Improving Long Term Myoelectric Decoding, Using an Adaptive Classifier with Label Correction

Sarthak Jain, Girish Singhal, Ryan J. Smith, Rahul Kaliki and Nitish Thakor

Abstract—This study presents a novel adaptive myoelectric decoding algorithm for control of upper limb prosthesis. Myoelectric decoding algorithms are inherently subject to decay in decoding accuracy over time, which is caused by the changes occurring in the muscle signals. The proposed algorithm relies on an unsupervised and on demand update of the training set, and has been designed to adapt to both the slow and fast changes that occur in myoelectric signals. An update in the training data is used to counter the slow changes, whereas an update with label correction addresses the fast changes in the signals. We collected myoelectric data from an able bodied user for over four and a half hours, while the user performed repetitions of eight wrist movements. The major benefit of the proposed algorithm is the lower rate of decay in accuracy; it has a decay rate of 0.2 per hour as opposed to 3.3 for the non adaptive classifier. The results show that, long term decoding accuracy in EMG signals can be maintained over time, improving the performance and reliability of myoelectric prosthesis.

I. INTRODUCTION

Over the past few years there has been a significant advance in the capabilities offered by prosthetic devices. Today, prosthetic devices more closely resemble human hands, and have dramatically improved from the previously available hooks. With the advances made in the capability of the mechanical parts and the various additional degrees of freedom offered, there is an increased interest in pattern recognition based devices, which promise to offer users control over various additional degrees of freedom. However, despite the limited functionality offered by proportional control based devices, as compared to pattern recognition based devices [1], proportional control based devices are still preferred by amputees since pattern recognition of EMG signals has not yet achieved clinical impact.

Pattern recognition based systems focus on creating a classification system, which is used to separate the incoming muscle signals into various discrete positions. This allows users the freedom to perform various discrete gestures with their prosthetic limb. Such techniques have been shown to demonstrate decoding accuracies over 90% for five or more distinct gestures [2]–[6].

This research was supported in part by National Institutes of Health R44 NS065495.

Sarthak Jain (email: sarthak@iitgn.ac.in) and Girish Singhal are with the Department of Electrical Engineering, Indian Institute of Technology Gandhinagar, Gujarat India 382424
Ryan J. Smith and Nitish V. Thakor are with the Department of Biomedical Engineering, The Johns Hopkins University, Baltimore, Maryland, USA.
Rahul Kaliki is with the Infinite Biomedical Technologies LLC, Baltimore, Maryland, USA.

However, these experiments were performed for a short duration, under strict laboratory settings and fail to to consider the long term effects. Most studies assume that the signal process remains stationary over time, therefore, they neglect modeling changes in the signals over time. A major reason for the lack of stable decoding performance, for such applications, is the change in the underlying probabilistic distributions generating the data. A change in the statistics of the signal is referred to as concept drift [7]–[10].

Two forms of concept drift have been described in literature, (1) gradual concept drift and (2) sudden (also referred to as fast, abrupt and instantaneous) concept drift. It is important to model both these kinds of drift with respect to the classifier, since both are detrimental towards the performance of the system [10], [11].

Concept drift in myoelectric coding may be caused due to variations in electrode conductivity (perspiration, humidity), electro-physiological factors (muscle fatigue, atrophy, or hypertrophy), spatial orientation (electrode movement on the skin or soft tissue fluid fluctuations), user intent (cognitive intent variations or contraction intensity changes), and potentially other factors [12].

To achieve a significant improvement in long term decoding accuracies, it is important to counteract the changes in the myoelectric signals. One of the first barriers in tackling changes in the underlying signal process is to detect when changes occur, known as concept drift detection [9], [13]. Once concept drift has been detected, the changes need to be categorized into slow or fast concept drift. If slow concept drift has occurred, an update in the classifier is performed based on the newly encountered data, retraining the classifier, making it capable of classifying recently encountered data with high accuracy. In the case where fast concept drift has occurred, the classifier does not have access to the correct labels for the recently processed data, as true labels are unavailable in an unsupervised setting. Fast concept drift is more likely to produce mislabeling in the new training set due to the sudden movement of data in the feature space. This leads to a significant problem, as updating the classifier based on mislabeled data leads to a badly trained classifier, which is sure to propagate error over time. However, if corrected labels are assigned to the mislabeled instances, it is possible to deal with the problem by updating the classifier based on recent data in the same way in which slow concept drift is dealt with. Various studies, in different fields, dealing with such phenomena have proposed schema for relabeling which provides the classifier with an improved label set as opposed to the case where there are misclassifications in the labels of
Fig. 1: The figures show the plot of time versus a particular dimension in which the classifier can separate the data. The figure shows the two types of concept drift, figure 1a shows the change in the distribution from t=0 where the classifier was trained to t=T when the classifier is used to make a prediction. As a result of slow concept drift, the data has moved but can still be classified correctly, as well as be used to train a new classifier as we have the correct labels available. In figure 1b demonstrates fast concept drift where the data moves a considerable distance and part of the data is misclassified and hence cannot be used to train a new classifier. Both the images are created using simulated data, using unimodal gaussian distributions.

II. METHODS

A. Experimental Procedure

The study involved muscle signal data collection from a single able-bodied user. The study was performed with the approval of the Johns Hopkins University Institutional Review Board, with prior consent of all the subjects.

We used 16 electrodes, in 8 bipolar pairs. The electrodes used were 3/8” stainless steel dome electrodes with snap connectors (Liberating Technologies INC., Holliston, MA) each of which were embedded inside a silicon cuff, placed on the dominant arm of the able-bodied subject. The electrode pairs were placed equidistantly over the cross section of the subject’s arm. An additional Cleartrace LT electrode (Conmed Corporation, Utica, NY), used as a ground electrode, was placed on the subject’s olecranon.

B. Recording Setup and Data Collection

Differential signal from the electrodes were sent to modified Otto Bock EMG Amplifiers (MYOBOCK, Otto Bock Healthcare, Minneapolis, MN), whose output was sent to a computer using a 16-bit data acquisition card (PCI-6040E, National Instruments, Austin, TX) using a custom-made electrical isolation board and I/O connector block (SCC-68, National Instruments, Austin, TX), the system sampled the EMG signals at 1 kHz.

The subject performed the various gestures as directed by cues appearing on a computer screen. The labels corresponded to eight distinct gestures: hand open/close (HO/HC), wrist flexion/extension (WF/WE), forearm pronation/supination (FP/FS), precision grasp (PG) and index finger point (IP). All subjects performed gestures on the limb ipsilateral to the electrode location.

The four and a half hour data collection period was broken down into fifteen separate sessions, the user was given fifteen minute breaks between sessions. The user continued to wear the silicon cuff with the electrodes for the entire experimental period. Each of these sessions comprised of the user performing 10 repetitions of each of the eight gestures, which appeared in a random order. The user was directed to hold the gesture for between 2.5 and 3.5 seconds, following a rest period of between 2.5 and 3.5 seconds as well.

C. Classifier: Linear Discriminant Analysis

The classifier used for this study was Linear Discriminant Analysis, which has been proven to be effective for myoelectric decoding through previous studies [18], [19]. Discrimination is done by projecting the incoming data into the dimension in which the inter class distance is maximized while minimizing the intra class variance. The feature extraction technique used was Time Domain Auto Regressive (TDAR) as demonstrated by [2], [20]. A total of 56 features were used, seven features per channel, including four coefficients of forth order auto regressive model and three time domain feature values, namely mean absolute value, variance and slope sign changes. We use a window of length 250ms and slide size of 25ms for extracting features.

While using Linear Discriminant Analysis as a classifier the posterior probability is given as a vector:

\[ \Phi = [\phi_1, \phi_2, ... \phi_p] \]  (1)
Fig. 3: Block Diagram of the Proposed System.

where $\phi_i$ represents the probability of an instance belonging to a particular class $i$, each $\phi_i$ is given by:

$$\phi_i = \frac{1}{1 + \exp^{-\sum_{j=1,j \neq i}^{k} w_{i,j}^T x}}$$  \hspace{1cm} (2)

where $w_{i,j}$ is the projection vector of the LDA to separate class $i$ from $j$, $x$ is the original input vector.

D. Concept Drift Detection

Concept drift is detected when the data moves away from its previously observed statistical state. This can be quantified as data moving closer to the decision boundary or crossing it. The probability function for a Linear Discriminant Analysis based classifier is shown to be a monotonic function in the distance from the decision boundary, given by $w_{i,j}^T x$. Therefore, we can use the probabilities directly as measure of concept drift. Entropy gives us a measure of the uncertainty associated with a random variable, which is a better indicator of how far data is from a decision boundary. Entropy is given by:

$$\eta = - \log_n \Phi \times \Phi^T$$  \hspace{1cm} (3)

Where $n$ is the number of distinct labels that can be assigned to the data point.

We use the entropy to predict whether concept drift has occurred, using two separate threshold based criteria. The detection of concept drift is done at defined intervals of time; by summation of entropy in windows, of length $\tau$, and comparing this value to the predefined thresholds. $\alpha$ is the threshold for slow concept drift, which if crossed, we retrain the classifier based on recently encountered data, using labels as predicted by the classifier output. If the entropy crosses $\beta$ we detect fast concept drift for which we retrain the classifier after having performed relabeling. We perform relabeling on the instances constituting the new training set, this is described in a later section. We set these thresholds based on previously observed statistics of the entropy for the user as given by:

$$\mu = \frac{1}{T} \sum_{t=1}^{T} \eta_t$$  
$$\sigma = \frac{1}{T} \sum_{t=1}^{T} (\eta_t - \mu)^2$$  
$$\alpha = \mu + \sigma$$  
$$\beta = \mu + 4 * \sigma$$  \hspace{1cm} (4)

Where $\eta_k$ represents the entropy for the $k^{th}$ instance, $\mu$ and $\sigma$ are calculated based on the entropy of the training set. Values of $\alpha$ and $\beta$ can be empirically determined offline. The value of $\tau$ chosen for this study was set to 48 gestures, which corresponded to 6 repetitions of each gesture; this corresponds to approximately 4 minutes of data collection. We check for an increase in entropy after every $\lambda$ number of movements, for our experiment we selected $\lambda$ as 24 gestures.

E. Update for Slow Concept Drift

A small increase in entropy implies slow concept drift has occurred. At this time we update the classifier by retraining it on the recently encountered data, this ensures the classifier can classify the data after concept drift has occurred. Constant update allows us to maintain a classifier which is capable of classifying the most recently encountered data with high accuracy. In its current version the algorithm does not have any inclusion rules for the new training set. We currently include the last $\tau$ encountered points where $\tau$ is size of the training set. The previous training set is completely discarded, however the rules for data inclusion and exclusion can be modified if required.

F. Relabeling for Fast Concept Drift

To tackle the problem of incorrectly labeled data, we present a relabeling algorithm which removes the misclassifications from the data, based on the knowledge of the updated statistics, along with information available about when the gestures were performed through the Teager Kaiser Energy Operator (TKEO) as described in [21]. The algorithm has three different parts, which are described as follows:

1) Majority Label Correction: Based on the assumption that a user performs only a particular gesture between gesture onset and gesture offset as determined by the TKEO operator, we correct for the spurious misclassifications that occur in between an isolated gesture. This information regarding the gesture allows us to easily correct for mislabeled instances in the duration of a gesture. We determine the most frequently occurring class for this gesture and assign it to the entire duration of the gesture.

In the case where abrupt concept drift occurs, it is possible that a part of the gesture is correctly labeled and a part, of comparable size, is mislabeled. In such a case we use a separate criteria, this is if the number of occurrences of the modal class are not sufficiently higher than the number of instances of the other class, in such a case, we assign the gesture two labels. We call such a case competitive relabeling, which assumes two labels for the data which are corrected through an additional step.
2) Statistics based Relabeling, using PCA based cross validation: Our given training data is now corrected for jitter and spurious misclassifications within gestures. It is however, still subject to misclassified instances which now occur over entire gestures, along with multiple labels for a single gesture. To recover from such a scenario we consider each individual gesture and attempt to relabel it based on the statistics of the available data belonging to the same class. It has been demonstrated in various studies that given the statistics of correctly labeled data it is possible to create a model based on these statistics and reassign labels to misclassified instances [14], [15], [17]. Considering a two class problem, we first project all of the data belonging to the two classes into m dimensional space from the original 56 dimensions, with the objective of capturing maximum variance. For such a projection we use a Principal Component Analysis as the preferred method for dimensionality reduction. It is preferred over Linear Discriminant Analysis since it is class independent [22], which is essential for class reallocation. A dimensionality reduction technique is required to maintain real-time performance as this step needs to be performed multiple times, once for each of the gestures.

We then select all the gestures belonging to the selected two classes (i, j), we refer to this data as \( Y_i, Y_j \), this data excludes the selected gesture \( k \), \( X^k \). We create a logistic regression model to separate the two selected classes based on data \( Y_i, Y_j \) which allows us to predict which class does \( X^k \) belongs to. If the gesture in question (\( X^k \)) initially belonged to the class i, we perform this analysis for all \( j \in 1, ..., 8 \) \( j \neq i \). Such an analysis produces a score for each class \( j \) based on the number of points output by the regression model belonging to class \( j \). We reassign the label to gesture \( X^k \) with the maximum score, if it is above a certain threshold \( \delta \). This process is carried out for all of the gestures and the entire analysis is repeated till no further relabeling occurs.

\[
d_{i,j,n} = u^T_{i,j} X^k_n
\]

\[
m^k_{i,j,n} = \begin{cases} 1 & \frac{1}{1 + e^{-(v^k_{i,j} T d^k_{i,j,n})}} > 0.5 \\ 0 & \text{otherwise} \end{cases}
\]

\[
s^k_{i,j} = \sum_{n=onset}^{offset} (m^k_{i,j,n}) \]

\[
if (\max_j s^k_{i,j}) > \alpha : \quad C'' (X^k) = \arg \max_j s^k_{i,j}
\]

where \( C'' (X^k) \) are the class labels corresponding to the gesture \( X^k \), \( s^k_{i,j} \) is the score attached with each class \( j \), \( u_{i,j} \) is the projection vector as determined by the principle component analysis, \( v^k_{i,j} \) is the projection vector given by the logistic regression model. n is the instance between onset and offset, \( m^k_{i,j,n} \) is the score output of the logistic regression model while considering gesture \( k \).

III. RESULTS

A. Overall Performance and Decay

The first major improvement is the decay in decoding accuracy of the system, which significantly improves for the adaptive system as compared to the non adaptive system. The decay in accuracy per hour for an adaptive system is 0.2 as opposed to a decay in accuracy of 3.3 for a non adaptive system. These results were determined by fitting a linear model through the available data. The improvement

Fig. 2: The figure shows the use of the Teager Keiser Energy Operator to detect movement onset and offset. The Raw signal after filtering, the signal after applying the TKEO and the signal after further low pass filtering are shown. As can be seen from the figure there is a delay caused due to response time of the subject which needs to be accounted for as well as the a delay caused by the TKEO based detection scheme.
Fig. 4: The figures show the key results of the proposed algorithm, 4a is the plot of the mean accuracy over a session versus the session after supervised training. It shows the improvement due to the adaptive system can be seen from the significant difference between the two methods towards the end. 4b shows the plot of the mean accuracy of the training set (the accuracy of the data used to create the classifier). The improvement in the accuracy of the training set over the sessions shows the reason why the decoding accuracy is higher for the proposed algorithm.

in performance due to the proposed method can be seen from the plot of the decoding accuracy vs the session after training 4a. In the figure the gap between the non adaptive and adaptive classifier is seen to widen in later sessions. The mean decoding accuracy over the entire four and a half hours, for the proposed method is $86.8 \pm 2.57\%$, as opposed to $78.4 \pm 2.33\%$ for a non adaptive system. Table I shows the improvement in decoding accuracy for each of the 8 classes, along with the decay in accuracy per hour for each of the classes.

<table>
<thead>
<tr>
<th>Movement</th>
<th>Improvement in Decoding Accuracy</th>
<th>Decay per Hour</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non Adaptive</td>
<td>Adaptive</td>
</tr>
<tr>
<td>HO</td>
<td>2.98</td>
<td>-0.54</td>
</tr>
<tr>
<td>HC</td>
<td>-4.28</td>
<td>-1.2</td>
</tr>
<tr>
<td>FP</td>
<td>8.35</td>
<td>0.45</td>
</tr>
<tr>
<td>FS</td>
<td>7.19</td>
<td>-2.97</td>
</tr>
<tr>
<td>EF</td>
<td>0.48</td>
<td>-0.39</td>
</tr>
<tr>
<td>WF</td>
<td>10.22</td>
<td>-5.52</td>
</tr>
<tr>
<td>WE</td>
<td>41.98</td>
<td>-9.45</td>
</tr>
<tr>
<td>IP</td>
<td>-1.32</td>
<td>-1.2</td>
</tr>
<tr>
<td>Overall</td>
<td>8.2</td>
<td>3.3</td>
</tr>
</tbody>
</table>

IV. DISCUSSION

In this study we have proposed a novel decoding algorithm for use with myoelectric prosthesis, the proposed algorithm consists of an Adaptive Linear Discriminant Analysis Classifier which has update rules based on the entropy of the current samples. Along with the adaptive classifier, we also proposed a relabeling mechanism which corrects for spurious labels that are output by the classifier. The relabeling occurs when the entropy levels cross a particular threshold after which retraining occurs. Such a method is considered unsupervised since we do not have access to the true labels and the only access we have to the true data is through the initial training session, which makes this algorithm suitable for real time implementation. The user would not be required to provide any input to the system after the initial training session which is essential for real time usage. The proposed algorithm significantly improved the stability of the decoding accuracies over time, which has been a major concern. The decay in accuracies over time has prevented prevalent use of pattern recognition based myoelectric prosthesis.

A major reason for improvement is that the adaptive sys-
tem ensures a model, capable of accurately decoding newly encountered data, is maintained throughout. The reason for the improved model is the adaptive classifier; having an adaptive system ensures the model is constantly updated to fit newly encountered data well. Also the relabeling process ensures the classifier is trained on correct data. Having an adaptive model that is trained on correct data and models recent data well is another major reason for the improved decoding accuracy.

However the efficacy of the system can only be validated through more experiments performed with both amputee and able bodied subjects. A major question mark for any myoelectric decoding system is the usability for amputees, since experiments with amputees provide a unique problem of lack of visual feedback. Since amputees do not receive visual feedback due to a missing appendage, the way they perform movements is likely to change as time progress, this can be thought of as another form of concept of drift, which is currently not captured in this experiment.

Another proposed challenge is modeling several kinds of concept drift; this experiment has looked at long term decoding accuracy, but is not an exhaustive set of all possible causes of concept drift, a large amount of data will need to be acquired from both amputee and able bodied subjects, which should help us better understand all the possible causes of concept drift. For example this experiment did not have any electrode drop out or electrode displacement, such problems can cause a major change in the signals which make it very difficult for the model to accurately decode.

V. CONCLUSION

To be viable solutions for amputees, pattern recognition based myoelectric devices require a significant increase in robustness with respect to long term stability in decoding accuracies. In this study we present an algorithm to solve this problem, using an adaptive decoding algorithm which using only training data provided at the beginning, adapts to changes and presents stable decoding in an unsupervised setting. The study demonstrates stable decoding accuracies for more than four hours an able bodied subject. Such an algorithm promises real time implementation, given the low computational complexity of the algorithm, since it only requires update of the classifier when an increase in the systems entropy is detected. However, the efficacy of the algorithm will need to be validated in future work through amputee subjects.

REFERENCES


537