Proceedings of the Sixth IASTED International Conference on Biomedical Engineering, held February 13 – 15, 2008 in Innsbruck, Austria

SPONSORS
The International Association of Science and Technology for Development (IASTED)
• Technical Committee on Biomedical Engineering
World Modelling and Simulation Forum (WMSF)

EDITOR
A. Hierlemann – Swiss Federal Institute of Technology Zurich (ETHZ), Switzerland

KEYNOTE SPEAKER
E. Simon – Siemens AG, Germany

SPECIAL SESSION ORGANIZERS
B. Jansen and R. Deklerck – Vrije University Brussel, Belgium
E. D. Übeyli – TOBB University of Economics and Technology, Turkey

INFORMATION ON PUBLICATION
For each IASTED conference, the following review process is used to ensure the highest level of academic content. Each full manuscript submission is double blind reviewed by a minimum of two separate reviewers on the International Program Committee/Additional Reviewers list. The review results are then compiled. If there are conflicting reviews, the paper is sent to a third reviewer.

Photo Credits: Background © Sachin Ghodke; Centre © Peter Rol; Bottom © Rodolfo Clix

Copyright © 2008 ACTA Press

ACTA Press
P.O. Box 5124
Anaheim, CA 92814–5124
USA

ACTA Press
B6, Suite #101, 2509 Dieppe Ave SW
Calgary, Alberta T3E 7J9
Canada

ACTA Press
P.O. Box 354
CH-8053 Zurich
Switzerland

Publication Code: 601
ABSTRACT

The myoelectric signals (MES) from human muscles have been utilized in many applications such as prosthesis control. The identification of various MES temporal structures is used to control the movement of prosthetic devices by utilizing a pattern recognition approach. Recent advances in this field have shown that there are a number of factors limiting the clinical availability of such systems. Several control strategies have been proposed but deficiencies still exist with most of those strategies especially with the Dimensionality Reduction (DR) part. This paper proposes using Particle Swarm Optimization (PSO) algorithm with the concept of Mutual Information (MI) to produce a novel hybrid feature selection algorithm. The new algorithm, called PSOMIFS, is utilized as a DR tool in myoelectric control problems. The PSOMIFS will be compared with other techniques to prove the effectiveness of the proposed method. Accurate results are acquired using only a small subset of the original feature set producing a classification accuracy of 99% across a problem of ten classes based on tests done on six subjects MES datasets.

KEY WORDS

Myoelectric control, feature selection, particle swarm.

1. Introduction

The Myoelectric signal (MES), also known as Electromyogram signal (EMG), is one of the biosignals generated by the human body [1]. It represents the muscles activity or the summation of the action potentials from many motor units [2, 3]. The MES signal is a one dimensional nonstationary signal that carries the distinct signature of the voluntary intent of the central nervous system (CNS) [4]. It has long been recognized as an efficient and promising resource for human-machine interaction (HMI). The MES is usually utilized in a non-invasive scheme, and used in many diverse applications including controlling prosthetic devices [5], speech recognition [6], and schemes of functional electrical stimulation [7].

Although the concept of myoelectric control has been known for many years [8], but the actual successfulliness is measured by means of the clinical availability of those systems. The most important factors that affect the clinical availability and users acceptance depends on the type of prosthesis, the control strategy, and user training [1]. The control strategy is of particular importance as it is usually accompanied with low acceptance rates by the patients in most of the cases [1]. This happens when the user perceives an inadequate controllability (the lack of intuitive and dexterous control). The basic parts of the control strategy are:

1. Features Extraction.
2. Feature Set Dimensionality Reduction.
3. A Suitable Classifier.

A significant amount of research focused on the first and the third parts throughout the literature. In the features extraction part various techniques and methods were developed and utilized to extract features from the myoelectric signals like mean absolute value [9], integral absolute value [10], and zero crossing [11]. However; the pattern recognition results using these feature vectors have not had a high success rate because such methods assumes that EMG signal is stationary, while in reality it is not [12]. The time-frequency analysis such as the short-time Fourier transform (FT) [13], Gabor representation, Winger-Ville (WVD) distribution, Wavelet transform (WT) [14], and Wavelet Packet transform (WPT) [15] also received considerable attention in the analysis of non-stationary signals. According to recent research in this field [16, 17], the TDAR features extracted by a 6'th order autoregressive (AR) model concatenated with Hudgins time domain features (TD) [9] can produce very powerful results in myoelectric control problems. In the classifier part, studies in this field indicates that Linear Discriminant Analysis (LDA) and Multilayer Perceptron neural network trained with back propagation algorithm (BPNN) are the most widely used in this field with a special focus on LDA.

One of the first studies of dimensionality reduction techniques in myoelectric control was made by Englehart[18], in which a comparison of both feature selection (FS) and projection (FP) on different MES datasets was presented. Englehart proved that features extracted by wavelet packet transform and later reduced in dimensionality by principal component analysis (PCA), a feature projection method, produced the best results.
when compared to other feature selection methods [19]. A fact that should be mentioned here is that, PCA is merely a dimensionality reduction tool that does not take into consideration the interaction between the features to discriminate between the output classes [20]. Recent research [16, 21] revealed that when the number of motion classes to be classified increases, the MES classification accuracies of the features projected with PCA does actually decrease. A combination of linear-nonlinear feature projection techniques was introduced by employing PCA and a self organizing feature map (SOFM) [21]. This method was compared with the PCA approach and proved to present better results. The disadvantage of this approach is that a SOFM was used along with each channel and thus increasing the computational cost especially when the number of MES channels increases.

This paper presents an approach to re-evaluate the significance of feature selection in MES classification problems. This study is motivated by the fact that feature selection methods that were available when Englehart [19] compared feature selection with feature projection methods on MES classification were not powerful enough to make this comparison fair with respect to the available techniques today. Also, the comparison utilized an FS method that computed the relevance of features only and this in turn was tested against PCA and it was found that PCA gave better performance. This formed a motivation to most of the researches in literature toward the use of only feature projection methods in myoelectric control neglecting the power of feature selection methods.

In this paper, a hybrid swarm based feature selection method is presented as a new dimensionality reduction tool that can be used in myoelectric control problems. The new method, termed as PSOMIFS, is based on modifying the canonical PSO algorithm with the inclusion of a filter measure based on mutual information to estimate the relevance and redundancy properties of the selected subset. The PSO method is employed to aid in computing the interaction property. The justification behind using the PSO method is the parallel computational nature of this method makes it attractive for such type of problems. As a result of this mixture a novel hybrid method is introduced for FS problems and is compared with other dimensionality reduction methods widely used in myoelectric control.

The paper is organized as follows: section 2 describes the canonical PSO and the modified PSOMIFS feature selection algorithm with mutual information. In section 3, the experimental results are presented and analyzed, and finally the conclusion is presented in section 4.

2. Background

2.1 Particle Swarm Optimization

Particle swarm optimization (PSO) is a population based stochastic optimization technique developed by Eberhart and Kennedy in 1995 [22]. It is inspired by social behaviour of bird flocking or fish schooling. In PSO, the potential solutions, called particles, can be considered as simple agents “flying” through a problem space. A particle’s location in the multi-dimensional problem space represents one solution for the problem. Each individual enters the coordinates of its current position into the formula imposed by the required fitness function that provides a quantitative value of the solution’s optimality and measure the error to the estimated target values. Then it moves to a new position and repeat until it is guided gradually to the optimum solution. PSO shares many similarities with evolutionary computation techniques such as Genetic Algorithms (GA). The system is initialized with a population of random solutions and searches for optimal solutions by updating its generations. However, unlike GA, PSO has no evolution operators such as crossover and mutation.

In general, the PSO algorithm consists of several factors which are [23]:

1- A Population of Particles: Each individual in the swarm is called an agent or a particle. The number of particles used ranges between twenty to fifty particles in a population which is far less than that in the usual evolutionary algorithms.

2- Position: The position represents the solution that the specific particle serves in the N-dimensional space.

3- Topology: Every particle has a topology, there are two topologies known those are the global best (gbest) sociometry and the local best (lbest) sociometry. In the gbest topology, every particle knows about the best solution discovered by the entire swarm. In contrast, in the lbest topology every particle remembers the location where it encountered the best solution.

4- Fitness: As in all evolutionary computation techniques there must be a function or method to evaluate the goodness of a position (solution).

The original formula developed by Kennedy and Eberhart was improved by Shi and Eberhart with the introduction of an inertia weight \( w \) that decreases over time, (typically rang from 0.9 to 0.4), to narrow the search that would induce a shift from an exploratory to an exploitative mode [24]. Though the maximum velocity of a particle \( V_{\text{max}} \) was no longer necessary for controlling the explosion of the particles, Shi and Eberhart continued to use it, often setting \( V_{\text{max}} = X_{\text{max}} \) that is the maximum velocity is equalled to the maximum value along the specific dimension, in order to keep the system within the relevant part of the search space. This was found to be a good idea that significantly improves the PSO performance and at the same time it costs very little computationally. During iterations each particle adjusts its own trajectory in the space in order to move towards its best position and the global best according to the following equations:

\[
V_i(t+1) = w \cdot V_i(t) + c_1 \cdot r_1 \cdot (p_{\text{best}_i} - x_i) + c_2 \cdot r_2 \cdot (g_{\text{best}} - x_i)
\]  

(1)
\[ x_j(t+1) = x_j(t) + v_j(t+1) \] (2)

where
- \( i \) is the particle index,
- \( j \) is the current dimension, where \( j \in d \),
- \( d \) is the total number of dimensions,
- \( x_i \) is the current position,
- \( v_i \) is the current velocity,
- \( w \) is the inertia weight,
- \( t \) is the current time step.

\( r1 \) and \( r2 \) are two random numbers uniformly distributed in the range \((0,1)\), \( c1 \) and \( c2 \) are cognitive and social parameters respectively, \( pbest \) is the local best position that is associated with the best fitness value the particle has achieved so far, and \( gbest \) is the global best position that is associated with the best fitness value found among all of the particles.

The personal best of each particle is updated according to the following equation:
\[
pbest_i(t+1) = \begin{cases} 
  \text{pbest}_i(t) & \text{if } f(\text{pbest}_i(t)) \leq f(x_i) \\
  x_i(t) & \text{if } f(\text{pbest}_i(t)) > f(x_i)
\end{cases}
\] (3)

Finally, the global best of the swarm is updated using the following equation:
\[
gbest(t+1) = \arg \min_{pbest_i} f(pbest_i(t+1))
\] (4)

where \( f \) is a function that evaluates the fitness value for a given position. This model is referred to as the \( gbest \) (global best) model.

2.2 PSO based feature Selection with MI.

A hybrid evaluation measure is proposed that is able to estimate the overall performance of subsets as well as the local importance of features. A classification algorithm, i.e. a wrapper evaluation function, is used to measure the performance of subsets. On the other hand, we employ the Mutual Information Evaluation function (MIEF) [20], i.e. a filter evaluation measure, to estimate the local importance of a given feature. The concept of Mutual Information (MI) is widely used in artificial intelligence to measure the stochastic dependence of discrete random variables [25]. This measure is suitable for assessing the information contents. It can be easily defined as the amount of information predictable from a variable when another variable is known.

To address the problem associated with the PSO algorithm, here we are presenting a new hybrid PSOMIFS method for feature selection utilizing a combination of modified PSO and the MIEF measure according to the following steps:

1. **Population Initialization**: The \( d \)-dimensional search space (where \( d \) represents the number of features to be selected) is initialized with the indices of the features in the problem. Although such a step will ends up with non-integer solution, a remedy is proposed is step 4.

2. **Particles Speed Initialization**: The speed of all particles is initialized to zeros.

3. **Specifying a Fitness Function**: The first part of the chosen fitness function is the mean error rate achieved by a specific classifier, Linear Discriminant Analysis (LDA) was employed for this purpose. Added to this is a simple measure based on the Euclidean distance that will give an indication in case of selecting the same feature twice or more by each particle. Since the number of dimensions is actually large, then nothing will prevent two or more dimensions of the PSO from settling at the same indices of the features, thus the value of the fitness function will increase. In this paper, a solution with large fitness value will not be chosen as one seeks the global optimum represented by the term with minimum error value and no redundancy in features indices. The proposed objective function is given by:
\[
f_{total}(i) = MSE(x_i) + f_{subtotal}(i)
\] (5)

where \( MSE(x_i) \) is the mean square error of the solution provided by particle \( i \) along the whole dimensions \( j \in 1 \ldots d \). The other term denoted as \( f_{subtotal}(i) \) will increase if any feature is selected more than once within the same solution. To give a clear picture to the reader, the pseudo code for this function is included as given below.

1. **procedure [OUT] = FSUBTOTAL(xsolution)**
2. \( W = \text{round}(xsolution) \) % conversion to discrete
3. \( Y = \text{dist}(W) \) % Euclidean distance function
4. \( Z = (Y = 0) \) % \( Y \) equal to zero, 1 if yes, else 0
5. \( \text{OUT} = \text{sum}(\text{sum}(Z)) \) % sum the results
6. **end procedure**

Consider the following example with three features only with the current solution represented by a specific particle given as \( xsolution = [1\ 3.23\ 5.293\ 41.431] \), then \( W = [1\ 5\ 41] \), and the value of \( Y \) will be \( Y = [0\ 0\ 40; 40\ 0\ 36; 40\ 36\ 0] \); and \( \text{OUT} = 3 \) as \( \text{OUT} \) is the number of zeros in \( Y \). On the other hand, if a solution of a specific particle is \( xsolution = [1.323\ 41.291\ 41.387] \), then \( W = [1\ 41\ 41] \), and \( Y = [0\ 40\ 40; 40\ 0\ 0; 40\ 0\ 0] \) and \( \text{OUT} = 5 \). Thus the second example is not chosen as an optimum as according to (5) the swarm has to follow the particle with minimum fitness value achieved, that is with the minimum error rates and no redundancy.

4. **Conversion from real-time to discrete optimization**: The canonical particle swarm is used for real-time optimization tasks while the proposed algorithm of feature selection is used for discrete optimization. A simple step is added that is to round the solution presented by each particle towards its nearest integer, thus representing the indices of the features to be tested. Equation (1) is modified to reflect this conversion since no more iterations are needed for a specific particle if the solution provided by that particle is for example \([1.4\ 5.2\ 41.33]\) while the optimum one found is \([1.1\ 5.4\ 41.19]\) as
the both solutions represent the same feature set. The new equation is given by.

\[ v_i(t+1) = w^r \cdot \text{round}(v_i(t)) + c_v \cdot r \cdot (\text{round}(\text{pbest}_i) - \text{round}(x_i)) + c_s \cdot r \cdot (\text{round}(\text{gbest}_i) - \text{round}(x_i)) \]

where \( I(\text{round}(\text{pbest}_i), \text{round}(x_i)) \) is the mutual information between the features represented by the rounded values of \( pbest_i \) and \( x_i \). Similarly, \( I(\text{round}(\text{gbest}_i), \text{round}(x_i)) \) is the mutual information between the features represented by the rounded values of \( gbest_i \) and \( x_i \). In the canonical PSO only the distance between the terms is considered, but in our case we included the mutual information also in the equation. In simple words, the original distance formula was kept to give the particle some speed for movement in the solution space but at the same time it is limited by the statistical distance between the features measured by mutual information.

5- Computation of new Velocity and Position: The whole swarm is divided into two parts, the \( E \) (Elite) particles (whose solutions are the best \( E \) among the whole swarm) and the rest of the swarm (usually \( 5 < E < 15 \)). The rest of the swarm will have their current solutions mixed with the solutions represented by the \( E \) particles according to the MIEF measure, thus randomly replacing parts of their solutions with features indices that the MIEF estimates to increase the whole subset performance. Later the whole swarm will follow the equation of the velocity and position given by (6) and (2). The MIEF measure is defined as:

\[ \lambda = \frac{2}{1 + \exp(-\alpha D)} - 1 \]  

where

\[ D = \min_{f \in \mathcal{F}} \left[ \frac{H(f) - I(f;f_1)}{H(f)} \right] \cdot \frac{1}{|\mathcal{K}|} \sum_{k \in \mathcal{K}} \exp \left[ \beta \left( \frac{I(C;\{f_1,f_f\})}{I(C;f_1)+I(C;f_f)} \right) \right] \]  

The parameters \( \alpha, \beta, \) and \( \gamma \) are constants. \( |\mathcal{K}| \) is the cardinal of \( \mathcal{K} \) that represents the selected feature subset. \( I(C;f_1) \) is the mutual information between feature \( f_1 \) and the class \( C \). \( I(C;\{f_1,f_f\}) \) is mutual information between two features and class. \( H(f) \) is the entropy of \( f \). \( I(f_1,f_f) \) is the mutual information between \( f_1 \) and \( f_f \).

For more information about the MIEF algorithm, the reader can refer to [20]. In this way there is no need for increasing the population size, as the \( E \) particles are enough to lead the other particles to the best solution.

6- Maintaining the Particles within the search space: To avoid particles from crossing their boundaries we followed the hybrid approach presented in [26]. When the particles cross their boundaries, parts of their velocities is absorbed by the boundary during the impact and the particles are then reflected back with lesser velocity of a reversal sign.

7 – Modifying Black listed Solutions: A black list is made of the solutions that have been already encountered by the swarm. In order to decrease the computational cost, only solutions that have not been added to the black list are checked with the wrapper method. If a new solution is found by a specific particle that is already existent in the black list then this solution will be exchanged with a randomly selected one.

3. Experimental Results

In this experiment, the MES datasets used to test the proposed method was acquired from the University of New Brunswick in Canada [16]. The dataset consisted of ten motions associated with three degrees of freedom (DOF’s) of the wrist, two different hand grips, and a rest state. In particular they were: forearm pronation, forearm supination, wrist flexion, wrist extension, radial deviation, ulnar deviation, key grip, chuck grip, hand open, and a rest state. Those datasets were acquired by using 16 electrodes mounted around the human forearm. Each session of the database consisted of two trials or two repetitions of each motion. Six subjects (abbreviated as AW, KS, LH, MW, SM, and WM) were prompted to complete medium force isometric contractions of 5 seconds duration followed by a brief rest period. Each record was 256 ms in duration (256 points sampled at 1024 Hz, pre-filtered between 10-500 Hz using a 4th order Bessel band pass filter with a gain of 2000 and a CMRR greater than 96 db/channel).

In the original research by Levi et al carried out based on using the same datasets [16], different feature extraction techniques were utilized and the performance was compared. The results indicated that (TDAR) features extracted mentioned earlier proved to present the best results when comparing with various techniques including the wavelet packet transform, the reader can refer to [16] for more information. Also in that experiment PCA was applied to the extracted data and the first 40 principal components were chosen to produce a result of 95%-99% in average.

Currently, we followed the same technique mentioned by Levi and extracted the TDAR features from 16 channel producing a total of 176 features (11 feature from each of the 16 channel). In this process features were computed from the MES using a sliding analysis window of 256 ms in length, spaced 32 ms apart. A single feature vector was produced for each analysis window. In specific the extracted feature were the same mentioned in [16, 17], those are: Autoregressive coefficients, root mean square, mean absolute value, integrated absolute value, zero crossings, and slope sign changes. The code for features extraction is available online by Chan [17].

The classification accuracies were computed using LDA classifier for the TDAR features reduced in
dimensionality with PCA, and ULDA [17] (a new method utilized in MES problems recently) as feature projection methods and the proposed PSOMIFS method as a feature selection method across all six subjects’ datasets. Half of the data exactly was used for training and the other half was employed for testing. When projecting the features with ULDA the resultant feature vector will have a dimensionality that is less than the number of classes (10 classes in this problem). ULDA attained minimum classification accuracy with only 9 features while Levi et al [16] pointed out that 40 principal components from PCA were necessary for the same datasets. In comparison, PSOMIFS results with only 9 features are given in Fig.1 without applying the Majority Vote (MV), that is necessary in MES classification problems so as not to overwhelm the robot arm with the classifier decisions. The MV represents a smoothing operation that removes superior misclassification. It is known from literature that applying MV in MES classification problems can achieve an enhancement of about 2% on the final results [27], based on the type of classifier chosen.

In another experiment, the same 9 features were utilized from all the DR methods. It was noted that most of the errors occurred during transitions between classes, which was expected as the myoelectric signal is in an undetermined state between contraction types [27]. Thus we eliminated the transition from the whole datasets during training and testing, and applied the MV on the final results. Table-II shows the results for this experiment.

The results from both cases reveals that feature selection with PSOMIFS can have comparable results with features projection based ULDA, and both methods outperform PCA for dimensionality reduction. When revealing the theory behind those methods, ULDA was found to project features in a way that optimize the class separability by maximizing the ratio of the determinant of the between-class scatter to the determinant of the within-class scatter. The PSOMIFS was actually selecting feature subsets that best interact together to produce powerful results. In comparison with PCA, the PCA was only adding components which carries smaller portion of the complete variance without taking into consideration the class separability.

Although the results for ULDA came slightly higher than PSOMIFS, but this is well justified when taking into considerations the small number of features employed. As a feature selection method the real power of such algorithm can appear when choosing few more features, thus we decided to employ 15 features from both the PSOMIFS and PCA only, as this wasn’t possible for ULDA since the resultant dimensionality is limited to less than the number of classes. Table III reveals the hit rates with both methods with transitions removed from data and with MV applied on the final results. The results have shown that powerful feature selection techniques such as the PSOMIFS algorithm can have a profound impact on enhancing the classification accuracy for MES recognition. The results have shown also that PSOMIFS outperformed the PCA and achieved the same results achieved by ULDA. This in turns proves that the current start of the art swarm intelligence based techniques can achieve very powerful results in the problem of features selection. The results also shows that effect of applying feature selection in MES classification can be as good as feature projection methods or better.

![Figure 1](image-url1) Classification results without MV with 9 features only

![Figure 2](image-url2) Classification results without transitions and with MV with 9 features only

4. Conclusion

This paper proposed a new swarm based feature selection algorithm utilizing both PSO and mutual information. The proposed PSOMIFS method was tested in the myoelectric control problem on six subject’s datasets. It was also compared with other state of the art and conventional feature projection methods (ULDA, and PCA respectively). Results indicated the high performance of the proposed method in a problem of ten classes. The results also indicated that PSOMIFS and ULDA can both achieve comparable results while both of them highly outperform PCA which needs at least twice the number of features selected with those methods to achieve comparable results. More experiments and tests should be done with transient and steady state MES data, as this experiment utilized steady-state data only. Also the
PSOMIFS should be tested with other types of features like the WPT based features which are known to be of high variance in addition to the high variance of the WPT thus complicating the problem. Research continues and more experiments are to be published soon.

![Figure 3 Classification results without transitions and with MV with 15 features only](image)

Acknowledgements

The authors would like to acknowledge the support of Levi Hargrove and Dr. Bernard Hudgins from the University of New Brunswick for providing the MES datasets.

References