A comparative study of surface EMG classification by fuzzy relevance vector machine and fuzzy support vector machine

To cite this article: Hong-Bo Xie et al 2015 Physiol. Meas. 36 191

View the article online for updates and enhancements.

Related content
- Use of the discriminant Fourier-derived cepstrum with feature-level post-processing
  Xinpu Chen, Xiangyang Zhu and Dingguo Zhang
- Support vector machines to detect physiological patterns for EEG and EMG-based human–computer interaction: a review
  L R Quitadamo, F Cavrini, L Sbernini et al.
- Classification of surface EMG signals
  Gang Wang, Zhiguo Yan, Xiao Hu et al.

Recent citations
- A comparison of sEMG temporal and spatial information in the analysis of continuous movements
  Michela Di Girolamo et al
- Support vector machines to detect physiological patterns for EEG and EMG-based human–computer interaction: a review
  L R Quitadamo et al
- Classifying neuromuscular diseases using artificial neural networks with applied Autoregressive and Cepstral analysis
  Sabri Koçer and Abdullah Erdal Tümer
A comparative study of surface EMG classification by fuzzy relevance vector machine and fuzzy support vector machine

Hong-Bo Xie¹, Hu Huang², Jianhua Wu¹ and Lei Liu¹

¹ Jiangsu Provincial Key Laboratory for Interventional Medical Devices, Huaiyin Institute of Technology, Huaian, Jiangsu Province, 223003, People’s Republic of China
² Biomedical Informatics and Computational Biology, University of Minnesota, Minneapolis, MN 55414, USA

E-mail: xiehb@sjtu.org

Received 21 October 2014, revised 2 December 2014
Accepted for publication 3 December 2014
Published 9 January 2015

Abstract
We present a multiclass fuzzy relevance vector machine (FRVM) learning mechanism and evaluate its performance to classify multiple hand motions using surface electromyographic (sEMG) signals. The relevance vector machine (RVM) is a sparse Bayesian kernel method which avoids some limitations of the support vector machine (SVM). However, RVM still suffers the difficulty of possible unclassifiable regions in multiclass problems. We propose two fuzzy membership function-based FRVM algorithms to solve such problems, based on experiments conducted on seven healthy subjects and two amputees with six hand motions. Two feature sets, namely, AR model coefficients and room mean square value (AR-RMS), and wavelet transform (WT) features, are extracted from the recorded sEMG signals. Fuzzy support vector machine (FSVM) analysis was also conducted for wide comparison in terms of accuracy, sparsity, training and testing time, as well as the effect of training sample sizes. FRVM yielded comparable classification accuracy with dramatically fewer support vectors in comparison with FSVM. Furthermore, the processing delay of FRVM was much less than that of FSVM, whilst training time of FSVM much faster than FRVM. The results indicate that FRVM classifier trained using sufficient samples can achieve comparable generalization capability as FSVM with significant sparsity in multi-channel sEMG classification, which is more suitable for sEMG-based real-time control applications.
Keywords: fuzzy relevance vector machine, sparse kernel machines, electromyography, fuzzy logic, pattern classification

(Some figures may appear in colour only in the online journal)

1. Introduction

During a voluntary contraction of skeletal muscles, the electrical activity of activated motor units can be detected by surface electrodes (Xie and Wang 2006). The resulting surface electromyographic (sEMG) signal is the summation of motor unit action potentials discharged by muscle fibers near the recording electrodes, and contains rich information on motor unit recruitment, firing, motion intention, and general physiological state of the neuromuscular system. sEMG pattern classification has been widely used in prosthetic hand and exoskeletal control, functional electrical stimulation devices, and other human–machine interface (HMI) control for the elderly, amputees, and those with various neuromuscular disorders (Yu et al 2002, Ahsan et al 2009, Khokhar et al 2010, Hung et al 2012).

Artificial intelligence and machine learning play important roles in sEMG pattern recognition, with many techniques based on these having been explored for the control of sEMG-based HMI. In the first pattern recognition-based prosthetic hand control schemes developed in 1970s, simple statistical classifiers were used to recognize hand motions from amplitude-based features, achieving about 75% accuracy in a four-class sEMG classification problem (Englehart and Hudgins 2003). This accuracy was then improved by using two artificial neural network (NN) classifiers, namely, a discrete Hopfield NN and a multi-layer perceptron (MLP), as described by Kelly et al (1990). Hudgins et al (1993) successfully applied the MLP NN, trained by a standard back propagation (BP) algorithm, to develop a real-time sEMG pattern control system with approximately 10% error rate in classifying four types of upper limb motion. Since sEMG signals are non-stationary and noisy, varying even they are belonging to the same motion, BP-based NNs are not able to achieve high learning and discrimination performance (Fukuda et al 2003, Xie and Wang 2006). Several other NN-based machine learning methods, such as radial basis functions network (Chaiyaratana et al 1996), time-delayed artificial NN (Au and Kirsch 2000) and self-organizing feature map (Eom et al 2002, Chu et al 2006), have also been evaluated for their applicability to sEMG classification.

sEMG signals are not always strictly repeatable, and may sometimes even be contradictory due to shift of electrodes, sweat, and muscular fatigue (Chan et al 2000). Since one of the most useful properties of fuzzy logic systems is that contradictions in the data can be tolerated, fuzzy logic systems are advantageous in sEMG signal classification. Compared with a MLP network, several fuzzy logic approaches have shown improved accuracy and robustness to noise in sEMG classification (Chan et al 2000, Kiguchi et al 2004, Ajiboye and Weir 2005).

NNs exhibit some problems inherent to their architecture, such as overtraining, overfitting, and the large number of controlling parameters. Other problems relate to the reproducibility of results, due mainly to random initialization of the networks and variability in stopping criteria (Xie et al 2009b). Support vector machine (SVM) classification which is based on the idea of structural risk minimization, is a new technique that has drawn much attention in the field of biomedical engineering in recent years. The good generalization ability of SVM is achieved by finding a large margin between two classes. Performance of binary SVMs can match or exceed MLP and linear discriminant analysis when combined in an efficient manner to classify sEMG signals of hand/wrist motions (Oskoei and Hu 2008, Yan et al 2008a, 2008b).
Despite the fact that SVM classifiers provide improved performance over traditional learning machines, a number of significant and practical disadvantages exist (Tipping 2001). Although relatively sparse, the number of support vectors (SVs) typically grows linearly with the size of the training set, and hence, SVM makes unnecessarily liberal use of basis functions. For some specific algorithms such as least square support vector machine (LS-SVM), the number of SVs equals to that of training samples without sparseness. SVM does not directly provide probability estimates, and therefore is not suitable for classification tasks in which posterior probabilities of class membership are necessary. In addition, estimation of the regularizing parameter in SVM construction, which generally entails a cross-validation procedure, is wasteful of computational time and data. Finally, the SVM kernel function must satisfy Mercer’s condition, namely, it must be a continuous symmetric kernel of a positive integral operator (Tipping 2001).

To overcome these problems of SVM efficiently, Tipping (2001) developed a new kernel based machine learning technique, termed relevance vector machine (RVM). The RVM shares many of the characteristics of SVM while avoiding its principal limitations. It uses the sparse Bayesian learning framework, in which a priori parameter structure is placed based on automatic relevance determination theory for removing irrelevant data points (MacKay 1992). Hence, it produces sparse models, as well as a comparable generalization performance to that of SVM. Most importantly, RVM classification requires dramatically fewer relevance vectors (RVs) compared with the number of SVs for SVM classification. This can significantly reduce the computational cost, making RVM more suitable for real-time applications (Majumder et al 2005, Williams et al 2005, Demir and Ertürk 2007, Wang et al 2009). Many sEMG-based control including prosthetic hand and exoskeletal control, as well as wheelchair and robotic control need to be performed in real-time (Englehart and Hudgins 2003, Fukuda et al 2003, Chu et al 2006, Ahsan et al 2009, Khokhar et al 2010, Hung et al 2012). RVM is thus a potentially promising tool to classify sEMG patterns. Similar to SVM, the original RVM is a binary classifier. As for multi-class recognition, several coding schemes have been proposed using binary classifiers (Tipping 2001, Oskoei and Hu 2008). However, indecisive regions often exist when a binary RVM classifier ensemble is used to accommodate a multi-class problem (explained in the next section). In order to solve for unclassifiable regions in RVM, we define two membership functions in a direction perpendicular to the optimal hyperplane that separates the pair of classes. Correspondingly, we construct fuzzy support vector machines (FSVMs) based on least squares algorithm (Yan et al 2008a, 2008b). The performance of the proposed fuzzy relevance vector machines (FRVMs) and FSVMs in classification of sEMG signals is widely compared in terms of classification accuracy, sparsity, training time, test delay, and the effect of training sample size.

2. Methods

In this section, we first introduce the basic RVM for binary classification. We then outline multi-class RVM and describe the FRVM method to avoid potential indecisive regions in multi-class RVM. The construction of FSVM is briefly presented. The experimental protocol and feature extraction are also described.

2.1. Binary RVM

As a sparse kernel technique, the central idea of RVM is to map a set of inputs to a high-dimensional feature space through kernel functions, providing posterior probabilistic outputs
of the class membership for constructing decision boundaries. The compelling feature of RVM is that it utilizes dramatically fewer kernel functions, whilst its generalization performance is comparable to the equivalent SVM (Tipping 2001). Given a data set \( \{ x_n, t_n \}_{n=1}^N \) where \( x_n \) denotes the input to be classified and \( t_n \) represents its class label, we write the targets as a vector \( t = (t_1, \ldots, t_N)^T \), and express it as the sum of an approximation vector \( \hat{y} = (y(x_1), \ldots, y(x_N))^T \) and an ‘error’ vector \( e = (\epsilon_1, \ldots, \epsilon_N)^T \):

\[
t = y + e = \Phi w + e,
\]

where \( w = (w_1, \ldots, w_M)^T \) is a ‘weight’ parameter vector and \( \Phi = [\Phi_1 \ldots \Phi_M] \) is an \( N \times M \) design matrix whose columns comprise the complete set of \( M \) ‘basis vectors’. Applying the logistic sigmoid link function \( \sigma(y) = (1 + e^{-y})^{-1} \) to \( y(x) \), and adopting the Bernoulli distribution, the likelihood of the complete data set can be represented as

\[
P(t \mid w) = \prod_{n=1}^N \sigma[y(x_n; w)]^{t_n} [1 - \sigma(y(x_n; w))]^{1-t_n},
\]

where the targets \( t_n \in \{0,1\} \). To control the complexity of the model and avoid over-fitting, a zero-mean Gaussian prior distribution is defined over \( w \).

\[
p(w \mid \alpha) = \prod_{i=0}^N \mathcal{N}(w_i \mid 0, \alpha_i^{-1}) = \prod_{i=0}^N \sqrt{\frac{\alpha_i}{2\pi}} \exp\left( -\frac{\alpha_i w_i^2}{2} \right),
\]

with \( \alpha = [\alpha_0, \alpha_1, \ldots, \alpha_N]^T \) a vector of \( N+1 \) hyperparameters. An individual hyperparameter is associated independently with every weight, moderating the strength of the prior, with the hyperparameter itself having a Gamma prior. The parameter \( \alpha \) for each \( w \) is intuitively called the ‘relevance’ of that feature, in the sense that the bigger \( \alpha \), the more likely the feature weight \( w \) is driven to zero. However, the weights \( w \) cannot be integrated out analytically, precluding closed-form expressions for either the weight posterior \( p(w \mid t, \alpha) \) or the marginal likelihood \( P(t \mid \alpha) \). Thus, the Laplace approximation procedure is utilized as described below (Tipping 2001).

Since \( p(w \mid t, \alpha) \propto P(t \mid w) p(w \mid \alpha) \), finding the optimal weights is equivalent to finding the maximum of

\[
\log \{ P(t \mid w) p(w \mid \alpha) \} = \sum_{n=1}^N \left[ t_n \log y_n + (1 - t_n) \log (1 - y_n) \right] - \frac{1}{2} w^T A w,
\]

for the most probable weight \( w_{\text{MP}} \), with \( y_n = \sigma \{ y(x_n; w) \} \) and \( A = \text{diag}(\alpha) \) for the current values of \( \alpha \). This represents a penalized logistic log-likelihood function and it requires iterative maximization, using an iterative reweighted least squares algorithm to find \( w_{\text{MP}} \).

To carry out the iterative procedure, we require the gradient vector and Hessian matrix of the log posterior distribution, which can be found by differentiating twice:

\[
\nabla_w \log p(w \mid t, \alpha)|_{w_{\text{MP}}} = \Phi^T (t - y) - A w,
\]

\[
\nabla^2_w \log p(w \mid t, \alpha)|_{w_{\text{MP}}} = - (\Phi^T B \Phi + A),
\]

where \( B \) is an \( N \times N \) diagonal matrix with elements \( b_n = y_n(1 - y_n) \), with vector \( y = (y_1, \ldots, y_N)^T \), and \( \Phi \) the design matrix with elements \( \Phi_{ni} = \phi_i(x_n) \). The approximation to the posterior distribution corresponding to the mean of the Gaussian approximation is obtained by inverting equation (6). The mean and covariance of the Laplace approximation can now be given as

\[
\Sigma = (\Phi^T B \Phi + A)^{-1},
\]
Using the statistics $\Sigma$ and $w_{MP}$ of the Gaussian approximation, we can follow MacKay’s approach to update the hyperparameters $\alpha_i$ by

$$\alpha_i^{\text{new}} = \frac{1 - \alpha_i \Sigma_{ii}}{w_{MP}},$$

where $\Sigma_{ii}$ is the $i$th diagonal element of the covariance matrix. During the optimization process, many $\alpha_i$ will have large values, and thus the corresponding model weights will be pruned out, leading to sparse representation. Those samples remaining with $w_i \neq 0$ are termed relevance vectors, corresponding to support vectors in SVM.

The above training procedure is typically slow. In order to speed up the training process, Tipping and Faul (2003) proposed a highly accelerated learning algorithm in which a single hyperparameter $\alpha_i$ is fully optimized at each step. If we define

$$\hat{t} = \Phi w_{MP} + B^{-1}(t - y),$$

the approximate log marginal likelihood can be written in the form

$$L(\alpha) = \log p(t \mid \alpha, \beta) := -\frac{1}{2} \left\{ N \log(2\pi) + \log|C| + (\hat{t})^T C^{-1} \hat{t} \right\},$$

where

$$C = B + \Phi A \Phi^T.$$

Considering the dependence of $L(\alpha)$ on a single hyperparameter $\alpha_i$, $i \in \{1, 2, \cdots, M\}$, the contribution from $\alpha_i$ in the matrix $C$ is then factored out to give

$$C = C_{-i} + \alpha_i^{-1} \Phi_i \Phi_i^T,$$

where $C_{-i}$ is $C$ with the contribution of basis vector $i$ removed. Established matrix determinant and inverse identities may be used to write the relevant terms in $L(\alpha)$ as

$$|C| = |C_{-i}| \left| 1 + \alpha_i^{-1} \Phi_i^T C_{-i}^{-1} \Phi_i \right|,$$

$$C^{-1} = C_{-i}^{-1} - \frac{C_{-i}^{-1} \Phi_i \Phi_i^T C_{-i}^{-1}}{\alpha_i + \Phi_i^T C_{-i}^{-1} \Phi_i}.$$

Using these results, the log marginal likelihood function (equation (10)) can be written in the form

$$L(\alpha) = L(\alpha_{-i}) + \frac{1}{2} \left\{ \log \alpha_i - \log(\alpha_i + s_i) + \frac{q_i^2}{\alpha_i + s_i} \right\} = L(\alpha_{-i}) + \lambda(\alpha_i).$$

Here, two quantities are introduced

$$s_i = \Phi_i^T C_{-i}^{-1} \Phi_i,$$

$$q_i = \Phi_i^T C_{-i}^{-1} t,$$

where $s_i$ the sparsity and $q_i$ the quality of $\Phi_i$. A large value of $s_i$ relative to $q_i$ means that the basis vector $\Phi_i$ is more likely to be pruned from the model. The ‘sparsity’ measures the extent to which basis vector $\Phi_i$ overlaps with the other basis vectors in the model. The ‘quality’ represents a measure of alignment of basis vector $\Phi_i$ with the error between the training set values
t = (t_1, t_2, \cdots, t_n)^T and the vector y_i of predictions that would result from the model with the vector \( \phi_i \) excluded (Tipping and Faul 2003).

### 2.2. Multiclass RVM

The RVM was originally developed for solving regression and binary classification problems. However, most practical applications need to handle multi-class discrimination problems. Several techniques have been proposed to extend a binary classifier to multi-class problems, including one-against-all (OAA), one-against-one (OAO, also known as pairwise), and error-correcting-output code (ECOC) (Tipping and Faul 2003, Mianji and Zhang 2011). In the OAA scheme, a \( k \)-class problem is converted into \( k \) two-class problems and for the \( i \)th two-class problem, class \( i \) is discriminated from the remaining classes. As for multiclass RVM, Tipping and Faul (2003) adopted this scheme and extended the original RVM to a multi-class model using a generalized multinomial form of likelihood for equation (2). However, raising the number of classes would lead to a significant increase in OAA computational load. More importantly, OAA generally produces a poor result (Tipping and Faul 2003, Vong et al 2013) since it does not consider the pairwise correlation and hence creates several indecisive regions shown in figure 1(a).

In the OAO scheme, the \( k \)-class problem is converted into \( k(k-1)/2 \) two-class problems which cover all pairs of classes. For an unknown sample \( x \), the inferred discriminant function for \( i, j \) class pair is given by

\[
D_{ij}(x) = \Phi w_{ij}. \tag{19}
\]

The logistic sigmoid function can be applied here to transfer \( D_{ij}(x) \) into the probability of \( P(x, C_i) \) and \( P(x, C_j) \). In practice, a ‘Max Wins’ strategy is used in OAO decision process, which first calculates the score function

\[
D_i = \sum_{j \neq i, j=1}^{k} \text{sgn}(D_{ij}(x)), \tag{20}
\]

and classifies \( x \) into the class

\[
\text{arg max}_{i=1, \cdots, k} (D_i(x)). \tag{21}
\]

Although this method is more computational efficient, an unclassifiable region may also exist for OAO if equation (21) is satisfied by multiple \( i \)'s. For example, if the discriminant functions satisfy \( D_{12}(x)<0, D_{23}(x)<0, \) and \( D_{13}(x)>0 \), the test sample \( x \) in the central region of figure 1(b) is obviously unclassifiable.

### 2.3. Fuzzy relevance vector machine

In order to solve the unclassifiable regions in OAO, we can define a choice of two fuzzy membership functions to indicate the probability of \( x \) belonging to each class. Let \( m_{i,i}(x) \) and \( m_{j,j}(x) \) represent membership values of sample \( x \) belonging to class \( i \) and \( j \) from the binary classifier RVM_\( i \). We define the first fuzzy membership function (MF I) as

\[
m_{i,j}(x) = \begin{cases} 
1 & \text{if } D_{ij}(x) \geq 1 \\
D_{ij}(x) - 1 & \text{if } 0 < D_{ij}(x) < 1 \\
-1 & \text{otherwise}
\end{cases} \tag{22}
\]
Figure 1. Indecisive regions in OAA (a) and OAO (b) classification strategy, and fuzzy scheme to solve the indecisive region in OAO strategy (c).
A second alternative membership function (MF II) is defined as

$$m_{i,j}(x) = \begin{cases} 
-1 & D_j(x) \geq 1 \\
-D_j(x) & 1 \leq D_j(x) < 1 \\
1 & \text{otherwise}
\end{cases} \quad (23)$$

Using either of these membership functions, we then implement a minimum operator to determine $m_i(x)$, the membership function of $x$ belonging to class $i$ from all possible RVMs:

$$m_i(x) = \min_{j=1,\ldots,k} m_{ij}(x). \quad (26)$$

The input sample $x$ is then classified into the class

$$\arg \max_{i=1,\ldots,k} (m_i(x)). \quad (27)$$

Using fuzzy functions, the unclassifiable region problem in OAO strategy of RVM is now solved as indicated in figure 1(c).

2.4. Fuzzy support vector machine

The basic idea of SVM for binary classification is similar to RVM, mapping the data $x_i$ into a high dimensional feature space via a nonlinear mapping and then constructing the optimal hyperplane that realizes the maximal margin in this space. Detailed descriptions of SVM can refer to the general introductions to SVM (Vapnik 1982, Cristianini and Shawe-Taylor 2000, Suykens et al 2002), and tutorials on support vector classification (Burges 1998). The LS-SVM algorithm is applied as the basic binary classifier in this study for comparison (Yan et al 2008a, 2008b). The scheme to extend binary LS-SVM to multi-class is the same as RVM described before. Correspondingly, two FSVM classification strategies based on fuzzy membership I and II can be constructed respectively, and the latter has already been successfully used in our previous study (Yan et al 2008b).

2.5. Data acquisition and feature extraction

We evaluate FRVM and FSVM techniques by classifying experimental sEMG signals, aiming at real-time control of various HMIs. Four channels of sEMG were collected from seven healthy subjects and two amputees (four males and five females aged between 22–48 years). Figure 2 is the experimental setting when recording signals from an amputee. Six classes of hand/wrist motions to be classified were grasp (GR), hand open (OP), wrist flexion (WF), wrist extension (WE), ulnar deviation (UD) and radial deviation (RD) as shown in figure 3. For two amputees, they had been trained to image and complete the movements accompanied...
by the visual guidance. Human subject ethics approval was obtained from the relevant committee in Huaiyin Institute of Technology and informed consent was obtained from all subjects prior to the experiment. sEMG signals were acquired from the forearm using bi-polar Ag-AgCl electrodes (Dual electrode #272, Noraxon USA Inc. AZ, USA). Electrodes were placed on the extensor digitorum, the extensor carpi radialis, the palmaris longus and the flexor carpi ulnaris around the forearm. The distance between two adjacent surface electrodes was 2 cm. Relevant skin areas were abraded beforehand with alcohol. An additional Ag–AgCl electrode was placed on the elbow to provide a common ground reference. sEMG signals were amplified by a custom-made amplifier with a gain of 2000, filtered using a 8–500Hz band-pass analog filter within the amplifier, then digitized by a 12-bit data acquisition card (NI PCI-6024E, National Instruments, Austin, TX), with a sample frequency of 1kHz.

During the experiment, subjects were required to perform three repeated 60 s continuous contractions. Within each session, each limb motion was held in a random order for about 10 s. The participants were allowed to relax between each session to avoid fatigue. The first session was used as training set, whist the second and the third were the validation and test sets respectively. All the data was segmented into consecutive 256 ms with 75% overlapping rate.

Two sets of features were extracted from the recorded sEMG signals in the time and time-frequency (TF) domains. The first set was composed of forth-order autoregressive model coefficients and the root mean square value, referred to as AR-RMS (Hudgins et al 1993, Englehart et al 1999). For time-frequency features, a set of wavelet transform (WT) energy values at each scale using a Coiflet wavelet family with four levels (Coif 4) was extracted (Hudgins et al 1993, Englehart et al 1999). A majority vote strategy was then conducted to improve the accuracy (Englehart and Hudgins 2003).

3. Results

3.1. Effect of membership functions

Two membership functions (equations (22)–(25)) were used to train $6 \times (6 - 1) / 2 = 15$ binary classifiers of both RVMs and SVMs. Their outputs were then processed by the proposed fuzzy strategies to recognize six patterns. Figure 3 illustrates the typical classifier output for subject 3, degree of fuzzy membership for FRVM with MF I. Figure 4 shows the averaged classification accuracy and standard deviation of two membership functions across all subjects. When using AR-RMS features, FRVMs exhibited similar performance over all subjects for both
membership functions. The averaged accuracy using MF I was slightly higher than that of MF II. In the case of WT features, averaged accuracy of the two membership functions was similar, whilst MF II demonstrated higher standard deviation. From the practical perspective, the performance of MF I was superior, and thus it was employed in the subsequent FRVMs evaluations. On the other hand, the results were quite different in the FSVM case. When using MF I, classification accuracy was much worse than MF II for both feature sets, indicating that MF II was more matched to FSVMs. In other words, the performance of FSVM was more sensitive to the membership functions than that of FRVM.

3.2. Classification performance

Since averaged accuracies of RVMs with MF I and FSVMs with MF II were higher than other combinations, we further compared their performance in terms of each motion pattern, sparsity, training time, and test delay. Figure 5 shows the classification accuracy and the percentage of vectors used in training the above two classifiers for subject 2. In terms of classification accuracy, results were slightly worse for FRVMs, but were still acceptable for both AR-RMS and WT features. For FRVMs, only 3.65 and 3.93% of the vectors were used in training for AR-RMS and WT features respectively. However, a LS-SVM was used as a base binary classifier for FSVMs (Yan et al. 2008a, 2008b), which utilized 100% of vectors as support vectors without any sparseness.

The mean classification accuracy and standard deviation of each motion across all subjects after majority vote is indicated in figure 6. When using AR-RMS features, the total classification accuracy and standard deviation were similar for both classifiers, i.e., 93.22 ± 4% for FRVMs and 93.28 ± 7% for FSVMs. In the case of WT features, although the discrimination capability of FRVMs was worse than FSVMs for the last two motions (UD and RD), the average classification accuracy of the former (90.86 ± 6%) was comparative to that of the latter.
However similar to subject 2, the RVs used in the training of other subjects varied from 3 to 5%, much lower than the 100% SV proportion for FSVMs.

FRVMs utilize multiple binary classifiers to achieve their final decision. The processing delay for sEMG classification is another important issue, since the response time should not introduce a delay perceivable by the user for many sEMG-based real-time control systems (Cristianini and Shawe-Taylor 2000). For example, the time threshold for acquiring sEMG data plus the processing time for generating classified control commands for prosthetic hand control is typically regarded to be roughly 300 ms (Xie et al 2009). Testing times for FRVMs and FSVMs were empirically evaluated using a 2.2 GHz Intel-based computer, with computations performed in Matlab (Version 8.0, The Mathworks, Natick, MA) and matrix multiplications built-in functions. Table 1 illustrates testing times for both AR-RMS and WT features.

---

**Figure 4.** Averaged classification accuracy and standard deviation across all subjects obtained by FRVMs and FSVMs using two membership functions.

**Figure 5.** Classification accuracy (left) and percentage of vectors used in training (right) for subject 2 using FRVMs and FSVMs with both AR-RMS and WT features.
across all subjects. The mean processing time of FSVMs for a motion was about 120 ms for both AR-RMS and WT features. In some extreme cases, the response time was larger than 200 ms. Normally, the minimal epoch length for sEMG data is at least 128 ms for a 1000 Hz sampling rate (Parker et al 2006). In such cases, FSVMs are potentially unable to satisfy the real-time demand since the total time of signal acquisition, feature extraction, and FSVM recognition would exceed 300 ms. However, FRVMs could discriminate these motions in much less time than FSVMs, due to fewer RVs used in the discriminant function. These results demonstrated that FRVMs could achieve comparable sEMG classification accuracy in a more computationally efficient manner, more suitable for real-time applications. Table 2 is a

![Figure 6](image-url)  
**Figure 6.** Mean classification accuracy and standard deviation of each motion across all subjects using FRVMs with MF I and FSVMs with MF II.

**Table 1.** The processing delay (ms) of FRVMs and FSVMs with both AR-RMS and WT features for all subjects.

<table>
<thead>
<tr>
<th>Subject</th>
<th>FRVM-AR-RMS</th>
<th>FSVM-AR-RMS</th>
<th>FRVM-WT</th>
<th>FSVM-WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15.63</td>
<td>84.93</td>
<td>5.26</td>
<td>89.14</td>
</tr>
<tr>
<td>2</td>
<td>105.27</td>
<td>234.91</td>
<td>20.32</td>
<td>229.72</td>
</tr>
<tr>
<td>3</td>
<td>49.55</td>
<td>84.96</td>
<td>12.51</td>
<td>88.51</td>
</tr>
<tr>
<td>4</td>
<td>15.14</td>
<td>90.17</td>
<td>12.56</td>
<td>90.68</td>
</tr>
<tr>
<td>5</td>
<td>18.29</td>
<td>88.54</td>
<td>6.34</td>
<td>88.52</td>
</tr>
<tr>
<td>6</td>
<td>20.35</td>
<td>182.82</td>
<td>22.94</td>
<td>186.56</td>
</tr>
<tr>
<td>7</td>
<td>15.17</td>
<td>91.17</td>
<td>4.74</td>
<td>85.97</td>
</tr>
<tr>
<td>8a</td>
<td>41.22</td>
<td>153.66</td>
<td>19.16</td>
<td>166.32</td>
</tr>
<tr>
<td>9a</td>
<td>29.71</td>
<td>102.43</td>
<td>9.83</td>
<td>107.25</td>
</tr>
<tr>
<td>Mean ± Std</td>
<td>34.47 ± 29.36</td>
<td>123.46 ± 54.35</td>
<td>12.52 ± 6.81</td>
<td>125.83 ± 54.16</td>
</tr>
</tbody>
</table>

* The amputee.
summarization of training time for both FRVM and FSVM across all subjects, indicating that FRVM required substantial longer time than FSVM to converge in training stage to prune out unnecessary features or samples.

3.3. Effect of training sample size

Finally, we evaluated the effect of training sample size on the recognition performance using both methods. We set the test sample size of each movement as 90 while training sample sizes as 30, 60, and 90, respectively. Figure 7 shows the boxplot of classification results in cases of 30, 60, and 90 training samples versus 90 test samples for FRVM and FSVM with both AR-RMS and WT features. In the case of 30 training samples, the performance of FRVM with

<table>
<thead>
<tr>
<th>Subject</th>
<th>FRVM-AR-RMS</th>
<th>FSVM-AR-RMS</th>
<th>FRVM-WT</th>
<th>FSVM-WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.66</td>
<td>0.28</td>
<td>0.25</td>
<td>0.17</td>
</tr>
<tr>
<td>2</td>
<td>2.38</td>
<td>0.73</td>
<td>55.08</td>
<td>0.71</td>
</tr>
<tr>
<td>3</td>
<td>30.25</td>
<td>0.25</td>
<td>1.00</td>
<td>0.25</td>
</tr>
<tr>
<td>4</td>
<td>0.38</td>
<td>0.25</td>
<td>136.45</td>
<td>0.25</td>
</tr>
<tr>
<td>5</td>
<td>22.25</td>
<td>0.23</td>
<td>0.36</td>
<td>0.25</td>
</tr>
<tr>
<td>6</td>
<td>4.19</td>
<td>0.50</td>
<td>84.59</td>
<td>0.45</td>
</tr>
<tr>
<td>7</td>
<td>1.03</td>
<td>0.23</td>
<td>0.34</td>
<td>0.25</td>
</tr>
<tr>
<td>8a</td>
<td>5.94</td>
<td>0.87</td>
<td>25.78</td>
<td>0.54</td>
</tr>
<tr>
<td>9a</td>
<td>8.82</td>
<td>0.31</td>
<td>44.28</td>
<td>0.29</td>
</tr>
<tr>
<td>Mean ± Std</td>
<td>8.43 ± 10.65</td>
<td>0.40 ± 0.24</td>
<td>38.68 ± 47.39</td>
<td>0.35 ± 0.17</td>
</tr>
</tbody>
</table>

* The amputee.

Figure 7: Boxplot of classification results in cases of 30, 60, and 90 training samples versus 90 test samples for FRVM and FSVM with both AR-RMS and WT features.
both AR-RMS and WT features was significantly worse than FSVM. With the increasing of sample size, each method achieved better recognition rate. When the sample size increased to 90, FRVM performed a comparable accuracy with FSVM. Our results demonstrated that the performance of FRVM was more dependent on training sample size. The classification accuracy significantly degraded in case of fixed and small training size.

4. Discussion and conclusions

This paper presented a FRVM learning mechanism and evaluated its application to classify hand motions using sEMG signals. Its performance was compared with FSVMs in terms of accuracy, sparsity, and response time. SVM is a binary classifier with a good generalization capability derived from structural risk minimization. Previous studies have extended SVM to FSVM and demonstrated that FSVMs were superior to artificial NN classifiers in several respects (Yan et al. 2008a, 2008b). However, both SVM and FSVM still suffer from a number of limitations. RVM is a Bayesian extension of SVM. Key attractions of RVM relative to SVM are the removal of the need to define regularizing parameter, a reduced sensitivity to hyperparameter settings, an ability to use non-Mercer kernels, the provision of a probabilistic output and a typical requirement for considerably fewer basis functions (relevance vectors) for a given analysis.

Focusing on the problem of unclassifiable regions in RVM, we proposed an improved FRVMs sparse kernel method. Two types of membership functions were presented as fuzzy logic functions to divide the indecisive regions. In an sEMG classification experiment, FRVMs showed similar accuracy over both membership functions, while FSVMs were more dependent on one over the other. The combination of RVM with membership function I was comparable to SVM with MF II.

FRVMs training resulted in much fewer RVs compared with the number of SVs obtained for FSVMs. Hence, classification could be carried out faster with FRVMs. Though various high speed micro-processors have been developed, the proposed computationally efficient algorithm is beneficial for saving hardware resources. It is particularly useful for many biosignal-based practical applications with real-time requirements including sEMG prosthetic control. However, it should be noted that FRVMs require longer training time to converge to fewer RVs in comparison with FSVMs. Another disadvantage of FRVMs related to fewer RVs is that it remains less robust to reduced training samples.

Acknowledgment

This research is partially supported by Jiangsu Provincial Key Laboratory for Interventional Medical Devices numbered (JR1412), PR of China.

References


Chaiyaratana N, Zalzala A M S and Datta D 1996 Myoelectric signals pattern recognition for intelligent functional operation of upper-limb prosthesis *Proc. 1st Int. Conf. on Disability, Virtual Reality & Associate Technologies* (Maidenhead, UK) pp 151–60


Tipping M E and Faul A C 2003 Fast marginal likelihood maximization for sparse Bayesian models *Proc. 9th Int. Workshop on Artificial Intelligence and Statistics* (Key West, FL)

