Boosting Training for Myoelectric Pattern Recognition using Mixed-LDA*

Jianwei Liu, *Student Member, IEEE,* Xinjun Sheng, *Member, IEEE,* Dingguo Zhang, *Senior Member, IEEE,* and Xiangyang Zhu, *Member, IEEE*

***Abstract*— Pattern recognition based myoelectric prostheses (MP) need a training procedure for calibrating the classifier. Due to the non-stationarity inhered in surface electromyography (sEMG) signals, the system should be retrained day by day in long-term use of MP. To boost the training procedure in later periods, we propose a method, namely Mixed-LDA, which computes the parameters of LDA through combining the model estimated on the incoming training samples of the current day with the prior models available from earlier days. An experiment ranged for 10 days on 5 subjects was carried out to simulate the long-term use of MP. Results show that the Mixed-LDA is significantly better than the baseline method (LDA) when few samples are used as training set in the new (current) day. For instance, in the task including 13 hand and wrist motions, the average classification rate of the Mixed-LDA is 88.74% when the number of training samples is 104 (LDA: 79.32%). This implies that the approach has the potential to improve the usability of MP based on pattern recognition by reducing the training time.**

I. INTRODUCTION

Surface electromyography (sEMG) is the electric potential measured on the surface skin of a muscle. It contains abundant control information from the central nervous system (CNS) and can be used to control electrical powered prostheses [1].

Recently, a large number of studies describing sEMG pattern recognition have been carried out to accomplish the multi-functional prosthetic control [2]–[4]. By extracting sEMG features with high discriminant power and designing appropriate classification techniques, the framework can achieve high classification accuracies (above 90%) on a large number of motion classes (above 10) [3]–[5].

However, it usually requires a long time to train (calibrate) the classifier for obtaining high recognition rate [2], [3], [5]. Furthermore, the durations of the experiments are relatively short (within a few hours), comparing with the practical usage of myoelectric prostheses (across days, months or years). It is known that long-term sEMG signals would change their characteristics, caused by electrode conductivity changes, electrophysiological changes, electrode shift, etc [6]. Such inherent non-stationarity would lead to deterioration on sEMG pattern recognition, as reported in [7] and [8]. To alleviate this effect, the classifier should be re-calibrated in the new day when the prostheses are re-donned [6], which is time-consuming and boresome to the users.

Despite intrinsic non-stationarity in long-term sEMG signals, we hypothesize that there are still some common characteristics in them for the same motion; once the myoelectric prostheses (MP) is custom-built for one individual amputee, the number of electrodes and the muscles from where the sEMG signals would be acquired might not change. Therefore, in the long-term use of MP, the classifiers trained in the earlier days may be partially consistent with sEMG features in the current (later) day. These earlier classifiers can be integrated into training process of the current day. By this method, we anticipate the classifier can be trained faster than learning from scratch.

To realize this idea, we propose a algorithm based on linear discriminant analysis (LDA) [9], which is called Mixed-LDA. This algorithm computes the parameters of LDA through combination of the model estimated on the incoming training samples of the current day with the prior models available from earlier days. An experiment ranged for 10 days, which simulated the scenario of long-term use of MP, was carried out to validate the proposed method. Results show that the Mixed-LDA is significantly better than the baseline method (LDA) when few samples are used as training set. The effort in this study has the potential for improving the usability of MP based on pattern recognition, given the promising results.

II. METHODS

A. Background

LDA might be the most popular classifier used in pattern recognition based MP because of its effectiveness and simplicity [4], [10], [11]. In [8] the authors also found that it was more robust than other classification techniques. Therefore, here we use the LDA to design the classifier.

Given an observation vector x, the Bayesian decision rule shows that $p(\omega_i|x)$, the posterior probability that x belongs to class ω_i, is defined as [9]:

$$p(\omega_i|x) = \frac{p(x|\omega_i)p(\omega_i)}{p(x)} \tag{1}$$

where $p(x|\omega_i)$ is the conditional probability density of the observation x from class ω_i, $p(\omega_i)$ is the prior probability of class ω_i and $p(x)$ is the unconditional probability density of

*This work is supported by the National Basic Research Program (973 Program) of China (Grant No.2011CB013305), the Science and Technology Commission of Shanghai Municipality (Grant No. 13430721600), and the National Natural Science Foundation of China (Grant No. 51375296).

Jianwei Liu (e-mail: sjtuljw520@sjtu.edu.cn; liujianwei-618@163.com), Xinjun Sheng (e-mail: xjsheng@sjtu.edu.cn), Dingguo Zhang (e-mail: dgzhang@sjtu.edu.cn), and Xiangyang Zhu (e-mail: mexyzhu@sjtu.edu.cn) are with State Key Laboratory of Mechanical System and Vibration, School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

the observation x. The Bayesian decision theory determines the class label for a measure x by choosing the class ω_i which produces the maximum posterior probability $p(\omega_i|x)$, from $i = 1\colon C$. Because the unconditional probability density is common for all classes, it can be omitted. The discriminant function can be written as:

$$g_i(x) = p(x|\omega_i)p(\omega_i). \tag{2}$$

The general assumption is that the conditional probability density for each class is Gaussian and can be represented as the multivariate normal distribution:

$$p(x|\omega_i) = \frac{1}{(2\pi)^{d/2}|\Sigma_i|^{1/2}} \exp\{-\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)\} \tag{3}$$

where d is vector dimension of x, and μ_i and Σ_i are the mean vector and covariance matrix for class i respectively.

Assuming that the class data are homoscedastic, the term Σ_i in equation (3) can be replaced by the pooled sample covariance matrix:

$$\Sigma = \sum_{i=1}^{C} \frac{n_i - 1}{N - C}\Sigma_i. \tag{4}$$

And then, taking the natural log of the discriminant function in equation (2) and removing constants without loss of information, after some manipulation, we obtain the discriminant function as:

$$g_i(x) = \mu_i^T\Sigma^{-1}x - \frac{1}{2}\mu_i^T\Sigma^{-1}\mu_i + \ln p(\omega_i). \tag{5}$$

Obviously, it is a linear function. That is why this kind of classification technique is called as LDA. In sEMG pattern recognition, generally the number of training samples for all class is the same and the prior probability for each class is assumed to be equal. Therefore, the term $\ln p(\omega_i)$ can be omitted and the pooled sample covariance matrix Σ can be calculated as the equal weighted average of the covariance matrices across all classes.

B. Mixed-LDA

From the description in Section II-A, we know that the model of LDA for sEMG pattern recognition is determined by the parameters, mean vector and covariance matrix of each class. Suppose there have been S pre-trained models stored in memory, which are trained on data acquired from S prior days. These models can be represented as $(\widehat{\mu}_i^j, \widehat{\Sigma}_i^j)$, $i = 1\colon C$; $j = 1\colon S$. In the new (current) day, we can integrate these prior models into the training process due to their consistence with the current data to some extent. The framework of such integration method is formulated as:

$$\mu_i = (1-r)\overline{\mu}_i + r\sum_{j=1}^{S}\omega_i^j\widehat{\mu}_i^j \tag{6}$$

$$\Sigma_i = (1-r)\overline{\Sigma}_i + r\sum_{j=1}^{S}\omega_i^j\widehat{\Sigma}_i^j \tag{7}$$

where $\overline{\mu}_i$ and $\overline{\Sigma}_i$ are the mean vector and covariance matrix of class i estimated on the incoming training samples of the current day; ω_i^j is the weight determining how much the model of jth day could be reused for class i. From equation (6) and equation (7), we see that the anticipated model is the combination of the current model and the weighted sum of prior models through a trade-off parameter, r. Therefore, we denote the method proposed as Mixed-LDA.

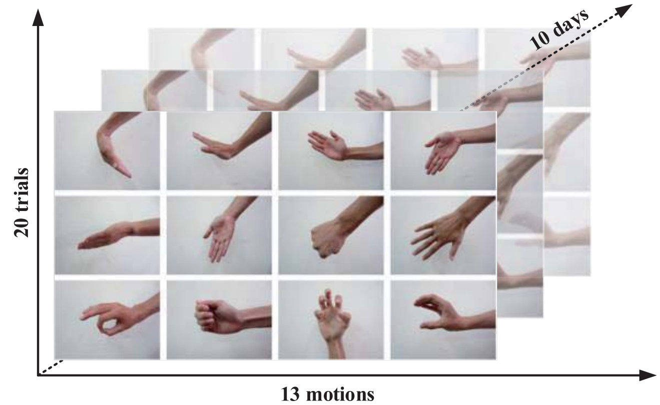

Fig. 1. Experiment procedure for each subject was presented. Each of thirteen motion classes was performed once in each trial; twenty trials were accomplished in each day. Experiment lasted for ten days. Twelve active motions were visible here, which were wrist flexion, wrist extension, radial deviation, ulnar deviation, pronation, supination, fist, open hand, fine pinch, key grip, ball grasp and cylinder grasp.

Here, the weight, ω_i^j, is computed as the inverse proportion to the Mahalanobis distance from the centroid of current model to that of jth prior model for class i. That is:

$$\omega_i^j = \frac{\frac{1}{D(\overline{\mu}_i, \widehat{\mu}_i^j, \widehat{\Sigma}_i^j)}}{\sum_{j=1}^{S}\frac{1}{D(\overline{\mu}_i, \widehat{\mu}_i^j, \widehat{\Sigma}_i^j)}} \tag{8}$$

where $D(\overline{\mu}_i, \widehat{\mu}_i^j, \widehat{\Sigma}_i^j)$ is defined as:

$$D(\overline{\mu}_i, \widehat{\mu}_i^j, \widehat{\Sigma}_i^j) = (\overline{\mu}_i - \widehat{\mu}_i^j)^T(\widehat{\Sigma}_i^j)^{-1}(\overline{\mu}_i - \widehat{\mu}_i^j). \tag{9}$$

We use this metric to quantify the consistence between the prior models and the current data. Larger $D(\overline{\mu}_i, \widehat{\mu}_i^j, \widehat{\Sigma}_i^j)$ means less consistence and results in less weight, ω_i^j.

For the trade-off parameter r, we will compare the classification performances for different values and choose the preferable one which obtains the best results.

C. Experimental Protocol

Five healthy, intact-limbed subjects, who have signed the informed consent, participated in this experiment. The experiment procedures conformed to the Declaration of Helsinki. Before the experiment, the forearm skin of the subject was rubbed with alcohol to provide good condition for sEMG signals acquisition. Four wireless sEMG sensors with bipolar configuration (Delsys INC, USA) were placed on the extensor carpi ulnaris (ECU), flexor carpi radialis (FCR), extensor carpi radialis longus (ECRL) and flexor carpi ulnaris (FCU) respectively. These four muscles were chosen because they were highly related to human wrist and hand motions. After the initial calibration, the sensors positions were marked to help reestablish the experimental setup on the following days.

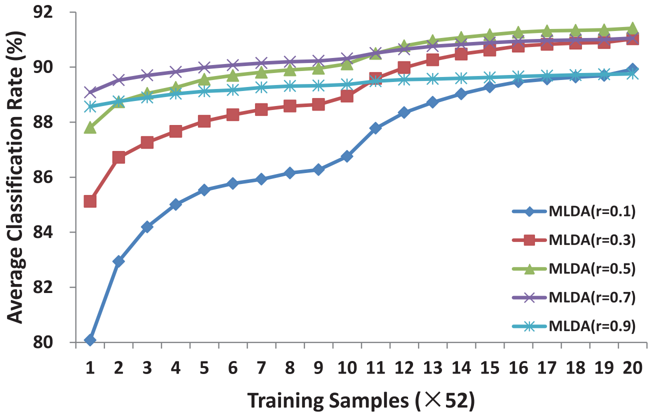

Fig. 2. Comparison of average classification accuracy with few training samples when different values of trade-off parameter are used. Results are obtained by averaging across all subjects and days. MLDA represents Mixed-LDA.

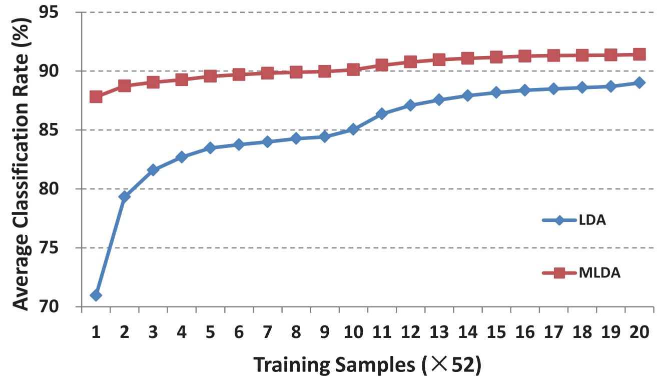

Fig. 3. Comparison of average classification accuracy with few training samples between LDA and Mixed-LDA. Results are obtained by averaging across all subjects and days.

The sEMG signals of thirteen motion classes were collected in the experiment. These motion classes are (m1) wrist flexion, (m2) wrist extension, (m3) radial deviation, (m4) ulnar deviation, (m5) pronation, (m6) supination, (m7) fist, (m8) open hand, (m9) fine pinch, (m10) key grip, (m11) ball grasp, (m12) cylinder grasp and (m13) "no motion". The subjects stood before the computer and naturally extended their arms toward the ground with palms facing inward in the preparatory stage. And then, they were instructed to perform motions with a consistent level of effort. Each motion was sustained for 5 seconds and subjects had a 5 seconds rest between subsequent motions to avoid fatigue. Every motion needed to be performed once in one trial. For each subject, there were totally 20 trials per day and the time interval between experiments of two subsequent days was kept about 24 hours. The entire experiment lasted as long as ten days to simulate the scenario of long-term use of MP. Fig. 1 depicted the whole experiment procedure.

We used a commercial wireless biological signal acquisition system, TrignoTM Wireless system (Delsys INC, USA), to record the sEMG signals. The sEMG signals were band-pass filtered (20–450Hz) by hardware and sampled at 2 KHz. All data were stored and analyzed offline using MATLAB (Mathworks, Inc.) in a 2.5GHz Intel Core Quad CPU computer.

D. Data Preprocessing and Preparation

The central 4 seconds part of each 5 seconds contraction data are used to analysis in order to remove the transient state of the contraction. Afterwards, the usable data are segmented into a series of 200 ms windows with 50% overlap. sEMG features are extracted in each window. The six order AR coefficients are used [2], [12]; hence, one sample in the classification stage is a feature vector with dimension 24 (6 coefficients per channel × 4 channels). After this preprocessing, in each trial 39 samples per motion class are obtained; that is, 780 samples per motion class for one subject are obtained in one day.

As mentioned in Section II-C, we have 10 days experiment data recordings for each subject. Out of these 10 days data, 9 days data are used to trained 9 models off-line, which are regraded as the S pre-trained models stored in memory. The remaining day is regraded as the new (current) day. We repeat this process 10 times to obtain 10 results similar to 10×10 cross-validation for each subject. The average classification accuracy across all motion classes is employed as the measure of performance.

III. Results

A. Classification Accuracy with Few Training samples

In each current (new) day, 20 successive training steps are considered. At the kth step, the number of available training samples per class is $4k$. The test is implemented over all the remaining samples ($780-4k$ samples per class).

1) Choosing Trade-off Parameter r: The trade-off parameter is set between 0.1 to 0.9, with tolerance of 0.2. The comparison between the classification performances of different r is illustrated in Fig. 2. At each training step, the classification rate of one curve is averaged across 50 results (5 subjects, 10 results for each subject). MLDA represents Mixed-LDA proposed. From Fig. 2, when the training samples are scarce, greater value of r (the weighted sum of prior models contributes more for the anticipated model) generally attains better results except for $r = 0.9$. On the other hand, as the increase of training samples, excessively great r can compromise the performance. We can see that the MLDA with $r = 0.5$ achieves a good balance despite the number of training samples. Basically, its performance is good at all training steps; therefore, the trade-off parameter r here should be chosen as 0.5.

2) Boosting Training with Mixed-LDA: Fig. 3 reports the comparison between classification performances of LDA and Mixed-LDA. LDA means the model whose parameters are estimated merely on the incoming training data; that is, the trade-off parameter r in equation (6) and (7) is equal to 0. This training strategy can be regarded as a non-adapted method which has been used in many previous studies. Mixed-LDA ($r = 0.5$) is the training strategy proposed in this study, which can be regarded as an adapted method. As showed in Fig. 3, Mixed-LDA outperforms LDA at each

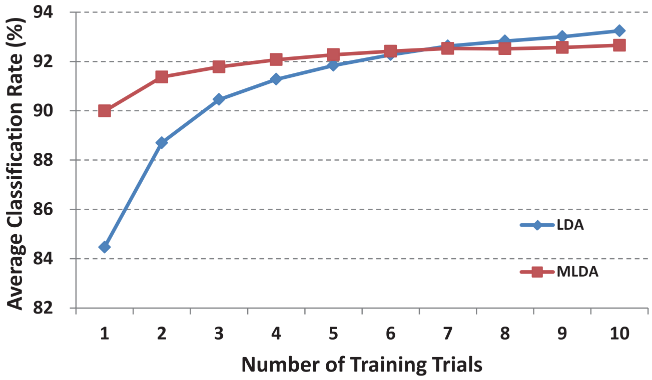

Fig. 4. Comparison of average classification accuracy with more training samples between LDA and Mixed-LDA. Results are obtained by averaging across all subjects and days.

training step. The advantage of Mixed-LDA is extremely evident when few training samples are used; however the advantage reduces as the number of training samples increases. Mixed-LDA surpasses LDA in classification rate of at least 6% when the training samples are not more than 260; at least 5% for 260-520 training samples; at least 2% for 520-1040 training samples.

B. *Classification Accuracy with More Training samples*

In the current (new) day, 10 successive training steps are considered. At the kth step, the available training samples are from k trials data; i.e., the number of training samples per class is $39k$. The remaining samples are used for test. As we can see in Fig. 4, as more and more training samples are used, the superiority of Mixed-LDA ($r = 0.5$) comparing with LDA becomes diminishing. The Mixed-LDA even becomes inferior to LDA when more than 7 trials data are adopted to training.

IV. Discussion and Conclusion

In this study we try to proportionally reuse the models trained on prior days to boost the training process of the new day for the long-term use of MP. Generally, the proposed method, Mixed-LDA, outperforms the baseline method, LDA, when few samples are used for training. For example, Mixed-LDA obtains average 88.74% classification rate using 104 training samples, which is similar to LDA with 988 training samples (88.69%). This indicates that the Mixed-LDA can be trained approximately 9 times faster than LDA. For another aspect, when more and more samples are employed to training, the advantage of Mixed-LDA becomes vanishing. However, the topic in this study is how to achieve an acceptable outcome using training samples as few as possible. The results show that the method proposed gives a possible way to realize this goal.

In [10] the authors proposed a self-enhancing approach based on LDA for improving myoelectric pattern recognition, which performed adaptation during the prediction stage. Meanwhile, the study here defines a way to boost the performance in the training, i.e., before the beginning of prediction. It is possible to obtain an ameliorate version via the combination of these two methods.

One limitation of the proposed method is that we determine the weight, ω_i^j, and the trade-off parameter r heuristically. Especially for the trade-off parameter r, we choose it by trial-and-error and finally fix it at 0.5. This gives an explanation to the phenomenon observed in Fig. 4: the performance of Mixed-LDA becomes inferior to that of LDA when more than 7 trials data are adopted to training. We can image that as the number of training samples increases, the model trained on the incoming training data becomes more reliable for the following testing data than the prior models; therefore, the trade-off parameter r should decrease. In the future, we plan to find an optimization method which can compute the trade-off parameter automatically.

Acknowledgment

The authors would like to thank all the people for participating in the experiments. The authors also thank Jiayuan He for help in experiment data acquisition.

References

[1] P. A. Parker and R. Scott, "Myoelectric control of prostheses." *Critical reviews in biomedical engineering*, vol. 13, no. 4, pp. 283–310, 1985.
[2] Y. Huang, K. Englehart, B. Hudgins, and A. Chan, "A Gaussian mixture model based classification scheme for myoelectric control of powered upper limb prostheses," *Biomedical Engineering, IEEE Transactions on*, vol. 52, no. 11, pp. 1801–1811, 2005.
[3] L. Hargrove, E. Scheme, K. Englehart, and B. Hudgins, "Multiple binary classifications via linear discriminant analysis for improved controllability of a powered prosthesis," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 18, no. 1, pp. 49–57, Feb 2010.
[4] E. Scheme, K. Englehart, and B. Hudgins, "Selective classification for improved robustness of myoelectric control under nonideal conditions," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 6, pp. 1698–1705, June 2011.
[5] J. Liu, J. He, X. Sheng, D. Zhang, and X. Zhu, "A new feature extraction method based on autoregressive power spectrum for improving sEMG classification," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, July 2013, pp. 5746–5749.
[6] J. Sensinger, B. Lock, and T. Kuiken, "Adaptive pattern recognition of myoelectric signals: Exploration of conceptual framework and practical algorithms," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 17, no. 3, pp. 270–278, june 2009.
[7] J. He, D. Zhang, X. Sheng, and X. Zhu, "Effects of long-term myoelectric signals on pattern recognition," in *Intelligent Robotics and Applications*. Springer, 2013, pp. 396–404.
[8] P. Kaufmann, K. Englehart, and M. Platzner, "Fluctuating EMG signals: Investigating long-term effects of pattern matching algorithms," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010, pp. 6357–6360.
[9] K. Fukunaga, *Introduction to statistical pattern recognition*. Academic press, 1990.
[10] 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, no. 1, p. 44, 2013. [Online]. Available: http://www.jneuroengrehab.com/content/10/1/44
[11] N. Wang, Y. Chen, and X. Zhang, "The recognition of multi-finger prehensile postures using LDA," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 706 – 712, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1746809413000918
[12] A. Young, L. Hargrove, and T. Kuiken, "Improving myoelectric pattern recognition robustness to electrode shift by changing interelectrode distance and electrode configuration," *Biomedical Engineering, IEEE Transactions on*, vol. 59, no. 3, pp. 645–652, March 2012.