Computers in Biology and Medicine 42 (2012) 30–38

Contents lists available at SciVerse ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Ant colony optimization-based feature selection method for surface electromyography signals classification

Hu Huang [a], Hong-Bo Xie [a,*], Jing-Yi Guo [b], Hui-Juan Chen [c]

[a] School of Electronic and Information Engineering, Jiangsu University, Xuefu Rd 301#, Zhenjiang 212013, PR China
[b] Department of Basic Science, New York Chiropractic College, 2360 State Route 89, Seneca Falls, NY 13148-3204, USA
[c] Jiangbin Hospital, Jiangsu University, Zhenjiang, PR China

ARTICLE INFO

Article history:
Received 31 March 2011
Accepted 16 October 2011

Keywords:
Ant colony optimization
Feature selection
Minimum redundancy maximum relevance
Surface electromyography
Pattern classification

ABSTRACT

This paper presented a new ant colony optimization (ACO) feature selection method to classify hand motion surface electromyography (sEMG) signals. The multiple channels of sEMG recordings make the dimensionality of sEMG feature grow dramatically. It is known that the informative feature subset with small size is a precondition for the accurate and computationally efficient classification strategy. Therefore, this study proposed an ACO based feature selection scheme using the heuristic information measured by the minimum redundancy maximum relevance criterion (ACO-mRMR). The experiments were conducted on ten subjects with eight upper limb motions. Two feature sets, i.e., time domain features combined with autoregressive model coefficients (TDAR) and wavelet transform (WT) features, were extracted from the recorded sEMG signals. The average classification accuracies of using ACO reduced TDAR and WT features were $95.45 \pm 2.2\%$ and $96.08 \pm 3.3\%$, respectively. The principal component analysis (PCA) was also conducted on the same data sets for comparison. The average classification accuracies of using PCA reduced TDAR and WT features were $91.51 \pm 4.9\%$ and $89.87 \pm 4.4\%$, respectively. The results demonstrated that the proposed ACO-mRMR based feature selection method can achieve considerably high classification rates in sEMG motion classification task and be applicable to other biomedical signals pattern analysis.

1. Introduction

Electromyography signal (EMG) is an electrical manifestation of skeletal muscle contractions. Since the EMG signal directly reflects the neuromuscular activity, it has been widely used in clinical diagnosis and the rehabilitation fields [1–3]. Furthermore, the surface EMG (sEMG) signal patterns can be utilized to control prosthetic hands or other human–machine interfaces [4–11]. Recently, many attempts have been made to extract the effective and informative features from the raw surface EMG signals, such as time domain features, frequency domain features, and time–frequency domain features [6,8,11]. In the context of EMG pattern classification, it has been demonstrated that classification accuracy is affected more profoundly by the choice of feature set than by the choice of classifier [6]. Moreover, multiple electrodes are now widely used in detecting surface EMG signals to improve frequency resolution [4,5,10]. This will result in the dramatic growth of the dimensionality of EMG features and computational cost. It is believed that more features should provide more discriminating power. However, with a limited amount of training data, excessive features may not only significantly slow down the learning process, but also cause the classifier to over-fit the training data because the irrelevant or redundant features may confuse the learning algorithm [12]. This will ultimately deteriorate the classification accuracy of motion pattern, which is an important aspect of controllability of an EMG-based prosthetic control system [5,13].

Feature selection is often used to limit the amount and dimensionality of the data or to select features that correlate closely with the target class. Feature selection methods can be subdivided into those that are unsupervised and supervised [14]. The former group is unaware of class attributes or class membership, and the most of feature projection methods can be classified into this group. In these approaches, the original N-dimensional data are projected into a reduced n-dimensional feature space, where $n < N$, and the n axes of the reduced feature space are determined according to the optimal criteria. Principal component analysis (PCA), which is a commonly used feature projection method, has showed its effectiveness to reduce the feature dimensionality in EMG pattern classification [5–7,10,13]. Englehart et al. [6] first employed PCA to classify the EMG patterns and obtained better separability among different motion classes than that by Euclidean distance class separability index.

* Corresponding author. Tel.: +86 15951288938.
E-mail address: xiehb@sjtu.org (H.-B. Xie).

0010-4825/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compbiomed.2011.10.004

Chu et al. [15] investigated a combination of PCA with a self-organizing feature map (SOFM). The combination not only improved the classification accuracy, but also shortened the processing delay so that it was applicable for real-time control of prostheses. Subsequently, they promoted the performance, in terms of both the classification accuracy and the processing time, by using a linear discriminant analysis (LDA) [16]. Bu et al. [17] proposed a reduced-dimensional log-linearized Gaussian mixture network (RD-LLGMN) feature extractor. RD-LLGMN used orthogonal transformation to project the original feature space into a lower dimensional space, and then calculated the posterior probabilities with a Gaussian mixture model (GMM) in the projected lower-dimensional space for classification. The result indicated that the RD-LLGMN approach outperformed traditional methods that are based on PCA [17].

Unlike the projection-based approaches, supervised feature selection is the process of choosing a subset of features from the original feature set before classification in order to reduce the dimensionality of the data and decrease noise. It is driven by the class information, which includes filter methods and wrapper approaches [18–20]. Exhaustive search is the simplest way, which finds the best subset of features by evaluating all the possible subsets but it may cause the overwhelming computational cost. Most recently, swarm intelligence algorithms like particle swarm optimization (PSO) and ant colony optimization (ACO), have attracted a lot of attention to avoid this prohibitive complexity [21–26]. In order to classify the complicated EMG patterns, Khushaba et al. proposed a hybrid binary PSO and mutual information measurement searching strategy (BPSO-MI). The classification results indicated that features selected by BPSO-MI outperformed PCA and the uncorrelated linear discriminant analysis (ULDA) feature projection techniques [27].

ACO is another promising approach to solve the combinatorial optimization problems and has been widely utilized in feature selection. Jensen and Shen applied ACO to the problem of finding optimal feature subsets in the fuzzy-rough data reduction process [28]. Al-Ani introduced two terms, i.e., "updated selection measure (USM)" and "local importance (LI)", in his ACO-based feature selection method [22]. The proposed algorithm was then applied to the texture segment and speech classification problems. The results revealed that this method outperformed the genetic algorithm-based feature selection. Khushaba et al. [29] combined this modified ACO with a differential evaluation operator (ANTDE) and studied its applicability in EMG classification. The results demonstrated that the proposed ANTDE achieved higher classification accuracy than PCA. Literatures on feature selection using ACO can also be found in many other fields, for examples, proteomics data classification and protein function prediction in bioinformatics [30,31], financial classification involving credit risk assessment and audit qualifications [32], and text categorization [21].

Though Khushaba and his colleagues' work has shown the potential power of ACO in EMG and other bio-signals pattern classification [29], the major problem of ANTDE algorithm is its high computational complexity. Suppose that there are N features in the original feature set, and k features of them have already been chosen in the subset. When choosing the remaining features, in its innermost loop, it requires the calculation of $(k+1)$ exponents, $3(k+1)$ pairs of mutual information (MI) and some divisions as well as multiplications. In each iteration, $(N-k)$ of these operations should be proceeded to calculate the term USM. The computational complexity grows with the increment of k after each iteration, which ultimately limits its further application in EMG pattern classification since the delay time is crucial for real-time EMG-based prosthetic control [13,33]. In order to solve this problem, we proposed a new hybrid ACO-minimum redundancy maximum relevance (ACO-mRMR) approach to select the significant features. The mRMR criterion itself is a fast and accurate feature selection method. In this study, it is used to approximate the heuristic values of features [34,35]. Using this novel variant form of ACO, the ACO-mRMR was utilized in the EMG pattern classification to show its high efficacy.

The rest of paper is organized as follows. Section 2 gives a brief review of the ACO algorithm and introduces the scheme of the proposed ACO-mRMR feature selection approach. The experimental protocols and the EMG recording procedures are conducted in Section 3 and the results are given in Section 4. Finally, Section 5 contains discussion and conclusions.

2. Methods

2.1. Ant colony optimization

ACO was firstly developed by Dorigo and his colleagues in the early 1990s to solve the hard combinatorial optimization (CO) problems, such as traveling salesman problem (TSP) [36,37]. It belongs to the class of population-based probabilistic Meta-Heuristics, which are approximate algorithms used to obtain good enough solutions to hard CO problems in a reasonable amount of computation time [38]. The very first idea of ACO was inspired by the foraging behavior of real ants. Ants can find the shortest route between food source and nest using chemical materials called pheromone rather than visual information, which they leave behind their trails. Artificial ants, also referred as agents, imitate their natural counterparts and find the optimal solutions to the problems [36–38].

In general, the ACO approach attempts to solve an optimization problem by repeating the following two steps: (1) candidate solutions are constructed using a pheromone model, that is, a parameterized probability distribution over the solution space, and (2) the candidate solutions are used to modify the pheromone values in a way that is deemed to bias future sampling toward high quality solutions [38]. Since the emergence of the original ACO, various modifications of ACO were proposed which generally differ from each other in these two steps [38]. One of the ACO extensions proposed by Dorigo et al. [36] was employed in this study. Sivagaminathan and Ramakrishnan [39] have already demonstrated its applicability and efficacy in feature selection using several benchmark data sets.

2.2. ACO for feature selection

The object of feature selection based on ACO is to find, if possible, the least number of features and to obtain high classification accuracy with low computational cost. The task must be represented as a form like TSP so that ACO can be utilized. Nodes or cities in TSP represent features in the original feature set and the edges between nodes denote the choice of next feature. The search for the optimal feature subset is then converted into the problem of finding the route with a minimum number of nodes using certain criteria, such as the classification accuracy.

At the first step of the ACO, each ant constructs its solution by the *pseudo-random-proportional* action choice rule:

$$s = \begin{cases} \arg\max\{[\tau(u)]\cdot[\eta(u)]^{\beta}\} & \text{if } q \le q_0 (\textit{exploitation}) \\ S & \text{otherwise } (\textit{exploration}) \end{cases} \quad (1)$$

where $\tau(u)$ is a pheromone level of the feature u, which indicates how informative the feature is; $\eta(u)$ is a heuristic desirability of choosing feature u. For an ant k, $J_k(r)$ represents the set of features still to be selected, that is, features are not in the current solution set. β is such a parameter that determines the relative importance of the heuristic values. If $\beta=0$, the search process will only utilize the amplified pheromone, which will lead to the rapid emergence of stagnation situation. Consequently, the ants will not find any good solution at all [37]. q is a uniformly distributed value in [0.1],

and $q_0(0 \le q_0 \le 1)$ is an exploitation probability factor parameter. S is a randomly chosen feature by maximizing the following probability function using the pheromone and heuristic values

$$p_k(s) = \begin{cases} \frac{[\tau(s)]\cdot[\eta(s)]^{\beta}}{\sum_{u \in J_k(r)}[\tau(u)]\cdot[\eta(u)]^{\beta}} & \text{if } s \in J_k(r) \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

If $q \le q_0$, ants will select the features which maximize the product of pheromone level and heuristic value. In this case, the information learned from the past experience will be exploited. Otherwise, a biased exploration will proceed. Tuning q_0 can modulate the trade-off between exploration and exploitation. In other words, a higher q_0 will less possibly lead to a random choice. Both functions of the action choice rule help the ants to keep exploring new states which are close to the optimal solution.

At the second step, the pheromone trails are iterated both globally and locally. Global updating is intended to reward the features which belong to the subset with the least classification error. Only these features will modify their pheromone level $\tau(u)$ by the following global updating rule:

$$\tau(u) \leftarrow (1-\alpha)\cdot\tau(u) + \alpha\cdot\Delta\tau \quad (3)$$

where $\alpha \in (0.1]$ is a pheromone trail decay coefficient, and $\Delta\tau$ is the inverse of the least classification error of the global best ants. By using this rule, the pheromone level of the features with the least classification error will increase frequently, which makes them more inclined to be selected during subsequent iterations by the ants.

The local updating rule is intended to avoid the features with higher pheromone level being chosen by all the ants so that the ants could explore the features which have never been chosen. In each iteration, the pheromone level $\tau(r)$ of those features which are not selected in the best subset are updated by the rule

$$\tau(r) \leftarrow (1-\alpha)\cdot\tau(r) + \alpha\cdot\tau_0 \quad (4)$$

where τ_0 is the initial pheromone level. By using this updating rule, the irrelevant features become less desirable and the pheromone level of the features that have been a part of the best subset of previous iterations will decrease slightly. On the other hand, for the features that have never appeared in the best feature subset, the value will remain the same. Consequently, the pheromone level of all features will never decrease to below τ_0. This prevents the ants from converging to a common path [39].

2.3. Calculation of heuristic values based upon mRMR

The heuristic desirability in ACO is a problem-dependent value and different approaches have been proposed to evaluate the heuristic information of each feature [22,32,40]. An easily computable form of heuristic evaluation does not affect the performance of the ACO algorithm due to the compensation of the pheromone for small potential errors in the heuristic values [40]. In this paper, mRMR criterion was used to determine the heuristic values. Compared with Al-Ani's mutual information method, it can nearly approximate the desirability of each feature with a more computationally effective form which has been mentioned before [22]. Moreover, compared with other stochastic or entropy-based methods to determine the heuristic values, mRMR criterion considers not only the minimum redundancy between features but also the maximum relevance between features and class labels [32,35,40].

The mRMR approach was proposed by Ding and Peng as a new solution to select the optimal feature subset which is also based on mutual information [34,35]. There are two constraints to be considered, i.e., the minimum redundancy in (5), and the maximum relevance in (6):

$$\min W_I, W_I = \frac{1}{|\Psi|^2}\sum_{i,j \in S} I(i,j) \quad (5)$$

$$\max V_I, V_I = \frac{1}{|\Psi|}\sum_{i \in S} I(h,i) \quad (6)$$

where $I(i,j)$ represents the mutual information between the two features, i and j; $I(h,i)$ represents the mutual information between the desired output class h and the feature i. Ψ is the candidate feature set to be selected, and $|\Psi|$ is the number of features in Ψ. The minimum redundancy condition makes the selected feature set a better representation of entire data set since the selected features are mutually maximally dissimilar. The maximum relevance condition maximizes the total relevance of all features in the selected feature subset. The mRMR criterion optimizes mutual information quotient Q by taking the form

$$Q = \max\left(\frac{V_I}{W_I}\right) \quad (7)$$

Readers may refer to Refs. [34,35] for the details of feature selection based on mRMR criterion. In this study, we applied the mRMR criterion to determine the approximate heuristic desirability of each feature. First, all the features in the original feature set are ranged according to the mutual information quotient Q and construct a new feature set Ψ. Then the heuristic values of the features are calculated as

$$\eta_m = \begin{cases} \lambda\cdot Q_2 + \theta & \text{if } m = 1 \\ \lambda\cdot Q_m & \text{otherwise} \end{cases} \quad (8)$$

where m denotes the order number of a feature in Ψ, η_m represents the heuristic value of the mth feature while Q_m is the mutual information quotient calculated by (7). λ is a parameter that regulates the range of heuristic values for the inappropriate heuristic values would affect the convergence of ACO. θ is a positive integral used to ensure that the heuristic function of the first feature in Ψ is larger than that of the second one, i.e., $\eta_1 > \eta_2$. The earlier the feature selected in Ψ, the more desirable it is to be selected in the optimal feature subset.

2.4. Proposed feature selection algorithm

The block diagram of the hybrid ACO-mRMR feature selection method applied to the sEMG pattern classification is shown in Fig. 1. The left part is the sEMG pattern classification section, and the right part is the feature selection section. In the feature selection section, each ant, coded as a binary vector, represents a feature subset. If the ith feature is selected, the corresponding bit in an ant is coded as 1, otherwise coded as 0. Each ant constructs a subset with n features, where n denotes the size of feature subset. The proposed method is to search for the least number of features that achieve the lowest classification error. The details of the procedure are depicted in the following steps:

Step 1: Initialize the parameters of ACO, including initial pheromone level τ_0, the heuristic values η_m of features, the number of ants m, the maximum number of iteration *Imax*, and the tunable parameters α, β, and q_0.

Step 2: Construct candidate solutions from a randomly selected feature and select the rest $n-1$ features using the *pseudo-random-proportional* rule.

Step 3: Evaluate the candidate feature subsets using the trained classifier by testing the classification accuracy on the validation set.

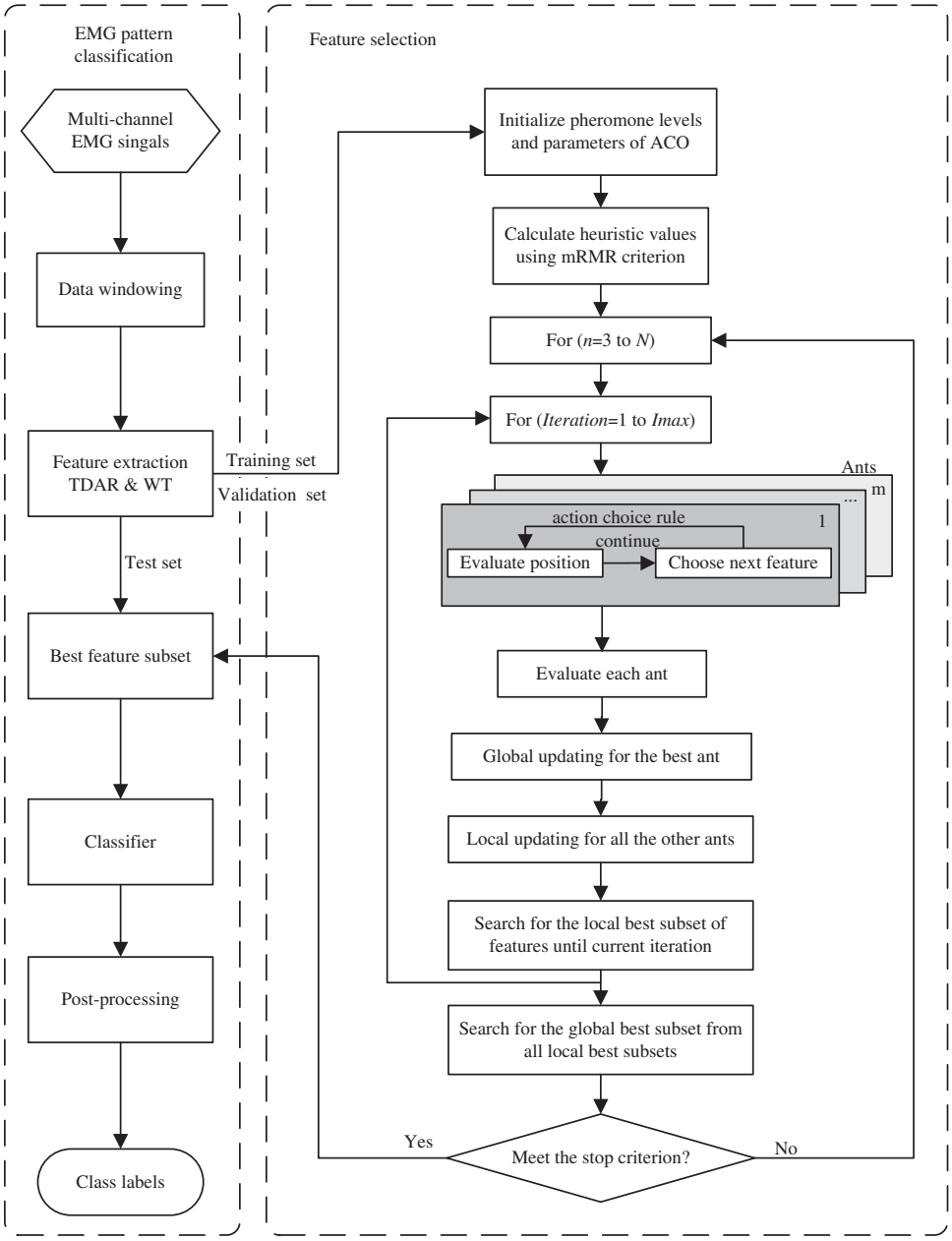

Fig. 1. The structure of the proposed ACO-mRMR based feature selection and the sEMG pattern classification. The left denotes the sEMG pattern classification, and the right denotes the ACO-mRMR based feature selection section.

Step 4: Apply the global updating rule (3) to the feature subset which produces the best classification accuracy, and utilize the local updating rule (4) to all the other solutions.
Step 5: Search for the best subset of local features that yields the highest classification accuracy. If the maximum number of iteration *Imax* is not reached, continue to construct candidate solutions using the *pseudo-random-proportional* rule and go back to Step 3; otherwise, go to the next step.
Step 6: Search for the global best subset which produces the highest classification accuracy among all local best solutions.
Step 7: Increase the number of features for the next searching step, i.e., $n \leftarrow n+1$. The number of features n ranges from three to N, where N is the number of features in the original feature set.
Step 8: Check the following stop criteria to determine whether to continue searching: (1) the classification accuracies using the last three best global solutions, respectively, are not increased, (2) n reaches N. If neither of the stop criteria meets, go to Step 2; otherwise return the best global feature subset that yields the highest classification accuracy and the trained classifier to the sEMG classification section.

3. Experimental protocols

The ACO-based feature selection method was evaluated on the sEMG data collected from the following experiment. Eight distinct wrist and hand motions were used: grasp (GR), hand open (OP), wrist flexion (WF), wrist extension (WE), ulnar deviation (UD), radial deviation (RD), pinch (PN), and thumb flexion (TF), as depicted in Fig. 2. These represent the commonly used wrist and hand movements in daily life.

In the experiment, the sEMG data were collected from ten non-amputee subjects (six males and four females with the age of 22–

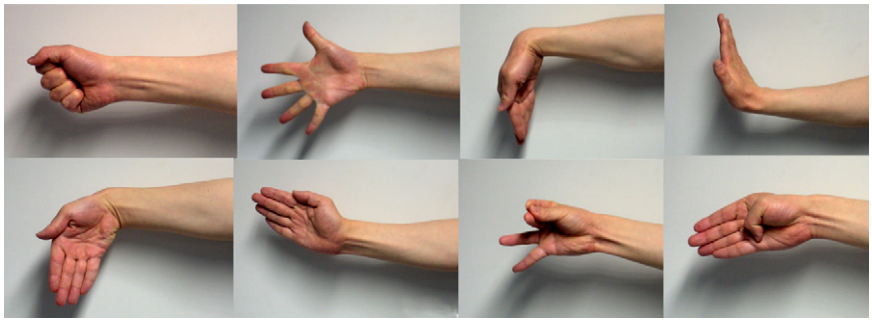

Fig. 2. Eight classes of motion were used in the experiment. From the left to right in the first row: grasp (GP), hand open (OP), wrist flexion (WF), wrist extension (WE), and in the second row: ulnar deviation (UD), radial deviation (RD), pinch (PN), thumb flexion (TF).

38). All the participants were right-hand dominant without any known neuromuscular disorders. The human subject ethical approval was obtained from the relevant committee in the authors' institution and informed consent was obtained from all subjects prior to the experiment. Four channels of sEMG signals were acquired from the forearm using the EMG 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 of two surface electrodes was 2 cm. Skin areas of interest were abrased beforehand with alcohol. An additional Ag–AgCl electrode was placed on the elbow to provide a common ground reference. EMG signal was amplified by an amplifier (RM-6280C, Chengdu Device Inc. Sichuan, China) with a gain of 2000, filtered by 8–500 Hz band-pass analog filter within the amplifier, and then digitized by a 12-bit data acquisition card (NI PCI-6024E, National Instruments, Austin, TX) with the sample frequency of 1 kHz.

During the experiment, subjects were required to perform three repeated 80 s continuous contractions. Within each session, each limb motion was held in a random order for about 10 s. The participants could relax between each session to avoid fatigue. The first session was used as training set while the second and the third as the validation and test set, respectively. All the data were segmented into consecutive 256 ms epochs.

In this paper two sets of features were extracted from the recorded sEMG signals in time and time–frequency (TF) domain. Autoregressive (AR) model coefficients (16 features=4 channels × 4 features/channel) and 6 time domain features [6,8], i.e., mean absolute value (MAV), zero crossing (ZC), slope sign changes (SSC), waveform length (WL), variance of sEMG (VAR), and Willson amplitude (WAMP) (24 features=4 channels × 6 features/channel, referred as TD) were extracted and combined (40 features=16 features+24 features, referred as TDAR). For the time-scale features, a set of wavelet transform energy coefficients at each scale using a Symmlet wavelet family (Sym5) with four levels (20 features=4channel × 5 features/channel, referred as WT) were extracted [41].

A three-layer back-propagation neural network (BPNN) was used as the classifier to perform the evaluation of the ants and the classification of motions. The BPNN had twenty hidden layer nodes; the hidden layer had a tan-sigmoid activation function, and the output layer had a linear activation function [42]. After the ANN classification, the accuracy was further improved by a post-processing procedure using majority vote (MV) [13].

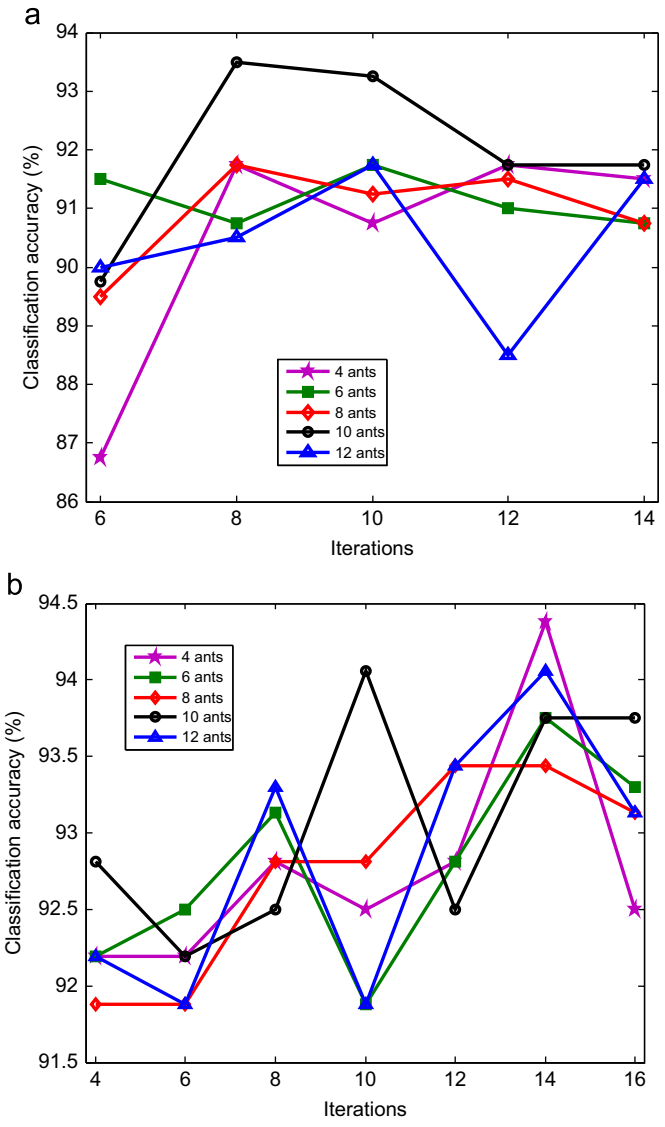

Fig. 3. Classification accuracies using different numbers of ants and iterations for (a) TDAR feature set and (b) WT feature set. Each curve represents the case that uses a certain number of ants, as given in different marks (4, 6, 8, 10, and 12 ants).

4. Results

It is important to assign appropriate value to the tunable parameters before implementing the ACO-mRMR based feature selection. Improper values can significantly affect the ACO convergent time or even lead to premature convergence. The number of ants and iterations should be sufficient to explore all potential solutions, while spending as little as possible time. In the experiment, we chose the parameters that yielded high classification accuracy in a comparatively short time, i.e., with fewer number of iterations. A number of trials were proceeded to

determine the initial parameters. The number of ants was varied from 4 to 12 in step of 2 for both feature sets. The number of iterations was varied from 6 to 14 in step of 2 for TDAR feature, and from 4 to 16 in step of 2 for WT feature. For each combination, three trials were performed. The classification accuracy on validation set by different number of ants and iterations for TDAR and WT features are shown in Fig. 3a and b, respectively. The classification accuracy was improved by increasing the number of ants and iterations though it fluctuated or even decreased when the number of ants or iterations reached a certain high value. This phenomenon is consistent with that of Sivagaminathan and Ramakrishnan's work [39]. On the other hand, this high values also resulted in a tremendous increase of computational time, which is improper for EMG based prosthetic control in real time. Therefore, the parameters were set by the trade-off between the accuracy and efficiency listed in Table 1.

In order to clearly illustrate the efficacy of the proposed feature selection method, the performance of feature subsets obtained by PCA method were also tested for comparison [5–7,10,13]. Fig. 4a and b showed the typical scattering plots of the first three TDAR features selected by PCA and ACO-mRMR using the sEMG data acquired from subject 4, respectively. Correspondingly, Fig. 5a and b were the scattering plots of the first three WT features selected by PCA and ACO-mRMR, respectively. It was indicated that the first three features selected by ACO-mRMR had a better class separability than that by PCA. Fig. 6a and b showed the comparable classification results of subject 4 between the two methods with increasing number of TDAR and WT features, respectively. The results demonstrated that the feature subsets selected by ACO-mRMR always gave higher classification accuracy than that of PCA. Moreover, the highest classification accuracy of PCA method (86.09% for TDAR and 85.94% for WT) was lower than the proposed ACO-mRMR method (91.56% for TDAR and 94.69% for WT). A majority vote scheme was then performed to improve the accuracy [13]. The detailed results of the two methods to classify each motion are listed in Table 2. It could be observed that the majority vote was an effective post-processing scheme to improve the classification accuracy for both ACO-mRMR and PCA feature selection methods. The error rate using a majority vote scheme was roughly half of the unprocessed stream of class decisions, regardless of the feature set (TDAR or WT) or feature selection method (ACO-mRMR or PCA). However, the accuracy of ACO-mRMR based method (96.25% for TDAR and 98.44% for WT) was still superior to that of PCA (92.81% for TDAR and 89.53% for WT) after majority vote post-processing. Fig. 7 showed the original sEMG data and the continuous stream of

Table 1
The best parameter set of ACO-mRMR for the sEMG pattern classification experiment.

Features	Ants	Iterations	β	α	q_0
TDAR	10	8	2	0.8	0.8
WT	4	14	2	0.8	0.8

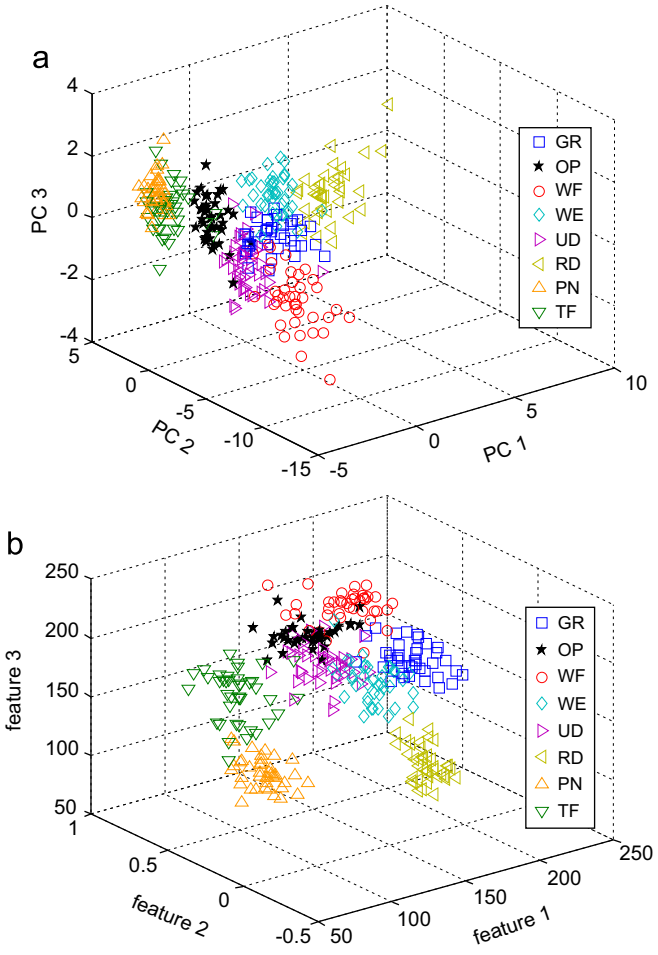

Fig. 4. Three-dimensional scatter plot of (a) the first three principal components of TDAR feature set and (b) the three most frequently selected TDAR features by ACO-mRMR. The parameters used in this experiment were concluded in Table 1.

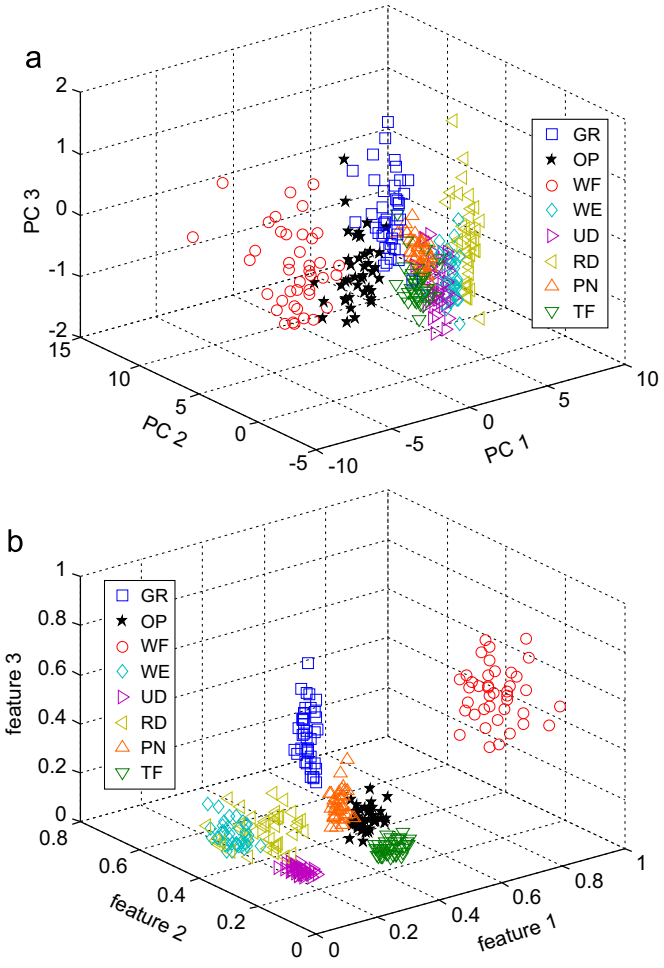

Fig. 5. Three-dimensional scatter plot of (a) the first three principal components of WT feature set and (b) three most frequently selected WT features by ACO-mRMR.

class decisions using TDAR and WT features reduced by ACO-mRMR. In this figure, red crosses indicated the positions where the MV scheme produced errors. These errors were almost exclusively restricted to the transition regions of the activity. The results of all the other subjects were similar to that of subject 4. The box-plot of overall average classification performances are shown in Fig. 8 and details are shown in Table 3. The average number of optimal features selected by ACO-mRMR was 9.9 for TDAR and 9.3 for WT feature set over the all subjects, and the mean accuracies for the two feature sets were 95.45 ± 2.2% and 96.08 ± 3.3%, respectively. While using PCA method, the average accuracies for the two feature sets were 91.51 ± 4.9% and 89.87 ± 4.4%, respectively, which are significantly lower than those of the proposed method. In order to statistically compare the performances between the two methods, a one-way analysis of variance (ANOVA) was performed [43]. The proposed ACO-mRMR method achieved significant improvement in classification accuracy as compared with the PCA method for both TDAR ($p=0.0332$) and WT ($p=0.0010$) features. Therefore, it could be concluded that ACO-mRMR feature selection method could not only significantly reduce the sEMG feature dimension but also the classification error.

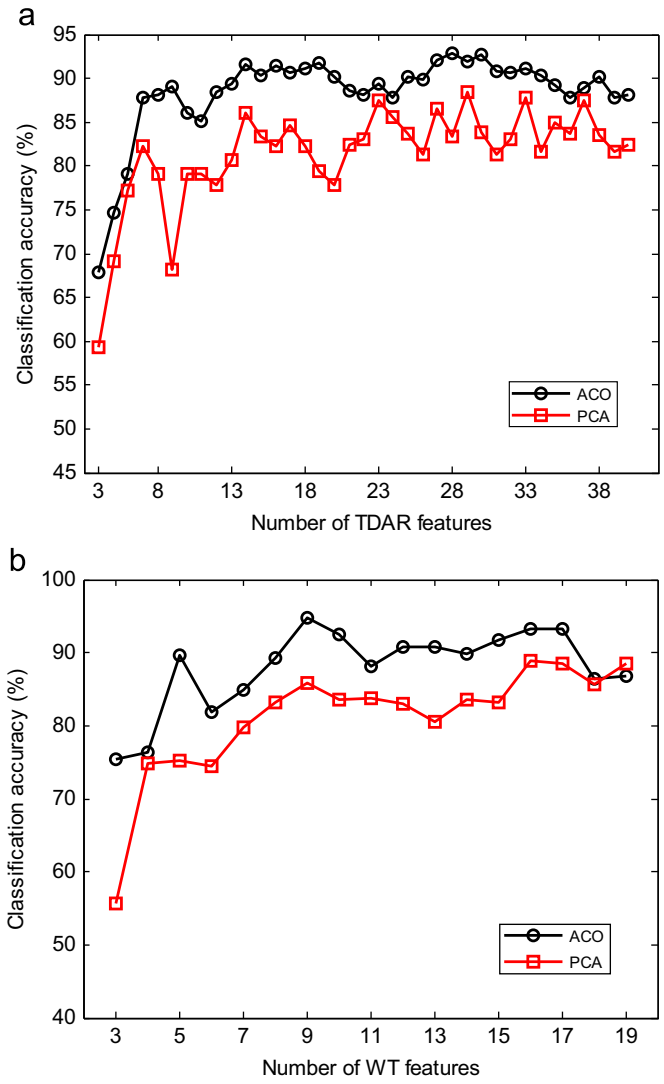

Fig. 6. Classification accuracies (subject 4) using the ACO-mRMR and PCA based features subsets with different number of features for (a) TDAR feature set and (b) WT feature set.

5. Discussion and conclusions

This paper presented a new swarm intelligence based feature selection algorithm utilizing both ACO and mRMR critera. The proposed method was tested in the surface EMG classification problem on ten subjects' data sets. It was also compared with PCA, the most commonly used feature reduction technique in EMG classification [5–7,10,13]. PCA is an unsupervised learning method and it only discovers correlation among patterns and their elements, as well as ordered intrinsic directions where the data patterns change most (with maximum variance). It does not consider the interaction between the features to discriminate the classes. Different from PCA, the proposed hybrid ACO-mRMR method took into account not only the relevance between the feature and the class pattern but also the redundancy among features. In addition, the mRMR criterion was adopted to measure the heuristic values of features, as the indicator of importance. It is more efficient than the earlier mutual information-based method to determine the heuristic values [22,40]. The fewer number of features selected by ACO-mRMR can not only improve the accuracy of EMG classification problem, but more importantly, it can result in a simplified structure of the classifier and reduce the computational cost.

An interesting result was that the average number of features selected in the optimum subset was 9.9 for TDAR and 9.3 for WT feature set. However, it varied from 5 to 14 for different subjects, which made sense for the reason that sEMG is a subject-dependent bio-signals [44] in some degree. Various factors may influence sEMG signal when it is collected, including not only external factors such as electrode–electrolyte interface, and electrode configuration, but also internal factors such as thickness of skins, tissues, and difference of the number of muscle fibers, which render different sEMG signal from person to person. Thus, one can hardly say that a specific feature subset of a specific subject, showing high separability between classes, will equally work well to other subjects [44].

To sum up, the present work has revealed the potential power of the ACO for sEMG feature selection. Apart from the sEMG signals, the method presented in this paper might be used for

Table 2
The classification accuracies (%) of each motion by ACO-mRMR and PCA methods for a typical subject (subject 4). Accuracy in the first column is before MV, and the other one given in parenthesis is after MV.

Features	Feature subsets	GR	OP	WF	WE	UD	RD	PN	TF	Average
TDAR	Original	94 (100)	73 (86)	87 (92)	91 (98)	87.5 (98.5)	88 (99)	74 (85.5)	79 (85)	84.19 (93)
	ACO	100 (100)	90 (97.5)	82.5 (92.5)	100 (100)	87.5 (95)	75 (87.5)	97.5 (97.5)	100 (100)	91.56 (96.25)
	PCA	97.5 (100)	100 (100)	77.5 (95)	85 (96.25)	77.5 (96.25)	58.75 (58.75)	97.5 (100)	95 (96.25)	86.09 (92.81)
WT	Original	100 (100)	95 (100)	80 (95)	100 (100)	87.5 (97.5)	82.5 (100)	90 (100)	97.5 (100)	91.56 (99.06)
	ACO	100 (100)	100 (100)	82.5 (95)	97.5 (100)	100 (100)	77.5 (92.5)	100 (100)	100 (100)	94.69 (98.44)
	PCA	98.75 (100)	68.75 (75)	92.5 (97.5)	98.75 (98.75)	96.25 (98.75)	77.5 (78.75)	56.25 (67.5)	98.75 (100)	85.94 (89.53)

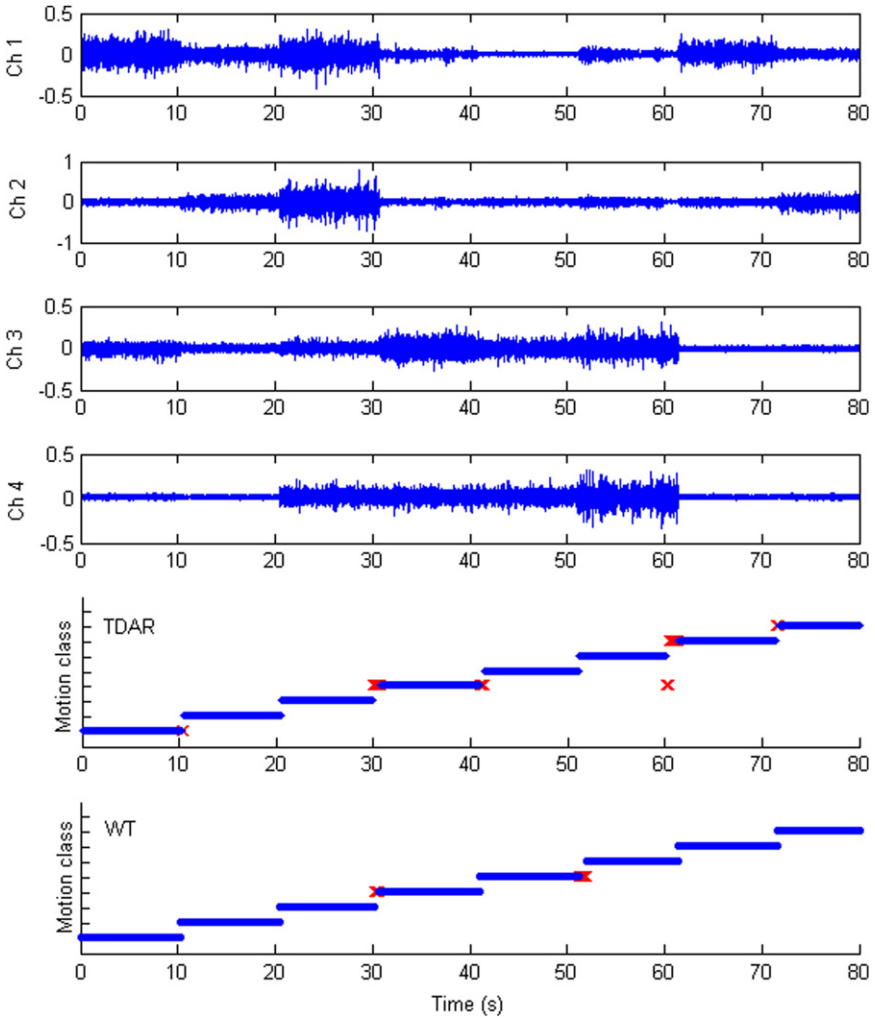

Fig. 7. The typical sEMG signals of 80 s duration and continuous classification performance (subject 4) using TDAR and WT feature subsets selected by ACO-mRMR. The cross represents the misclassified motions.

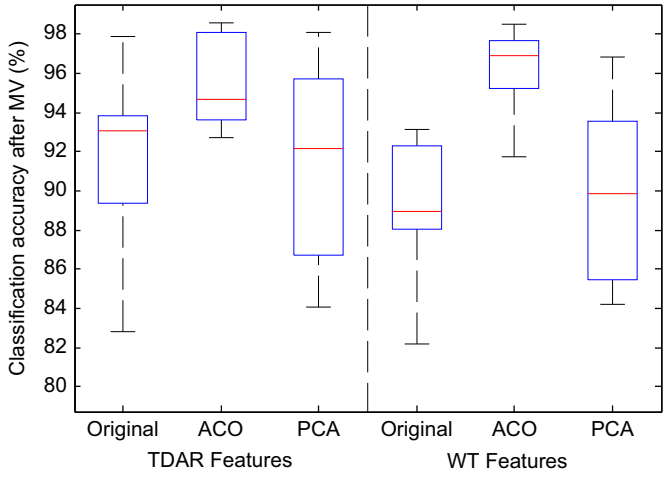

Fig. 8. Box-plot of classification accuracy after MV over all subjects using feature sets (TDAR and WT) without reduction (original), and feature subsets reduced by ACO-mRMR and PCA methods.

analyzing other physiological signals that could be easily generalized to other applications. For further study, different models of ACO as well as other swarm intelligence based optimization algorithms may be investigated to find a more effective and

Table 3
Classification results of all ten subjects by ACO-mRMR and PCA based feature subsets.

Features	Subjects	Before MV		After MV	
		ACO	PCA	ACO	PCA
TDAR	1	92.25	83.38	98.58	88.38
	2	86.09	84.69	94.06	94.06
	3	89	90.58	94.83	95.75
	4	91.56	86.09	96.25	92.81
	5	90	80.94	94.53	84.06
	6	91.56	83.44	98.12	91.56
	7	93.44	88.75	98.28	98.16
	8	92.19	83.28	93.44	86.72
	9	91.77	91.04	93.65	97.29
	10	89.79	81.25	92.71	86.35
	Average	90.77 ± 2.1	85.34 ± 3.7	95.45 ± 2.2	91.51 ± 4.9
WT	1	88.08	78.75	91.75	84.25
	2	88.75	84.38	95.21	90.21
	3	89.67	88.17	96.75	95.08
	4	94.69	85.94	98.44	89.53
	5	91.25	83.28	95.47	93.59
	6	86.56	79.53	92.5	85.47
	7	90.31	87.03	97.03	96.88
	8	93.91	86.88	97.66	91.88
	9	93.13	83.02	97.5	86.46
	10	93.75	80.83	98.54	85.31
	Average	91.01 ± 2.8	83.78 ± 3.3	96.08 ± 2.4	89.87 ± 4.4

applicable approach for feature selection. Moreover, ACO can also be applied simultaneously to optimize the structure of the neural network to improve both the classification accuracy and the efficiency further [45].

Conflict of interest statement

None of the authors has any conflict of interest in this study.

Acknowledgments

This work is supported by the Jiangsu Natural Science Foundation (BK2009198), PR China.

References

[1] A. Holtermann, P.J. Mork, L.L. Andersen, H.B. Olsen, K. Søgaard, The use of EMG biofeedback for learning of selective activation of intra-muscular parts within the serratus anterior muscle: a novel approach for rehabilitation of scapular muscle imbalance, J. Electromyogr. Kines 20 (2010) 359–365.
[2] R. Neblett, T.G. Mayer, R.J. Gatchel, Theory and rationale for surface EMG-assisted stretching as an adjunct to chronic musculoskeletal pain rehabilitation, Appl. Psychophys. Biof. 28 (2003) 139–146.
[3] H.B. Xie, Z.Z. Wang, Mean frequency derived via Hilbert-Huang transform with application to fatigue EMG signal analysis, Comput. Meth. Prog. Biomed. 82 (2) (2006) 114–120.
[4] A.B. Ajiboye, R.F. Weir, A heuristic fuzzy logic approach to EMG pattern recognition for multifunctional prosthesis control, IEEE Trans. Neural Syst. Rahabil. Eng. 13 (2005) 280–291.
[5] K. Englehart, B. Hudgins, A.D.C. Chan, Continuous multifunction myoelectric control using pattern recognition, Technol. Disabil. 15 (2) (2003) 95–103.
[6] K.B. Englehart, B.S. Hudgins, P.A. Parker, M. Stevenson, Classification of the myoelectric signal using time–frequency based representations, Med. Eng. Phys. 21 (1999) 431–438.
[7] Y. Huang, K.B. Englehart, B.S. Hudgins, A.D.C. Chan, A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses, IEEE Trans. Biomed. Eng. 52 (11) (2005) 1801–1811.
[8] B. Hudgins, P.A. Parker, R.N. Scott, A new strategy for multifunction myoelectric control, IEEE Trans. Biomed. Eng. 40 (1993) 82–94.
[9] M.A. Oskoei, H. Hu, Myoelectric control systems—a survey, Biomed. Signal Process. Control 2 (4) (2007) 275–294.
[10] P.A. Parker, K.B. Englehart, B. Hudgins, Myoelectric signal processing for control of powered limb prostheses, J. Electromyogr. Kines 16 (6) (2006) 541–548.
[11] M. Zecca, S. Micera, M.C. Carrozza, P. Dario, Control of multifunctional prosthetic hands by processing the electromyographic signal, Crit. Rev. Biomed. Eng. 30 (4–6) (2002) 459–485.
[12] L. Yu, H. Liu, Efficient feature selection via analysis of relevance and redundancy, J. Mach. Learn. Res. 5 (2004) 1205–1224.
[13] K. Englehart, B. Hudgins, A robust, real-time control scheme for multifunction myoelectric control, IEEE Trans. Biomed. Eng. 50 (7) (2003) 848–854.
[14] P. Smialowski, D. Frishman, S. Kramer, Pitfalls of supervised feature selection, Bioinformatics 26 (3) (2010) 440–443.
[15] J.U. Chu, I. Moon, M.S. Mun, A real-time EMG pattern recognition system based on linear-nonlinear feature projection for a multifunction myoelectric hand, IEEE Trans. Biomed. Eng. 53 (11) (2006) 2232–2239.
[16] J.U. Chu, I. Moon, Y.J. Lee, S.K. Kim, M.S. Mun, A supervised feature-projection-based real-time EMG pattern recognition for multifunction myoelectric hand control, IEEE/ASME Trans. Mechatron. 12 (3) (2007) 282–290.
[17] N. Bu, J. Arita, T. Tsuji, A novel pattern classification method for multivariate EMG signals using neural network, in: Proceedings of the First International Conference on Advances in Natural Computation, Part II, Springer, 2005, pp. 165–174.
[18] G.H. John, R. Kohavi, K. Pfleger, Irrelevant features and the subset selection problem, in: Proceedings of the 11th International Conference on Machine Learning, Morgan Kaufmann Publishers, 1994, pp. 121–129.
[19] R. Kohavi, G.H. John, Wrappers for feature subset selection, Artif. Intell. (1–2) (1997) 273–324.
[20] I.H. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufman, San Francisco, 2005, pp. 288–296.
[21] M.H. Aghdam, N. Ghasem-Aghaee, M.E. Basiri, Text feature selection using ant colony optimization, Expert Syst. Appl. 36 (3) (2009) 6843–6853.
[22] A. Al-Ani, Feature subset selection using ant colony optimization, Int. J. Comput. Intell. 2 (1) (2006) 53–58.
[23] C.L. Huang, J.F. Dun, A distributed PSO–SVM hybrid system with feature selection and parameter optimization, Appl. Soft. Comput. 8 (4) (2008) 1381–1391.
[24] Y. Liu, Z. Qin, Z. Xu, X. He, Feature selection with particle swarms, in: Proceedings of the First International Symposium on Computational and Information Science, 2004, pp. 425–430.
[25] W. Pedrycz, B.J. Park, N.J. Pizzi, Identifying core sets of discriminatory features using particle swarm optimization, Expert Syst. Appl. 36 (3) (2009) 4610–4616.
[26] A. Unler, A. Murat, A discrete particle swarm optimization method for feature selection in binary classification problems, Eur. J. Oper. Res. 206 (3) (2010) 528–539.
[27] R.N. Khushaba, A. Alsukker, A. Al-Ani, A. Al-Jumaily, A.Y. Zomaya, A novel swarm based feature selection algorithm in multifunction myoelectirc control, J. Intell. Fuzzy Syst. 20 (4) (2009) 175–185.
[28] R. Jensen, Q. Shen, Fuzzy-rough data reduction with ant colony optimization, Fuzzy Set Syst. 149 (1) (2005) 5–20.
[29] R.N. Khushaba, A. Al-Ani, A. AlSukker, A. Al-Jumaily, A combined ant colony and differential evolution feature selection algorithm, in: Proceedings of the Sixth International Conference on Ant Colony Optimization and Swarm Intelligence, Springer, 2008, pp. 1–12.
[30] S. Nemati, M.E. Basiri, N. Ghasem-Aghaee, M.H. Aghdam, A novel ACO-GA hybrid algorithm for feature selection in protein function prediction, Expert Syst. Appl. 36 (2009) 12086–12094.
[31] W. Zhao, C.E. Davis, Swarm intelligence based wavelet coefficient feature selection for mass spectral classification: an application to proteomics data, Anal. Chim. Acta 651 (2009) 15–23.
[32] Y. Marinakis, M. Marinaki, M. Doumpos, C. Zopounidis, Ant colony and particle swarm optimization for financial classification problems, Expert Syst. Appl. 36 (7) (2009) 10604–10611.
[33] T.R. Farrell, R.F. Weir, The optimal controller delay for myoelectirc prostheses, IEEE Trans. Neur. Syst. Rahabil. Eng. 15 (1) (2007) 111–118.
[34] C. Ding, H. Peng, Minimum redundancy feature selection from microarray gene expression data, J. Bioinf. Comput. Biol. 3 (2) (2005) 185–205.
[35] H. Peng, F. Long, C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern. Anal. Mach. Intell. 27 (8) (2005) 1226–1238.
[36] M. Dorigo, G.D. Caro, L.M. Gambardella, Ant algorithms for discrete optimization, Artif. Life 5 (1999) 137–172.
[37] M. Dorigo, V. Maniezzo, A. Colorni, Ant system: optimization by a colony of cooperating agents, IEEE Trans. Syst. Man Cybern. B Cybern. 26 (1) (1996) 29–41.
[38] M. Dorigo, C. Blum, Ant colony optimization theory: a survey, Theor. Comput. Sci. 344 (2005) 243–278.
[39] R.K. Sivagaminathan, S. Ramakrishnan, A hybrid approach for feature subset selection using neural networks and ant colony optimization, Expert Syst. Appl. 33 (2007) 49–60.
[40] B. Liu, H.A. Abbass, B. McKay, Classification rule discovery with ant colony optimization, in: Proceedings of the IEEE/WIC International Conference on Intelligence Angent Technology, IEEE Computer Society, 2003, pp. 83–88.
[41] X. Hu, Z. Wang, X. Ren, Classification of surface EMG signal using relative wavelet packet energy, Comput. Meth. Prog. Biol. 79 (3) (2005) 189–195.
[42] S. Haykin, Neural Networks and Learning Machines, Prentice Hall, Upper Saddle River, New Jersey, 2008.
[43] P.A. Castillo-Valdivieso, J.J. Merelo, A. Prieto, I. Rojas, G. Romero, Statistical analysis of the parameters of a neuro-genetic algorithm, IEEE Trans. Neural Network 13 (2002) 1374–1394.
[44] J.S. Han, W.C. Bang, Z.Z. Bien, Feature set extraction algorithm based on soft computing techniques and its application to EMG pattern classification, Fuzzy Optim. Decis. Making 1 (2002) 269–286.
[45] K. Socha, C. Blum, An ant colony optimization algorithm for continuous optimization: application to feed-forward neural network training, Neural Comput. Appl. 16 (3) (2007) 235–247.

Hu Huang is now pursuing his Master degree in biomedical engineering. His research interest is biomedical signal processing.

Hong-Bo Xie got his Ph.D. degree of Biomedical Engineering in Shanghai Jiao Tong University. He joined The Hong Kong Polytechnic University as a research associate in 2007 and then a postdoctoral fellow in 2008. He is now the Associate Professor with Jiangsu University. He has published more than 50 peer reviewed journal papers. His research interest includes chaotic time series analysis, biomedical signal processing, and artificial intelligence.

Jing-Yi Guo got her Ph.D. degree of Biomedical Engineering in The Hong Kong Polytechnic University in 2010. She is now a postdoctoral fellow in New York Chiropractic College. Her research interest is neuromuscular assessment by EMG and ultrasound techniques.

Hui-Juan Chen is a neurophysiological clinician. Her research interest is prosthetic hand control and neural rehabilitation.