

Automatic Misclassification Rejection for LDA classifier using ROC curves

Radhika Menon, Gaetano Di Caterina, Heba Lakany, Lykourgos Petropoulakis, Bernard A. Conway and John J. Soraghan

Abstract— This paper presents a technique to improve the performance of an LDA classifier by determining if the predicted classification output is a misclassification and thereby rejecting it. This is achieved by automatically computing a class specific threshold with the help of ROC curves. If the posterior probability of a prediction is below the threshold, the classification result is discarded. This method of minimizing false positives is beneficial in the control of electromyography (EMG) based upper-limb prosthetic devices. It is hypothesized that a unique EMG pattern is associated with a specific hand gesture. In reality, however, EMG signals are difficult to distinguish, particularly in the case of multiple finger motions, and hence classifiers are trained to recognize a set of individual gestures. However, it is imperative that misclassifications be avoided because they result in unwanted prosthetic arm motions which are detrimental to device controllability. This warrants the need for the proposed technique wherein a misclassified gesture prediction is rejected resulting in no motion of the prosthetic arm. The technique was tested using surface EMG data recorded from thirteen amputees performing seven hand gestures. Results show the number of misclassifications was effectively reduced, particularly in cases with low original classification accuracy.

I. INTRODUCTION

Myoelectric control of upper limb prosthetic devices has evolved over the last few decades and continues to be researched [1]. The use of surface electromyography (sEMG) signals provides a non-invasive means to identify the user's motion intention, whereby restoring lost arm function to amputees. Although several prosthetic devices are commercially available, there is yet much to be done to develop a multi-articulating, intuitive, comfortable, light weight, cost effective and aesthetically designed prosthetic device, which would truly replicate the human arm. This study aims to enhance the functionality of a myoelectric prosthetic device, which is an important factor affecting user rejection rate [2], by minimising the formation of unwanted gestures.

Pattern recognition based myoelectric control of prostheses has the potential for superior functionality by providing an intuitive and dexterous control [1]. It is based on the hypothesis that a distinct EMG signal results in the formation of a specific forearm gesture. But physiologically, this is not the case because EMG signals are complex and time-varying. Hence classifiers have been trained to

recognise specific features in EMG signals and predict the associated gesture. A multitude of classifiers have been employed for interpreting hand, wrist and forearm gestures, such as Linear Discriminant Analysis (LDA), Support Vector Machines, Artificial Neural Networks, Gaussian Mixture Model, Hidden Markov model, k-nearest neighbour etc. The LDA classifier has been widely used in this field because it provides good classification accuracy despite its simplicity of implementation and low training time [3]. Several researchers have extended the LDA in order to overcome the complex nature of EMG signals, which poses significant challenges to classification. Chen *et al.* [4] developed a 'self-enhancing LDA' classifier to account for the non-stationary property of EMG that can be attributed to muscle fatigue, electrode shift, sweating etc. They used a continuous classification model and retrained the classifier by updating the mean vector and covariance matrix of the class into which the latest EMG pattern was classified. This method served the purpose of enlarging the training data set to improve classification. In [5] a similar strategy was used and feature vectors were replaced in a part of the training data set with new ones, leaving the other part untouched to preserve classifier stability.

Another challenge in pattern recognition of human arm motion is the manifold gestures possible in a real world scenario, while in a research environment only a limited number of gestures can be used to train the classifier. A robust classifier must be capable of identifying and discarding an unknown pattern, instead of misclassifying it as a trained class. False motions of the prosthetic arm are frustrating for the user as corrections must be enforced, whereas, absence of motion would go unnoticed [6]. Post classification strategies using ensemble or majority voting to reject motion decisions generated by an LDA classifier are reported in [6, 7, 8]. An alternative technique has been put forth by Scheme *et al.* [9], where a confidence score, computed from the posterior probability of LDA classification, is compared to a fixed threshold. The threshold is determined empirically to provide a reasonable trade-off between classification accuracy and the frequency of motion rejections. Amsuss *et al.* [10] has trained an artificial neural network to detect misclassifications of LDA, the base classifier.

The technique presented in this paper aims to reject misclassifications generated by an LDA classifier using an approach similar to that proposed in [9]. However, in this paper a confidence threshold is calculated automatically for each individual gesture class, with the help of receiver operating characteristic (ROC) based curves. An output from the classifier is rejected if its confidence level is below the threshold associated with the given class, thus preventing the

R. Menon, G. Di Caterina*, L. Petropoulakis and J.J. Soraghan are with the Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, United Kingdom (*corresponding author - e-mail: gaetano.di-caterina@strath.ac.uk).

H. Lakany and B.C. Conway are with the Department of Biomedical Engineering, University of Strathclyde, Glasgow, United Kingdom.

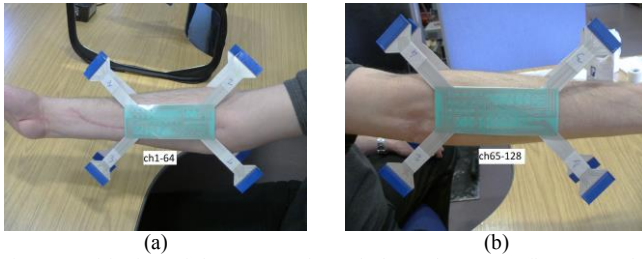


Fig. 1. Positioning of the two 64-channel electrodes on (a) flexor muscles and (b) extensor muscles.

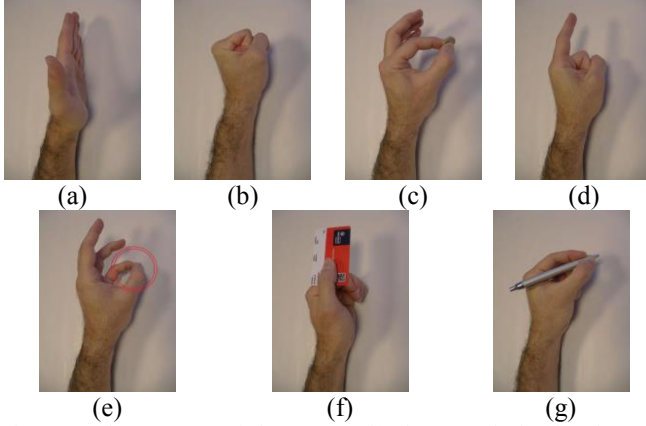


Fig. 2. Hand gesture recorded: (a) open, (b) close, (c) pinch, (d) point, (e) thumb-pinky, (f) lateral grip, (g) tripod

prosthesis from performing a possibly unwanted gesture. The proposed technique has been tested using surface EMG data recorded from thirteen amputees performing seven hand gestures. Experimental results demonstrate that such a technique can effectively reduce the number of misclassifications, particularly in cases with low original classification accuracy.

The remainder of the paper is organized as follows. Section II describes recording protocol, data processing and threshold computation. Section III reports experimental results and discussions, while Section IV concludes the paper.

II. METHODS

A. Data Acquisition

Eight trans-radial amputees (including three congenital) and five partial hand amputees participated in an experimental study, which was approved by the University of Strathclyde Ethics Committee. Surface electromyography (sEMG) signals were acquired from the forearms of the volunteers with their informed consent. Two high density electrodes of 64 channels each were used, one being placed on the flexor muscles and the other on the extensor muscles, as illustrated in Fig. 1. The electrodes were connected to an EMG-USB2 bioelectrical signal amplifier from OT Bioelectronica, and the EMG signals were recorded using a floating monopolar configuration. A sampling frequency of 2048 Hz was used, and built-in hardware filters of 3 Hz, for Low-Pass, and 900 Hz for High-Pass, were adopted. The gain used was either 500, 1000 or 2000, depending on the individual participant's EMG signal strength.

The experimental protocol during recording was as follows. Participants were seated comfortably with their elbows flexed and forearms placed on a desk in front of them. They were prompted by written messages displayed on a monitor in front of them to perform hand gestures (open, close, pinch, point, thumb-pinky, lateral grip, tripod) as in Fig. 2, in random order with five repeats per gesture. A mirror box was positioned between their arms (in the case of unilateral amputees), which enabled them to visualise the intended formation of gestures using their amputated arms, by viewing the reflection of the gesture performed by the contra-lateral arm. All gestures were formed from a relaxed position in a three second time interval, followed by the gesture being held from four to five seconds randomly, and finally returning to rest, which was held from three to four seconds randomly. Hence the length of one gesture repeat was between ten and twelve seconds.

B. Data Processing

The EMG data from each participant is processed offline in Matlab, by first segmenting the signal, followed by feature extraction from individual blocks of the segmented data, and finally training an LDA classifier to recognise EMG patterns associated with seven forearm gestures, and testing it.

The myoelectric signal from each of the 128 channels pertaining to one gesture repeat is segmented discarding the first 1.5s of gesture formation time and the relaxed time. The EMG signals are filtered in software, with three 3rd order Butterworth filters, with the following characteristics: 1) a 48-52 Hz stop-band to remove the 50 Hz main; 2) a 20 Hz high-pass to remove motion artifacts; 3) a 400 Hz low-pass to discard unwanted frequency content and further prevent aliasing. Then time domain features are extracted from the resultant EMG signal of length 5.5s-6.5s using the overlapped windowing technique (Englehart and Hudgins, 2003) with an analysis window of length 200ms and 50% overlap. The window length lies within the optimal range for good controllability of the prosthetic arm (Smith et al, 2011). For each analysis window the mean absolute value (MAV), slope sign change (SSC), waveform length (WFL) and zero crossings (ZC) are computed (Hudgins et al., 1993). Features from the same analysis window and pertaining to different channels were concatenated to form a single feature vector. Given 4 features and N channels, the length of a feature vector is therefore $L=4N$.

Four of the 5 repeats of each gesture (~80% of the data) are selected randomly to train the classifier. The rest of the data is used for testing. This process is repeated a thousand times and, at each iteration, the classification accuracy is calculated as well as the average classification accuracy for all completed iterations. The process converges when the difference between the average classification accuracies for two consecutive iterations is less than 0.0005. This usually occurs around the 500th iteration.

C. Threshold Computation

Along with the classification result, the LDA classifier provides in output the posterior probability of the classification [9]. This value is compared with a threshold to determine if the classified result is usable or should be discarded. Hence the posterior probability acts a level of confidence. For each gesture class, one can plot a histogram

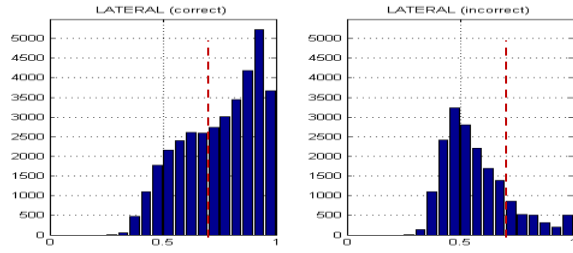


Fig. 3. Histogram of correct (left hand-side) and incorrect (right hand-side) classifications for the 'lateral' gesture. The red dashed line is the best threshold in this case.

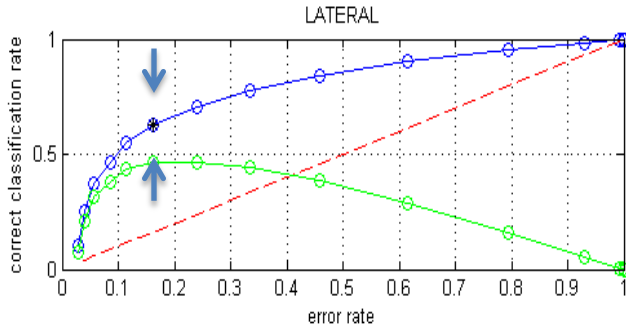


Fig. 4. Threshold computation using ROC curves: in blue is the correct rate plotted against the sorted error rate; the red dashed line is the ROC space diagonal; in green is the difference between blue line and diagonal. The selected point is in correspondence of the black star on the blue line.

of the number of times in which that class has been selected by the classifier, against different levels of the posterior probability. Therefore, two histograms can be computed, i.e. one for correct classifications and the other for incorrect decisions by the classifier. Fig. 3 illustrates such two histograms for a sample gesture, 'lateral' in this case.

The threshold to be computed can be imagined as a vertical line at the same location, in both histograms: all the classifications to the left of such a line are discarded as erroneous; while all classifications to its right are deemed as correct. In the 'incorrect' histogram, one would want to set the threshold high, in order to discard a high number of misclassifications. On the contrary, in the 'correct' histogram, the threshold should be low, to retain as many correct classifications as possible. The wanted value of threshold must be such that the loss in correct classifications is at its lowest compared to the rejections of misclassifications. The computation of the threshold is done automatically per gesture using the following technique.

For possible values of threshold ranging from 0 to 0.95 in 0.05 increments, correct classification rate and error rate are computed, therefore obtaining two vectors. The error rate vector is then sorted in ascending order, while the correct rate vector is rearranged in the same order as the error rate vector. Then the correct rate is plotted against the sorted error rate, to obtain an ROC curve (blue line in Fig. 4), and the diagonal of the ROC space (red dashed line) is subtracted from the plotted curve. The wanted value for the threshold is selected as the one corresponding to the highest correct classification rate on the difference curve (green line). In case of multiple points on the difference curve having the same highest correct rate, the right-most one is chosen, as

TABLE 1. COMPARISON OF OVERALL CLASSIFICATION ACCURACY AND ERROR BEFORE AND AFTER APPLYING ROC ANALYSIS FOR PARTICIPANT TT1.

	Before ROC analysis	After ROC analysis
Accuracy(%)	41.68	29.82% (-11.86)
Error(%)	58.32	34.71% (-23.61)

TABLE 2. CLASSIFICATION ACCURACIES ACQUIRED FROM CONFUSION MATRICES FROM PARTICIPANT TT1.

Gestures	Classification accuracies (%)		
	Before ROC analysis	After ROC analysis and introducing 'empty' class	After ROC analysis and 'empty' class removed
Close	65.5	54.6	86.1
Lateral	20.8	15.7	21.5
Open	54.5	24.5	61.9
Pinch	16.2	16.2	16.2
Pinky	64	45.9	94.6
Point	54.6	39.1	63
Tripod	14.6	11.9	18.2
Overall	41.7	29.8	46.2

this point corresponds to a higher correct rate on the original ROC curve. For the example illustrated in Fig. 3 and Fig. 4, the best threshold value is 0.7, corresponding to the red dashed line in Fig. 3.

In order to handle the case where the prediction of the LDA classifier is rejected as a misclassification, an additional gesture class is introduced and labelled as 'empty' gesture. Hence, if the confidence level fails the threshold check, the EMG pattern would be assigned to the 'empty' class and the prosthetic device would not form a new gesture.

III. RESULTS AND DISCUSSION

The confidence scores of predicted gestures and the classification accuracy and error of the LDA classifier with and without the use of the ROC based technique are analysed offline using sEMG data from all thirteen participants. Data analysis is explained with the help of results from one participant who is a traumatic trans-radial amputee (herein referred to as participant TT1). This dataset is an example of an originally poor classification with classification accuracy 41.68% and error 58.32% as reported in Table 1.

First, the confusion matrix for the LDA classification of the seven gesture classes is generated. The overall and per-gesture classification accuracies are provided in Table 2 respectively. Then histograms of the posterior probabilities of correct and incorrect predictions are plotted to visualise the confidence levels. Subsequently, ROC analysis is applied for each gesture, generating the ROC curves. The confusion matrix is computed again using an additional 'empty' class into which ROC patterns not passing the threshold check are classified. The inclusion of the empty class causes a decrease in the overall classification accuracy and error as seen in Table 1. However, in a successful classification with ROC based post-processing, the decrease in correct classifications must be less than the decrease in errors as shown in Table 2. Finally, the empty class is removed and a new confusion matrix is generated. The improvement in overall and per-

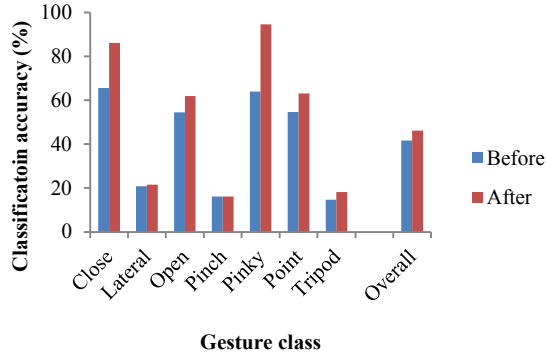


Fig. 5. Classification accuracies before and after applying the ROC analysis (after excluding the empty class) for participant TT1. This example is a case where classification accuracies are not initially very high and the ROC analysis is most beneficial.

TABLE 3. COMPARISON OF OVERALL CLASSIFICATION ACCURACY AND ERROR BEFORE AND AFTER APPLYING ROC ANALYSIS FOR PARTICIPANT TP1.

	Without ROC analysis	With ROC analysis
Accuracy(%)	90.18	84.29% (-5.88)
Error(%)	9.82	4.93% (-4.89)

gesture classification accuracies reported in Table 2 can be visualized as in Fig. 5.

An observation made from the data analysis is that the proposed post-processing technique is more beneficial for improving classification scores when the original score is not high. When initial classification score is high (generally >90%), the decrease in correct classification may occasionally be slightly higher than the decrease in error as reported in Table 3, for a traumatic partial hand participant (TP1). Such a decrease consists of classifications with low confidence, which are therefore discarded and assigned to the ‘empty’ class. However, in this case the ROC analysis does not negatively affect the final per-gesture and overall accuracy, once the ‘empty’ class is removed from the gesture set, as it can be appreciated from the graph in Fig. 6.

IV. CONCLUSION

A post-classification technique to reduce LDA misclassifications has been provided in this paper. The results demonstrate the improvement in classification scores of gesture recognition for thirteen amputees. This improvement is more prominent in the case of classifications which initially have low scores and, usually, there is little improvement in classifications where the score is already high. In the latter case, however, classification scores are not negatively affected by the ROC analysis. Hence it is recommended that the ROC based technique be incorporated to enhance the performance of an LDA classifier, as it can only improve classification results, by discarding classification output with poor confidence. In future work, the computation of the threshold in the case of multiple recordings will be investigated.

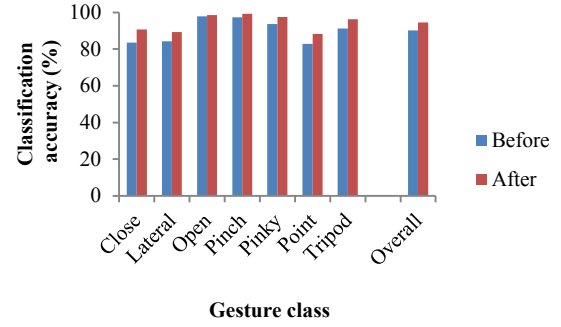


Fig. 6. Classification accuracies before and after applying the ROC analysis (after excluding the empty class) for participant TP1. Since initial classification accuracy is high, the ROC analysis does not markedly improve per-gesture results, as in the case of TT1.

ACKNOWLEDGEMENT

This work has been funded by Innovate UK.

REFERENCES

- [1] M. A. Oskoei, and H. Hu, “Myoelectric control systems—A survey,” *Biomedical Signal Processing and Control*, vol. 2, pp. 275-294, October 2007
- [2] E. A. Bidiss and T. T. Chau, “Upper limb prosthesis use and abandonment: A survey of the last 25 years,” *Prosthetics and Orthotics International*, vol. 31, pp. 236-257, September 2007
- [3] K. Englehart, and B. Hudgins, “A robust, real-time control scheme for multifunction myoelectric control,” *IEEE Transactions on Biomedical Engineering*, vol. 50, pp. 848-854, July 2003
- [4] X. Chen, D. Zhang, and X. Zhu, “Application of a self-enhancing classification method to electromyography pattern recognition for multifunctional prosthesis control,” *Journal of Neuroengineering and Rehabilitation*, vol. 10, pp. 1-13, May 2013
- [5] H. Zhang, Y. Zhao, F. Yao, L. Xu, P. Shang, and G. Li, “An adaptation strategy of using LDA classifier for EMG pattern recognition,” *Engineering in Medicine and Biology Society (EMBC), 35th Annual International Conference of the IEEE*, July 3-7, 2013, Osaka, Japan.
- [6] L. J. Hargrove, E. J. Scheme, K. B. Englehart, and B. S. Hudgins, “Multiple Binary Classifications via Linear Discriminant Analysis for Improved Controllability of a Powered Prosthesis,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, pp. 49-57, January 2010.
- [7] K. Scheme, B. Englehart, and B. S. Hudgins, “Selective classification for improved robustness of myoelectric control under nonideal conditions,” *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 1698-1705, February 2011.
- [8] Z. Li, B. Wang, C. Yang, Q. Xie, and C-Y. Su, “Boosting-Based EMG Patterns Classification Scheme for Robustness Enhancement,” *IEEE Journal of Biomedical and Health Informatics*, vol. 17, pp. 545-552, April 2013.
- [9] K. Scheme, B. S. Hudgins, and B. Englehart, “Confidence-Based Rejection for Improved Pattern Recognition Myoelectric Control,” *IEEE Transactions on Biomedical Engineering*, vol. 60, pp. 1563-1570, January 2013.
- [10] S. Amsuss, P. M. Goebel, N. Jiang, B. Graimann, L. Paredes, and D. Farina, “Self-Correcting Pattern Recognition System of Surface EMG Signals for Upper Limb Prosthesis Control” *IEEE Transactions on Biomedical Engineering*, vol. 61, pp. 1167-1176, December 2013
- [11] B. Hudgins, P. Parker, and R. N. Scott, “A new strategy for multifunction myoelectric control,” *IEEE Transactions on Biomedical Engineering*, vol. 40, pp. 82-94, January 1993.
- [12] L. H. Smith, L.J. Hargrove, B. A. Lock, and T. A. Kuiken, “Determining the optimal window length for pattern recognition-based myoelectric control: balancing the competing effects of classification error and controller delay,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 19, pp. 186-192, April 2011.