental tasks can be recognized from spontaneous EEG signals [1–6].

If a BCI system is considered to be a pattern recognition system, the EEG recognition procedure mainly involves feature extraction and classification. The performance of a recognition system depends on the features and the classification algorithm employed. Most of the schemes in existence for extracting spontaneous EEG features are based on autoregressive (AR) models, fast Fourier transform (FFT), short-time Fourier transform (SFT) and wavelet transform (WT) [7–16]. The AR model or FFT model cannot capture the transient features in a given signal and the time–frequency information is not readily seen in the AR parameters or transformed Fourier coefficients. The SFT or Gabor transform is a time–frequency analysis method; information regarding time and frequency is localized by a uniform time window for all frequency ranges. The WT adapts the window size according to the frequency. If the basic wavelet function has a finite duration, the frequency information obtained from the WT is localized in time. Therefore, for non-stationary transient signals such as EEG, the WT is superior to FFT and SFT. In recent years, investigation and evaluation of classification algorithms in EEG-based BCI studies have been reported [17,18]. The µ rhythm (8–13 Hz) and β rhythm (13–22 Hz) originating in the sensorimotor cortex have been postulated to be good signal features for EEG-based BCIs [19–20]. Pfurtscheller et al. investigated the on-line classification of EEG signals by using adaptive autoregressive (AAR) parameters [9]. Qin and He developed a wavelet-based time–frequency analysis approach to extract the event-related desynchronization/event-related synchronization patterns for classifying the motor imagery in EEG-based BCI applications [21]. Lemm et al. proposed the probabilistic modeling of sensorimotor µ rhythms to classify the imaginary hand movements in the BCI Competition 2003 [22]. Zhou et al. stated that by using the bispectrum-based feature extraction, all three classifiers (i.e., linear discriminant analysis (LDA), support vector machine (SVM) classifier, and neural network (NN) classifier) outperformed
the winner of the BCI Competition 2003 on the same Graz dataset [23, 18]. In [24], AAR parameters extracted by Kalman filtering were applied to four classifiers, i.e., minimum distance analysis, LDA, and k-nearest neighbor classifiers as well as SVM classifiers, to recognize the four-class motor imagery EEG data for the BCI Competition 2005. Other existing classification techniques, not currently used for BCI study, could be explored and may prove to be rewarding in online or offline analysis. Once BCIs are more widely used in clinical practice, new properties will have to be taken into consideration, such as the availability of large datasets or long-term variability of EEG signals.

Obtaining a sufficient number of training samples is time-consuming and costly in BCI operations, so the problem of efficient learning from a limited training set becomes increasingly important. SVM as a recent classification approach was used to address this issue within the framework of statistical learning theory [25]. The SVM classifier has been applied to a relatively large number of synchronous BCI problems [15, 17, 18, 23, 24], but SVM was very sensitive to the dynamic noise. The fuzzy support vector machine (FSVM) introduced a fuzzy membership to each training sample of SVM, so that different training samples could make different contributions to the learning of decision surface. FSVM therefore enhances SVM in reducing the effect of outliers and noises in training samples [26, 27]. EEG-based BCI features are often noisy and likely to contain outliers because EEG signals have a poor signal-to-noise ratio (SNR). The noise and artifacts due to electromyographic (EMG) activity (from scalp and facial muscles) and electrooculographic activity (from eye movements) are a serious problem for pattern recognition of EEG signals [28–30]. In this study, a comparison of two classification approaches, i.e., SVM and FSVM, was presented when wavelet-based features are used as classifier inputs. SVM and FSVM classifiers were tested on the Graz BCI datasets used in the BCI Competitions of 2003 and 2005.

The error rate was the most commonly used criterion to evaluate the performances of different methods in pattern recognition. The error rate considered only the signs of the classifier outputs, not the degrees of memberships of patterns belonging to each class, so the error rate measure provided only the classification accuracy of the classifier, not the confidence information of the classifying result [31]. To combine classification accuracy and confidence, in the BCI Competition 2003 on the Graz dataset, the entropy-based mutual information (MI) obtained from classifying results was used as the criterion to compare the performances of different methods. Greater MI of classifying results indicated that the classifier produced results with higher confidence. The time delay in the online analysis must also be considered. Although the time delay did not matter in the offline analysis, it became important for fast and accurate online feedback. Therefore, in the BCI Competition 2005 on the Graz dataset, the time course of the MI was of interest. The maximum increase of MI was used for validating the performances of different classifiers [32]. In the present study, the classifying results obtained by the SVM and FSVM classifiers using the wavelet-based features were compared with the ones achieved by the BCI Competition-winning methods [18, 32] on the same Graz datasets from the BCI Competitions of 2003 and 2005 in terms of the competition criteria of the MI or the maximum increase of MI, whereas the criteria of misclassification rate and classification time were selective measures of classifier performance.

2. Graz dataset of BCI Competitions

The dataset III of BCI Competition 2003 and the dataset IIIB of BCI Competition 2005 provided by Dr. Gert Pfurtscheller from Graz University of Technology were used to evaluate the classification performance of classifiers [33, 34]. All EEG data were collected during a motor imagery task. The subjects were asked to control a visual feedback setting by means of imagery left or right movements. The order of left and right cues was random.

In BCI Competition 2003 the Graz dataset was recorded over C3, Cz and C4 (Fig. 1b) from a normal female subject during a feedback session. The EEG was sampled at a 128-Hz sampling rate and was filtered between 0.5 and 30 Hz. Each trial (Fig. 1a) lasted 9 s. At t = 3 s, a left or right arrow was displayed as a cue, and the subject was, at the same time, asked to do motor imagery along the direction of the cue. The experiment consisted of 7 runs with 40 trials each. In the available 280 trials, 140 labeled trials were used to train the classifiers, whereas the other 140 trials are now available in [33] for testing the generalization performance of the trained classifiers.

In the Graz dataset of BCI Competition 2005 the EEG data were recorded over the positions C3, Cz, and C4 (Fig. 1b) from three subjects (O3, S4, and X11) during a motor imagery task [34]. The duration of each trial was 8 s, with a random interval between trials from 1 to 2 s. At t = 3 s, a left or right arrow was presented as a visual cue, and the subject was, at the same time, asked to do motor imagery along the cued direction. The EEG was sampled with 125 Hz and filtered between 0.5 and 30 Hz. The BCI experiment consisted of three sessions for each subject. Each session consisted of 4–9 runs with 40 trials each. The trials from each subject for training and testing were randomly chosen. The distribution of available trials in the training and testing dataset for three subjects is summarized in Table 1.

<table>
<thead>
<tr>
<th>Subject (feedback setting)</th>
<th>Training dataset</th>
<th>Test dataset</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>O3 (virtual environment)</td>
<td>242</td>
<td>159</td>
<td>401</td>
</tr>
<tr>
<td>S4 (basket)</td>
<td>540</td>
<td>540</td>
<td>1080</td>
</tr>
<tr>
<td>X11 (basket)</td>
<td>540</td>
<td>540</td>
<td>1080</td>
</tr>
</tbody>
</table>

Table 1 Distribution of trials in the training and test datasets from the 2005 BCI Graz dataset. The Graz dataset was generated by different experimental settings and three subjects (O3, S4 and X11). The trials from each subject for training and testing were randomly chosen.

<table>
<thead>
<tr>
<th>Subject (feedback setting)</th>
<th>Training dataset</th>
<th>Test dataset</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>O3 (virtual environment)</td>
<td>242</td>
<td>159</td>
<td>401</td>
</tr>
<tr>
<td>S4 (basket)</td>
<td>540</td>
<td>540</td>
<td>1080</td>
</tr>
<tr>
<td>X11 (basket)</td>
<td>540</td>
<td>540</td>
<td>1080</td>
</tr>
</tbody>
</table>

Table 1

The error rate was the most commonly used criterion to evaluate the performances of different methods in pattern recognition. The error rate considered only the signs of the classifier outputs, not the degrees of memberships of patterns belonging to each class, so the error rate measure provided only the classification accuracy of the classifier, not the confidence information of the classifying result [31]. To combine classification accuracy and confidence, in the BCI Competition 2003 on the Graz dataset, the entropy-based mutual information (MI) obtained from classifying results was used as the criterion to compare the performances of different methods. Greater MI of classifying results indicated that the classifier produced results with higher confidence. The time delay in the online analysis must also be considered. Although the time delay did not matter in the offline analysis, it became important for fast and accurate online feedback. Therefore, in the BCI Competition 2005 on the Graz dataset, the time course of the MI was of interest. The maximum increase of MI was used for validating the performances of different classifiers [32]. In the present study, the classifying

<table>
<thead>
<tr>
<th>Subject (feedback setting)</th>
<th>Training dataset</th>
<th>Test dataset</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>O3 (virtual environment)</td>
<td>242</td>
<td>159</td>
<td>401</td>
</tr>
<tr>
<td>S4 (basket)</td>
<td>540</td>
<td>540</td>
<td>1080</td>
</tr>
<tr>
<td>X11 (basket)</td>
<td>540</td>
<td>540</td>
<td>1080</td>
</tr>
</tbody>
</table>

Table 1 Distribution of trials in the training and test datasets from the 2005 BCI Graz dataset. The Graz dataset was generated by different experimental settings and three subjects (O3, S4 and X11). The trials from each subject for training and testing were randomly chosen.

<table>
<thead>
<tr>
<th>Subject (feedback setting)</th>
<th>Training dataset</th>
<th>Test dataset</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>O3 (virtual environment)</td>
<td>242</td>
<td>159</td>
<td>401</td>
</tr>
<tr>
<td>S4 (basket)</td>
<td>540</td>
<td>540</td>
<td>1080</td>
</tr>
<tr>
<td>X11 (basket)</td>
<td>540</td>
<td>540</td>
<td>1080</td>
</tr>
</tbody>
</table>

Table 1 Distribution of trials in the training and test datasets from the 2005 BCI Graz dataset. The Graz dataset was generated by different experimental settings and three subjects (O3, S4 and X11). The trials from each subject for training and testing were randomly chosen.
3. Method

3.1. Feature extraction using discrete wavelet transform

In the present study, we developed a wavelet-based time–frequency scheme for classifying EEG-based motor imagery tasks. The discrete wavelet transform (DWT) analyzed the signal at different frequency bands with different resolutions by decomposing the signal into a coarse approximation and detail information [11–12]. In the procedure of multi-resolution decomposition of a signal \( x[n] \), each stage consisted of two digital filters and two down-samplers by 2. The down-sampled outputs of first high-pass and low-pass filters provided the detail, D1 and the approximation, A1, respectively. The first approximation, A1, was further decomposed and this process was continued.

Selection of the suitable wavelet and the number of decomposition levels was very important in the analysis of EEG signals using the DWT. The number of decomposition levels was chosen based on the dominant frequency components of the EEG signal. Table 2 presents frequencies corresponding to different levels of wavelet decomposition on the EEG signals with a 128-Hz sampling rate from the Graz dataset of BCI Competition 2003. Table 3 describes frequency ranges corresponding to different levels of wavelet decomposition on the EEG signals with a 125-Hz sampling rate from the Graz dataset of BCI Competition 2005. The sub-bands D2 (16–32 Hz or 15.62–31.25 Hz) and D3 (8–16 Hz or 7.81–15.62 Hz) mostly included the \( \beta \) rhythm (18–26 Hz) and the \( \mu \) rhythm (8–12 Hz) respectively, whereas the sub-bands D4 (4–8 Hz or 3.91–7.81 Hz) and A4 (0–4 Hz or 0–3.91 Hz) would represent the \( \theta \) rhythm (5–7 Hz) and the \( \delta \) rhythm (0–4 Hz) separately. The \( \mu \) rhythm (8–12 Hz) and \( \beta \) rhythm (18–26 Hz) originating in the sensorimotor cortex were considered to be good signal features for EEG-based BCIs, so the number of decomposition levels was chosen to be 3. Fig. 2 shows the approximation (A3) and details (D1–D3) of an EEG signal taken from the C3 electrode location during left and right motor imagery. The wavelet coefficients of D1 (32–64 Hz or 31.25–62.5 Hz) could be eliminated because all EEG recordings were previously filtered between 0.5 and 30 Hz in the Graz BCI datasets. The features based on the sub-bands D2 and D3 reflecting the \( \beta \) and \( \mu \) rhythms, respectively, were used for the classification of EEG signals. The tests were usually done with different types of wavelets, and the one that gave maximal efficiency was selected for the particular application.

3.2. FSVM classification

3.2.1. SVM classifier

The SVM proposed by Vapnik has been studied extensively for classification and regression [25]. A SVM classifier uses a discriminant hyperplane to identify classes. The selected hyperplane is the one that maximizes the margin, i.e., the distance from the nearest training points (see Fig. 3). Maximizing the margin is known to increase the generalization capabilities. Training the SVM is a quadratic optimization problem. The construction of a hyperplane \( w^T x + b = 0 \) (\( w \) is the weight vector, \( b \) is an offset parameter) was achieved so that the margin between the hyperplane and the nearest point was maximized (see Fig. 3). The only free parameter, \( C \), in

![Fig. 2. Approximation and detailed coefficients of Daubechies order four-wavelet decomposition on an EEG signal taken from the C3 electrode location during left motor imagery (a) and right motor imagery (b).](image-url)
SVMs controlled the trade-off between the maximization of margin and the amount of misclassifications.

Such a SVM enabled classification using linear decision boundary and is known as a “linear SVM”. A nonlinear decision boundary and the amount of misclassifications.

SVMs controlled the trade-off between the maximization of margin like the classical SVM classifier, while treating the noises or outliers with less importance and setting lower fuzzy membership for these points to prevent noisy data points from making a narrower margin. This equipped the FSVM classifier with the ability to train data with noises or outliers by setting lower fuzzy memberships to the data points that are considered as noises or outliers with higher probability [26,27].

Let us consider a two-class classification task with training data points \((x_1, y_1, s_1), \ldots, (x_l, y_l, s_l)\). For each training data point \(x_i \in \mathbb{R}^N\), a label \(y_i \in \{-1, 1\}\) and a fuzzy membership \(s_i \in [0, 1]\) were presented with \(i = 1, \ldots, l\), and sufficient small \(\sigma > 0\), because the data point with \(s_i = 0\) meant nothing and could be removed from training set without affecting the result of optimization.

Because the fuzzy membership \(s_i\) was the attitude of the corresponding data point \(x_i\) toward one class and the parameter \(\xi_i\) could be viewed as a measure of error in the SVM, the term \(s_i \xi_i\) was a measure of error with different weighting. The optimal hyperplane problem was then regarded as the solution to

\[
\begin{aligned}
\text{minimize} & \quad \frac{1}{2} \mathbf{w} \cdot \mathbf{w} + C \sum_{i=1}^{l} s_i \xi_i \\
\text{subject to} & \quad y_i \left( \mathbf{w} \cdot \mathbf{x}_i + b \right) \geq 1 - \xi_i, \quad i = 1, \ldots, l, \\
\xi_i & \geq 0, \quad i = 1, \ldots, l,
\end{aligned}
\]

where \(C\) was a constant. A smaller \(s_i\) reduced the effect of the parameter \(\xi_i\) in Eq. (4) such that the corresponding data point \(x_i\) was treated with less importance. By using the Kuhn–Tucker conditions and lagrangian theorem, the optimal hyperplane could be found by solving the quadratic programming problem.

To choose the appropriate fuzzy memberships in a given problem was very important in the design of FSVM classifiers. The rule to assign appropriate membership values to data points could depend on the relative importance of data points to their own classes. In this work, to fuzz the training set efficiently, the training set \(S\) was initially divided into two sets: the positive training set \(S^+\) with \(y_i = 1\) and the negative training set \(S^-\) with \(y_i = -1\). The density \(\rho(x_0)\) of the point \(x_0\) was defined as the number of data points in its neighborhood, that is

\[
\rho(x_0) = N(X) = N(\{|x-x_i| \leq T\})
\]

where \(|x-x_i|\) denotes the Euclidean distance, and \(N(X)\) indicates the cardinality of the set \(X\). \(T\) is the threshold depending on the distance between two class \((S^+\) and \(S^-\)\) centers. Then the positive density \(\rho^+(x_0)\) and the negative density \(\rho^-(x_0)\) were respectively defined by

\[
\begin{align*}
\rho^+(x_0) &= \begin{cases} N(\{|x-x_i| \leq r \cdot d, x \in S^+\}) & \forall x_i \in S^+ \\ N(\{|x-x_i| \leq r \cdot d, x \in S^-\}) & \forall x_i \in S^- \end{cases} \\
\rho^-(x_0) &= \begin{cases} N(\{|x-x_i| \leq r \cdot d, x \in S^+\}) & \forall x_i \in S^- \\ N(\{|x-x_i| \leq r \cdot d, x \in S^-\}) & \forall x_i \in S^+ \end{cases}
\end{align*}
\]

where \(d\) was the distance between two class \((S^+\) and \(S^-\)\) centers, and \(r\) was a predetermined fractional multiplier corresponding to the threshold \(T\) in Eq. (5). It appeared that \(\rho^+(x_0)\) indicated the attitude of the point \(x_i\) to its own class in the training set. In principle, both \(\rho^+(x_0)\) and \(\rho^-(x_0)\) of the data point with outliers were small, and for the normal data points without noise and outliers, \(\rho^+(x_0)\) was large and \(\rho^-(x_0)\) was small. For the data points with noise, \(\rho^+(x_0)\) was small and \(\rho^-(x_0)\) was large. Therefore, the assigned membership value of data point \(x_0\) could be calculated by the membership
function:
\[ s_i = \frac{\rho^r(x_i)}{\rho_i^+ + \rho_i^-} \]
where \( \rho^+ \) is the maximum of \( \rho^r(x_i) \) for all training data points.

### 3.3. Parameter selection

#### 3.3.1. Wavelet function and decomposition level

The number of decomposition levels chosen was 3 (described in Section 3.1). To select an appropriate wavelet for the feature extraction of EEG signals, several well-known wavelet functions were tested on the Graz dataset of BCI Competition 2003: Symlet (sym2), Daubechies (db4), Biorthogonal (Bior3.1), Coiflet (coif3). Because the electrode Cz showed its independence of the motor imagery, only the electrodes C3 and C4 were used for feature extraction. The feature vectors obtained by the wavelet coefficients in the sub-bands D2 and D3 reflecting the \( \beta \) and \( \mu \) rhythms, respectively, were fed into the SVM classifier with RBF kernel functions, while the kernel parameter \( \sigma \) of 1.5 and the trade-off parameter \( C \) of 2 were chosen [25]. Classification accuracies were obtained by 140 testing trials. The number of support vectors, and the margin as well as the classification accuracy obtained by the SVM classifier with different wavelet functions, were shown in Table 4. Fewer support vectors would reduce the calculation time of the SVM classification. A larger margin indicated better generalization ability for the SVM classifier. The “db4” wavelet gave higher classification accuracy and lower computation cost than the others. Therefore, the wavelet coefficients were computed using the “db4” in this study.

#### 3.3.2. Classifiers

For training the SVM and FSVM classifiers with the RBF kernel functions, one had to predetermine the kernel parameter \( \sigma \) and the trade-off parameter \( C \) that could be chosen based solely on training data [25]. We firstly used the original algorithm of SVM to get the optimal \( \sigma \) value in Eq. (3) and the trade-off parameter \( C \). The parameters were sought in the two-dimension grid by \( C = (1, 2, 10, 50, 100, 1000) \) and a set of 100 equidistant values of \( \sigma \) between 1 and 5. A low fraction of support vectors was used as the criterion for choosing \( C \) and \( \sigma \) [25]. The fraction of support vectors as a function of \( \sigma \) values for the training trials in the Graz dataset of BCI Competition 2003, is shown in Fig. 4, where the typical “db4” wavelet was used to decompose the EEG signals to three levels. For too low values of \( C \), a large fraction of support vectors was obtained. Such a solution was less parsimonious and required more computations for evaluating the decision function. The chosen example values of \( C \) between 50 and 100 should provide good and parsimonious solutions. The minimal fraction of support vectors obtained was 0.41 corresponding to \( C = 50, \sigma = 0.2 \). Therefore, the RBF kernel parameter \( \sigma \) of 0.2 and the trade-off parameter \( C \) of 50 were chosen for the RBF kernel function in the SVM and FSVM classifiers.

The value of the tuning membership parameter \( r \) in Eqs. (6) and (7) for the FSVM classifier using the RBF kernel function \( C = 50, \sigma = 0.2 \) was searched in the set of \( r = (1, 1/2, 1/4, 1/8, 1/16) \) based solely on training data. We selected the value of \( r \) associated with the minimal number of support vectors. The minimal number of support vectors obtained was 69 (49% of the whole training dataset) corresponding to \( r = 1 \).

### 3.4. Description of the classification procedure

Based on the definitions in Section 3.1, features were extracted at every sampling point, with the sliding window size of 2 s to capture the rich frequency information in the EEG signals from the Graz dataset of BCI Competitions. The procedure of the classifiers with wavelet-based feature extraction could be divided into two parts; the first one performing the optimization of the classifiers based on the training trials and the second one being for the analysis of testing trials. The general flow chart was shown in Fig. 5, where the left of the flow chart was for the offline training of classifiers and the right was for the classification of the testing trials.
The procedure of online classification for the motor imagery task with the wavelet-based features was as follows:

1. We selected the same wavelet function and decomposition level as the training trials;
2. The DWT of EEG signals was conducted on a time window with a window shift of one sample, and the result was assumed for the hind boundary time of the window;
3. We constructed the time-varying feature vector of electrodes C3 and C4 for every trial using the statistics over the set of wavelet coefficients, and normalized each feature vector to an interval of [0,1];
4. We calculated the time-varying signed distance function \( D(x) \) in Eq. (2) of each trial, and classified each trial by the SVM or FSVM classifier, using the optimal parameters (described in Section 3.3.2), with a result denoted by \( f_t(x) \) belonging to \(-1,1\);
5. We evaluated the classification performance of the classifier on the testing dataset in terms of BCI Competition criteria.

4. Results and discussion

The experimental results were presented on a benchmark EEG dataset which was used in the BCI Competition 2003, and only the data between \( t = 3 \) and \( 9 \) s were used. The results obtained by the SVM and FSVM classifiers using the wavelet-based features were compared with those of the competition winners and other similar studies. To combine classification accuracy and confidence, on the Graz dataset in BCI Competition 2003, the MI was used as the criterion to compare the performances of different methods. The classification time was defined as the time when the MI is a maximum. Table 5 ranks the performances of the SVM and FSVM classifiers, the BCI Competition 2003 winning methods and the methods in [23] in terms of the MI criterion. Fig. 6 shows the MI time courses of the SVM and FSVM classifiers using the proposed wavelet-based features. The increase of MI indicates an increase in separation ability between left- and right-hand motor imagery.

In terms of the criterion of BCI Competition 2003, the FSVM and SVM classifiers based on the proposed features achieved the maximal MI 0.66 and 0.65, respectively, both of which were greater than the one achieved by [23] and the first winner of the BCI Competition 2003 on the Graz dataset. The proposed EEG-based mental task classification system combining the wavelet-based features with the FSVM classifier outperformed the SVM classifier with the wavelet-based feature extraction, and the winning methods of BCI Competition 2003 together with the methods in [23] on the Graz dataset. This demonstrated the effectiveness of the proposed FSVM classifier.

Another benchmark EEG dataset in the BCI Competition 2005 was used to validate the proposed approaches for classifying the motor imageries. The dataset was generated by different experimental settings and three subjects (O3, S4 and X11). Only the data between \( t = 3 \) and \( 8 \) s were used. Because the time delay is important for the online analysis, on the Graz dataset in BCI Competition 2005, the maximal increase of MI (maximum steepness calculated as MI(t)/|t - 3 s| for \( t > 3.5 \) s) was used for measuring the performance of different classification methods. To avoid a stimulus–response mechanism, \( t > 3.5 \) s will be evaluated. The “steepness” of MI quantifies the response time [32,34].

The kernel function and the optimal values of tuning parameters in the SVM and FSVM classifiers remain the same as the ones in the experiments described above based on the Graz dataset of BCI Competition 2003. Table 6 ranked the performances of the proposed SVM and FSVM classifiers, together with the BCI Competition 2005 winning methods, according to the averaged maximal MI steepness of all subjects on the Graz dataset of BCI Competition 2005. The FSVM and SVM classifiers, based on the proposed features, achieve the maximal MI steepness 0.6711 and 0.6051 on the EEG signals from the subject O3, respectively, both of which are much greater than the one achieved by the winners of the BCI Competition 2005. The FSVM classifier can yield a better classification performance on the subject O3 than the SVM classifier. Table 6 shows that the FSVM classifier could not ensure that all subjects obtained the best results owing to the difference between subjects.

The classification performance depends on the selection of several parameters, i.e., the wavelet function and decomposition level for the feature extraction using DWT, the kernel parameter \( \sigma \) and the trade-off parameter \( C \) in SVM, and the membership parameter \( r \) in FSVM. In this study, all parameters have been chosen based solely on training EEG data recorded from the subject in the Graz dataset of BCI Competition 2003, and were not adaptively selected for each subject in the Graz dataset of BCI Competition 2005. From a physiological standpoint, no two people are identical, so EEG signals vary between subjects. The non-subject-based method with uniform parameter settings for all subjects, therefore, cannot accomplish an adaptive parameter optimization in the FSVM and SVM classifiers for each subject. The subject-based method should be taken into

![Fig. 6. The MI time courses of the SVM (top) and proposed FSVM (bottom) classifiers with the wavelet-based features; at t = 3 s the cue (left or right in random order) was presented.](image-url)

<table>
<thead>
<tr>
<th>Table 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ranking</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
</tbody>
</table>
consideration in future investigations to provide optimal (or near optimal) parameters in the proposed F SVM model for each subject. It is known that EEG signals with a poor SNR are generated from the human brain (which is a system with highly nonlinear dynamics), while the efficient learning from a limited training set becomes very important in BCI operation. We believe that the proposed F SVM classifier may outperform all existing approaches that solve the specific task on the Graz dataset of BCI Competitions. The key advantages of the proposed F SVM method are detailed below.

(1) The F SVM classifier is robust to dynamic noise by setting lower fuzzy membership for the data points with the noises or outliers, whereas the SVM classifier is sensitive to the noise. Some data points with outliers and noises could be support vectors in the SVM classifier. In users who eventually acquire EEG control, early target-related EMG contamination may be the most prominent for unsuccessful trials. The F SVM classifier is probably the most appropriate classifier to deal with artifacts present during the new user’s initial BC I training sessions.

(2) The proposed F SVM classifier has lower computation cost than the SVM classifier due to the fewer support vectors in Eq. (2), which is critical for online classification in effective BCI operations.

The present study has shown that the proposed F SVM classifier is very effective for identifying different mental tasks from EEG signals.

5. Conclusion

This contribution presented a new application of the F SVM classifier for classifying EEG-based left and right motor imagery, while the feature vectors calculated by the wavelet coefficients of the EEG signals in two sub-bands reflecting the and rhythms were used as inputs. F SVM and SVM classifiers with the wavelet-based feature extraction outperform the winners of BCI Competition 2003 and other similar studies on the same Graz dataset in terms of the MI criterion, whereas the F SVM classifier performs better than the SVM approach. The F SVM classifier with the proposed features ranked second behind the first winning method of the BCI Competition 2005 in terms of mean maximal MI steepness in three subjects of Graz dataset, whereas F SVM and SVM classifiers perform much better than the first winning method on the EEG signals from the subject 03 according to the maximal MI steepness criterion. The F SVM classifier outperforms the SVM method. Therefore, the F SVM classifier is a promising method and provides a new way for the online classification of EEG signals in BCIs.

An even better approach would be to use the obtained model online in a subsequent closed-loop experiment. During the procedure of training SVM, most of the computational effort was spent in solving the quadratic programming problem in order to find the support vectors. Because there is only one free parameter in SVM while the number of free parameters for F SVM is equivalent to the number of training data points, one difficulty encountered in this research was that the training time of F SVM was much more than that of SVM due to the extra parameters. Reducing the computation cost for training the classifier, while moving the proposed F SVM model from the offline classification to online BCI applications, merits further study.

Acknowledgements

The authors would like to thank Graz University of Technology for providing their data. This study has been supported by the Natural Science Foundation of China (grant numbers 60874035 and 60674105), in part by the Hi-tech Research and Development Program of China (2007AA042204 and 2008AA042207), in part by the Key Laboratory of Image Processing and Intelligent Control (200704) and in part by the Natural Science Foundation of Hubei Province (2007ABA027).

Conflict of interest

There is no financial and personal relationships with other people or organisations that could inappropriately influence (bias) our work in this paper.

References


| Table 6
<table>
<thead>
<tr>
<th>Ranking</th>
<th>Methods</th>
<th>Maximal MI steepness (bit/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BCI,Comp2005,1st winner</td>
<td>0.1698</td>
</tr>
<tr>
<td>2</td>
<td>SVM with wavelet-based features</td>
<td>0.0711</td>
</tr>
<tr>
<td>3</td>
<td>SVM with wavelet-based features</td>
<td>0.6051</td>
</tr>
<tr>
<td>4</td>
<td>BCI,Comp2005,2nd winner</td>
<td>0.1626</td>
</tr>
<tr>
<td>5</td>
<td>BCI,Comp2005,3rd winner</td>
<td>0.2030</td>
</tr>
</tbody>
</table>

864 Q. Xu et al. / Medical Engineering & Physics 31 (2009) 858–865


