Swarm based Fuzzy Discriminant Analysis for Multifunction Prosthesis Control

Rami N. Khushaba, Ahmed Al-Ani, and Adel Al-Jumaily

Faculty of Engineering, University of Technology, Sydney, Australia
{rkhushab, ahmed, adel}@eng.uts.edu.au

Abstract. In order to interface the amputee's with the real world, the myoelectric signal (MES) from human muscles is usually utilized within a pattern recognition scheme as an input to the controller of a prosthetic device. Since the MES is recorded using multi channels, the feature vector size can become very large. In order to reduce the computational cost and enhance the generalization capability of the classifier, a dimensionality reduction method is needed to identify an informative moderate size feature set. This paper proposes a new fuzzy version of the well known Fisher's Linear Discriminant Analysis (LDA) feature projection technique. Furthermore, based on the fact that certain muscles might contribute more to the discrimination process, a novel feature weighting scheme is also presented by employing Particle Swarm Optimization (PSO) for the weights calculation. The new method, called PSOFLDA, is tested on real MES datasets and compared with other techniques to prove its superiority.

Key words: Discriminant Analysis, Myoelectric Control, Particle Swarm Optimization

1 Introduction

The myoelectric signal (MES), also known as the Electromyogram (EMG), is a one dimensional non-stationary signal that carries the distinct signature of the voluntary intent of the central nervous system. It is usually recorded in a noninvasive scheme utilizing a set of surface electrodes mounted on the human forearm. One of the most important application of the MES is its use in controlling prosthetic devices functioning as artificial alternatives to missing limbs. Advances in myoelectric signals studies revealed that the MES exhibits different temporal structure for different kinds of the arm movements. This in turn facilitated the use of a pattern recognition based myoelectric control strategies for prosthetics control. To this end, a wide set of pattern recognition methods were proposed in the literature to produce a computationally efficient and accurate MES recognition systems [1].

In order to capture the complete muscle activity, a multi channel approach is usually utilized when measuring the MES signal to capture novel motor information from different muscles. However, this will increase the number of extracted

features (variables that describe these movements) and hence it will increase the learning parameters of the classifier and may degrade its performance. A straight forward solution to these problems is to project the data onto low-dimensional subspaces to extract the most significant features. Many feature projections techniques were used in myoelectric control with the aim to produce a statistically uncorrelated or independent feature set, a desirable goal in any pattern recognition system. Various approaches in dimensionality reduction were utilized in myoelectric control. These include, principal component analysis (PCA), a well-known feature projection method which, according to [2], achieved good performance in the myoelectric control problem , a combination of PCA and a self organizing feature map (SOFM), proving better results than PCA alone, according to [3], linear discriminant analysis (LDA) based feature projection, proving better results than PCA with SOFM, according to [4].

Although LDA is a well known projection technique, but there are many limitations with the classical LDA [5]. The first is that it requires the scatter matrices to be nonsingular, while in real world problems they can be singular. The second limitation with LDA is that it treats all the data points equivalently where as in the real world problems each sample may belong to each of the different classes with to certain degree. Finally, classical LDA pays no attention to the decorrelation of the data, which is a desirable property in many applications. One possible approach to overcome the first and the third problems was utilized in myoelectric control, this is based on the use of the uncorrelated linear discriminant analysis (ULDA) that requires the reduced features to be statistically uncorrelated with one another [6].

As a variation to the ULDA approach which is based on Singular Value Decomposition (SVD) that is known to be expensive in terms of time and memory requirements for large datasets, this paper proposes a new mixture of fuzzy logic and discriminant analysis as a novel dimensionality reduction technique. The proposed method aims to reduce the dimensionality of the extracted feature set and cluster features, such that the classification accuracy is improved. Due to the fact that most of the biosignals generated by the human body tend to produce patterns that are fuzzy in nature (i.e., belongs to different classes with certain degrees), then the incorporation of the concept of fuzzy memberships is required to reduce the effect of overlapping and outliers points. The new method, called PSOFLDA, unlike the current available variations to Fisher's linear discriminant analysis (LDA), accounts for the different contribution of different muscles into the discrimination process. Thus it assumes that certain features are more important than others. In order to reflect this importance, a novel feature weighting scheme is introduced employing Particle Swarm Optimization (PSO) technique for the weights calculation. Also in order to overcome the singularity problem, a regularization parameter is included within each particle (i.e., member of the population).

This paper is structured as follows: Section 2 explains the proposed methodology. The experimental results are given in section 3. Finally the conclusion is given in section 4.

2   Methodology

In this section, first the theory behind the proposed fuzzy discriminant analysis is introduced. Secondly, an introduction to the Particle Swarm Optimization (PSO) technique is presented as one possible way to reach the near optimal solution.

2.1   Weighted Fuzzy Discriminant Analysis

Consider a classification problem with c classes, in which the data set of labeled training samples is given as:

$$S = \{(x_1, y_1), (x_2, y_2), ..., (x_l, y_l)\} \subseteq (X, Y)^l \tag{1}$$

Where X is the input space and Y is the output space. $X \subseteq \Re^n$, and l is the number of samples. Each training point x_i originally belongs to one of the c classes and is given a label $y_i \in \{1, 2, 3, \ldots, c\}$ for $i = \{1, 2, 3, \ldots, l\}$. The goal is to find an optimal hyper-plane using the training samples that can recognize the test points, i.e., the classifier will have a good generalization capability. In PSOFLDA each point, x_i, belongs to each of the c classes with a certain degree of membership. The fuzzy within class scatter matrix S_W, fuzzy between class scatter matrix S_B, and the fuzzy total class scatter matrix S_T are given as follows:

$$S_W = \sum_{i=1}^{c} \sum_{k=1}^{l} u_{ik}^m (x_k - v_i)(x_k - v_i)^T \odot (ww^T) \tag{2}$$

$$S_B = \sum_{i=1}^{c} \sum_{k=1}^{l} u_{ik}^m (v_i - \overline{x})(v_i - \overline{x})^T \odot (ww^T) \tag{3}$$

$$S_T = \sum_{i=1}^{c} \sum_{k=1}^{l} u_{ik}^m (x_k - \overline{x})(x_k - \overline{x})^T \odot (ww^T) \tag{4}$$

where u_{ik} is the membership of pattern k in class i, m (given that $m > 1$) is the fuzzification parameter, x_{kj} is the value of the k'th sample across the j'th dimension, v_i is the mean of the patterns belonging to class i, and v_{ij} is its value across the j'th dimension. $\odot$ refers to the *Hadamard product* operation, w is the weight vector associated with all features, i.e., $w = \{w_1, w_2, ..., w_f\}$, where f is the total number of features. $\overline{x}$ is the mean of the training samples which is given in Eq.(5) below.

$$\overline{x} = \frac{1}{l} \sum_{k=1}^{l} x_k \tag{5}$$

In this paper, the value of the membership u_{ik} is calculated using a possibilistic fuzzy clustering approach. The cost function of the possibilistic clustering approach is adopted from [7], as given in Eq.(6) below.

$$J(\theta, U) = \sum_{k=1}^{l} \sum_{i=1}^{c} u_{ik}^{m}(x_k - \theta_i)^2 + \sum_{i=1}^{c} \eta_i \sum_{k=1}^{l}(1 - u_{ik})^m \tag{6}$$

where θ_i is the i'th cluster center, η_i are positive constants that are suitably chosen. The first term in Eq. (6) is the same objective function used in probabilistic clustering approach, while the second term is added to reduce the effect of outliers. In order to find the membership values from the above equation, then the values of the clusters centers are needed. A direct way would be to differentiate Eq. (6) with respect to θ_i, but this in turn would cancel the second term leaving only the first term. A general look at the first term of Eq. (6) reveals that it represents the classical within class scatter matrix S_W given in Eq. (2) if the weight is removed. Thus applying the values of the clusters means ensures that the objective function given by Eq. (6) would settle at a global optimum value. Then in order to compute the membership values, a differentiation of the resultant function with respect to u_{ik} needs to be done as follows.

$$\frac{\partial J(\theta, U)}{\partial u_{ik}} = m u_{ik}^{m-1}(x_k - v_i)^2 - m\eta_i(1 - u_{ik})^{m-1} = 0 \tag{7}$$

This would in turn result in the following function

$$u_{ik} = \frac{1}{1 + \left(\frac{(x_k - v_i)^2}{\eta_i}\right)^{\frac{1}{m-1}}} \tag{8}$$

The values of η_i, where $i = \{1, 2, 3, \dots, c\}$ were chosen to be equal to the maximum distance between the samples belonging to that class and the class center.

After computing all the variables, PSOFLDA finds the vector G that would maximize the ratio of the between class scatter matrix to the within class scatter matrix by solving the following equation:

$$G = arg \ \max_{G} \ trace\left((G^T S_W G)^{-1} G^T S_B G\right) \tag{9}$$

The solution can be readily computed by applying an Eigen-decomposition on $S_W^{-1} S_B$, provided that the within class scatter matrix S_W is nonsingular. In this paper, we are using a regularized version of S_W given by $S_W = S_W + zI$, for some $z > 0$ that is included in the particle representation of the weights, where I is an identity matrix. In this way the scatter matrix is guaranteed to be nonsingular. Since the rank of the between class scatter matrix is bounded from above by c - 1, there are at most c - 1 discriminant vectors by PSOFLDA.

2.2   PSO based Weight Optimization

One possible solution for finding the best values of the weights is to employ evolutionary algorithms, or EAs. Powerful EA algorithms include genetic algorithm (GA) and Particle Swarm Optimization (PSO). PSO is an effective continuous

function optimizer as it encodes the parameters as floating-point numbers and manipulate them with arithmetic operators. By contrast, GAs are often better suited for combinatorial optimization because they encode the parameters as bit strings and modify them with logical operators. There are many variants to both approaches, but because PSO is primarily a numerical optimizer, the PSO is considered in this paper.

Particle swarm optimization, is a population based stochastic optimization technique developed by Eberhart and Kennedy in 1995 [8]. It represents an example of a modern search heuristics belonging to the category of *Swarm Intelligence* methods. PSO mimics the behavior of a swarm of birds or a school of fish. The swarm behavior is modeled by particles in multidimensional space that have two characteristics: position (p) and velocity (s). These particles wander around the hyper space and remember the best position that they have discovered. A particle's position in the multi-dimensional problem space represents one solution for the problem. They exchange information about good positions to each other and adjust their own position and velocity with certain probabilities based on these good positions. The original formula developed by Kennedy and Eberhart was improved by Shi and Eberhart with the introduction of an inertia weight ϖ that decreases over time, (typically from 0.9 to 0.4), to narrow the search that would induce a shift from an exploratory to an exploitative mode. Though the maximum velocity of a particle (s_{max}) was no longer necessary for controlling the explosion of the particles, Shi and Eberhart continued to use it, often setting $s_{max} = p_{max}$ that is the maximum velocity is equaled 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:

$$s_{ij}(t+1) = \varpi s_{ij}(t) + c_1 r_1 (pbest_{ij} - p_{ij}) + c_2 r_2 (gbest_{ij} - p_{ij}) \qquad (10)$$

$$p_{ij}(t+1) = p_{ij}(t) + s_{ij}(t+1) \qquad (11)$$

Where
i: is the particle index
j: is the current dimension under consideration
p_i: is the current position,
s_i: is the current velocity
ϖ: is the inertia weight
t: is the current time step
r_1 and r_2 are two random numbers uniformly distributed in the range (0,1), c_1 and c_2 are cognitive and social parameters respectively, $pbest_i$ is the local best position, the one associated with the best fitness value the particle has achieved so far, and $gbest_i$ is the global best position, the one 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:

$$pbest_i(t+1) = \begin{cases} pbest_i(t) & \text{if } (pbest_i(t)) \leq f(p_i) \\ p_i(t) & \text{if } (pbest_i(t)) > f(p_i) \end{cases} \tag{12}$$

Finally, the global best of the swarm is updated using the following equation:

$$gbest(t+1) = arg \min_{pbest_i} f(pbest_i(t+1)) \tag{13}$$

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. In this paper, each particle will represent a vector whose elements are the weights assigned to each feature plus the regularization parameter z. The idea here is to generate a new weight vector by utilizing a set of particles that wander through the solution space searching for the best possible representation achieving the minimum error rates. In such a system, the fitness function was chosen to be the error rates achieved by a suitable classifier. The details of the classifiers chosen will be given in the experiments section.

3 Experiments and Results

In order to present a fair comparison with the available techniques, we include many of them in the experiments. The details of the experiments carried on are listed below:

- **Comparison with other methods**: The PSOFLDA will be compared against two groups of other techniques: the first has already been applied into myoelectric control like ULDA [5], and PCA [2]. The second group include technique that were not used within the myoelectric control problems, like Orthogonal Linear Discriminant Analysis (OLDA) [9], and Fuzzy Discriminant Analysis (FLDA) [10]. These were included because they represent new variations to Fisher's LDA.
- **Testing method employed**: The general testing scheme employed is a three way data split. The dataset utilized is divided into three sets: training, validation, and testing. An initial projection matrix is calculated based on the training set. Then a validation set is used in order to optimize the weights to produce the optimum projection matrix that can minimize the mean of the training and validation errors. Finally a completely unseen testing set is utilized to measure the generalization capability of the proposed system.
- **Parameters of PSO**: Specifically the following parameters values were used: maximum number of generations, 100; maximum velocity s_{max}: 20% of the range of the corresponding variable; maximum value along a specific dimension $p_{max} = 1$ and minimum $p_{min} = 0$; w decreases linearly from 0.9 to 0.4; and acceleration constants c_1, c_2 are set to 2.0.

The MES dataset utilized in this research is the same one that was originally collected and used by Chan et al [6]. Eight channels of surface MES were collected from the right arm of thirty normally limbed subjects (twelve males and eighteen females). Each session consisted of six trials. Seven distinct limb motions were used, hand open (HO), hand close (HC), supination (S), pronation (P), wrist flexion (WF), wrist extension (WE), and rest state (R). Data from the first two trials were used as training set and data from the remaining four trials were divided equally into two trials for validation (trails 1 & 2) and two trails for testing (trails 3 and 4).

As a first part of the MES pattern recognition system, two sets of features were extracted from the original dataset in order to test the performance of the proposed method with different feature extraction techniques. The first set of extracted features included a combination of the first four autoregressive (AR) coefficients and the root mean square value (Time-Domain (TD) feature) as the feature vector (dimensionality is 40 = 8 channels 5 features/channel). This feature set was referred to as the **TDAR** feature set. The second feature set extracted included the mean of the square values of the wavelet coefficients using a Symmlet wavelet family with five levels of decomposition (dimensionality is 48 = 8 channels 6 features/channel). This feature set was referred to as **WT** feature set. The analysis window size was 256 msec. Data that were 256 msec before or after a change in limb motion were removed from the training set to avoid transitional data. As a dimensionality reduction part, all of the following five methods: PSOFLDA, ULDA, OLDA, FLDA, and PCA were utilized to compare their performance with different feature sets. The final step of the MES recognition system involves a suitable classifier that can be chosen at the disposal of the designer. In the current experiments a Linear Discriminant Analysis (LDA) classifier was chosen. The advantage of this classifier is that it does not require iterative training, avoiding the potential for under- or over-training [6].

The classification results averaged across thirty subjects (with one standard deviation) using both the TDAR and the WT feature sets reduced in dimensionality with PSOFLDA, ULDA, OLDA, FLDA, and PCA are shown in Fig.1. The number of extracted features from all methods was set to $c - 1$, where c is the number of classes, since the discriminant analysis based techniques usually ends up with c - 1 features. The results shown for both the validation and the testing sets were given first without post processing (referred to as Initial), then with a majority vote (MV) as a post processing step, followed by the transitional data between classes removed (NT), and finally with both majority vote and the removal of the transitional data (MV+NT). The results for both the validation and testing sets are given in the Table-1.

It is clear from the results that the PSOFLDA was able to outperform all other methods. This is due to the fact the PSOFLDA is assigning higher importance to good features compared to those that are less useful. At the same time, the using of the classification accuracy as a judgment criterion on the weights values moved the projection matrix closer toward the optimal projection matrix than all other techniques. Also the PSOFLDA assigns lower fuzzy membership

Table 1. Classification Results Achieved by Different Methods

Feature Set	Divisions	PSOFLDA	FLDA	ULDA	OLDA	PCA
WT	Validate	96.26	93.28	93.35	93.35	89.90
	Test	94.60	92.32	92.44	92.44	88.21
TDAR	Validate	95.02	92.25	92.38	92.38	83.51
	Test	93.68	91.67	91.79	91.79	81.93

values to the outliers points, thus reducing their effect. Another issue to be mentioned here is that with all of the feature projection techniques, the WT features achieved higher accuracies than that achieved using the simple TDAR features. But from the computational cost point of view the performance of the system with the TDAR features is still highly accepted.

In order to provide a rigorous validation or comparison with existing techniques for dimensionality reduction, the confusion matrix for all the subjects was also computed for the different feature sets. A plot of the diagonal values of the confusion matrices (class wise classification accuracy) with both the TDAR and the WT feature are presented in Fig.2 respectively, each with the validation and testing sets results. All the results indicate that there were more significant enhancements when applying the PSOFLDA method than that of the other techniques.

4   Conclusion

In this paper, a novel feature projection technique based on a mixture of fuzzy logic and Fisher's LDA was developed. Unlike the typical variations to LDA, The new technique assigned higher importance to good features compared with others. The importance was based on a weighting scheme that was optimized with PSO technique. This in turn caused the PSOFLDA's projection matrix to be closer to the optimal. The proposed PSOFLDA technique was fairly compared with other techniques like FLDA, ULDA, OLDA, and PCA proving to present better results on real MES datasets. This in turn proves the ability of the proposed technique in enhancing the performance of the multifunction myoelectric hand control system.

References

1. Asghari Oskoei, M., Hu, H.: Myoelectric Control Systems–A Survey. Biomedical Signal Processing and Control, 2, 275–294 (2007)
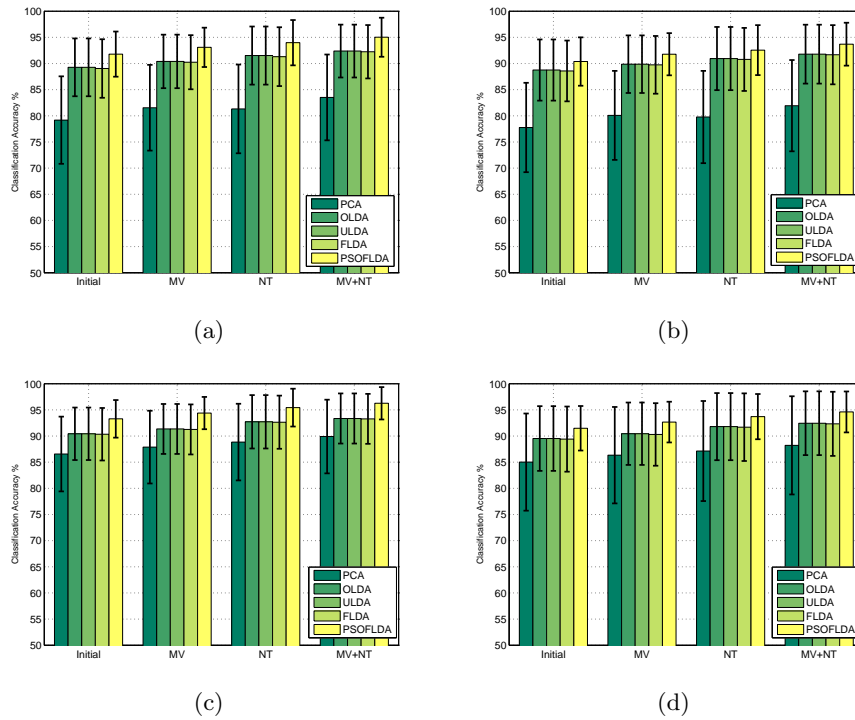
Fig. 1. Classification accuracies using different feature sets averaged across 30 subjects with different dimensionality reduction techniques (a) Using the TDAR validation set and (b) Using the TDAR testing set (c) Using the WT validation set and (d) Using the WT testing set

2. Englehart, K., Hudgins, B., Parker, P. A., Stevenson, M.: Time-Frequency Representation for Classification of the Ttransient Myoelectric Signal. In: The 20th EMBS Annual International Conference, pp. 2627–2630. China (1998)

3. Chu, J. U., Moon, I., Mun, M. S.: A Real-Time EMG Pattern Recognition System based on Linear-Nonlinear Feature Projection for a Multifunction Myoelectric Hand. IEEE Transactions on Biomedical Engineering, 53, 2232-2239 (2006)

4. Chu, J. U., Moon, I., Mun, M. S.: A Supervised Feature Projection for Real-Time Multifunction Myoelectric Hand Control. In:The 28th IEEE EMBS Annual International Conference, pp. 2417–2420. New York City (2006)

5. Ye, J., Janardan, R., Li, Q., Park, H.: Feature Reduction via Generalized Uncorrelated Linear Discriminant Analysis. IEEE Transactions on Knowledge and Data Engineering, 18(10), 1312–1322 (2006)

6. Chan, A. D. C., Green, G. C.: Myoelectric Control Development Toolbox. In. Proceedings of The 30'th Conference of the Canadian Medical & Biological Engineering Society, Toronto, ON (2007)

7. Oliveira, J. V. d., Pedrycz, W.: A Comprehensive, Coherent, and in Depth Presentation of the State of the Art in Fuzzy Clustering. John Wiley & Sons Ltd. (2007)

(a)

(b)

(c)

(d)
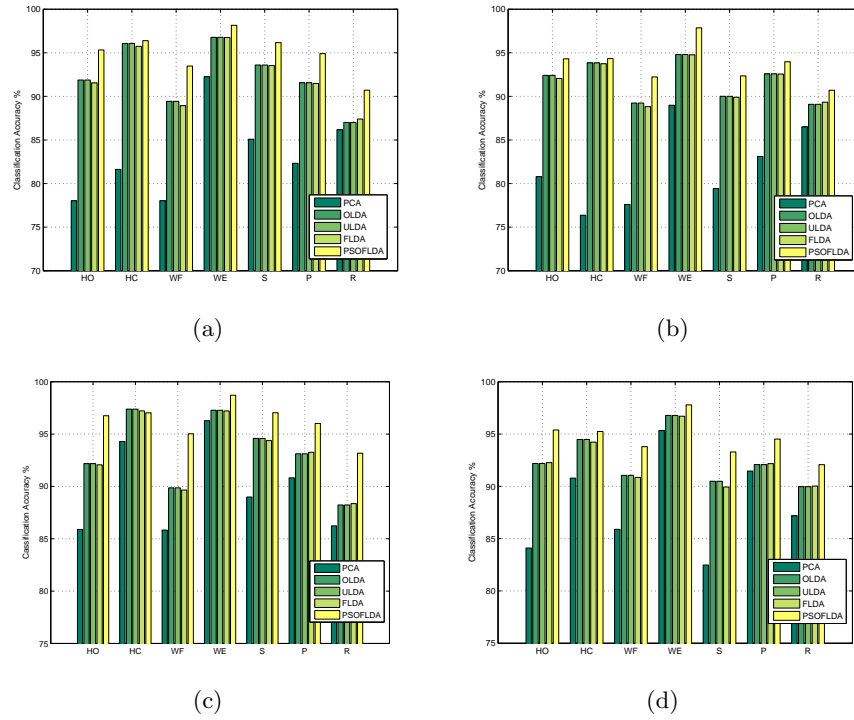
Fig. 2. Diagonal values of the confusion matrix averaged across 30 subjects using the proposed PSOFLDA in comparison with FLDA, ULDA, OLDA, and PCA (a) Using the TDAR validation set and (b) Using the TDAR testing set (a) Using the WT validation set and (b) Using the WT testing set

8. Kennedy, J., Eberhart, R. C., Shi, Y.: Swarm Intelligence. The Morgan Kaufmann Series in Artificial Intelligence, Morgan Kaufmann Publishers, London (2001)
9. Ye, J.: Characterization of a Family of Algorithms for Generalized Discriminant Analysis on Undersampled Problems. Journal of Machine Learning Research, 6, 483–502 (2005)
10. Chen, Z. P., Jiang, J. H., Li, Y., Liang, Y. Z., Yu, R. Q.: Fuzzy Linear Discriminant Analysis for Chemical Data Sets. Chemometrics and Intelligent Laboratory Systems, 45 (1-2), 295–302 (1999)