33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Hierarchical Domain Adaptation for SEMG Signal Classification across Multiple Subjects

Rita Chattopadhyay#, Narayanan C Krishnan+ and Sethuraman Panchanathan#
#Center for Cognitive Ubiquitous Computing,
School of Computing, Informatics, and Decision Systems Engineering,
Arizona State University, Tempe, Arizona, 85287, USA.

+ Center for Advanced Studies in Adaptive Systems,
School of Electrical Engineering and Computer Science,
Washington State University, Pullman, Washington, 99163, USA.

Abstract— **Large variations in Surface Electromyogram (SEMG) signal across different subjects make the process of automated signal classification as a generalized tool, challenging. In this paper, we propose a domain adaptation methodology that addresses this challenge. In particular we propose a hierarchical sample selection methodology, that selects samples from multiple training subjects, based on their similarity with the target subject at different levels of granularity. We have validated our framework on SEMG data collected from 8 people during a fatiguing exercise. Comprehensive experiments conducted in the paper demonstrate that the proposed method improves the subject independent classification accuracy by 21% to 23% over the cases without domain adaptation methods and by 14% to 20% over the existing state-of-the-art domain adaptation methods.**

I. INTRODUCTION

Daily life activities such as typing on the keyboard, dusting, brooming, ironing or work in assembly lines, involve repetitive movements of different parts of the body. It has been proved that repetitive tasks make work particularly hazardous as it is the primary cause of muscle fatigue [1]. According to the US Bureau of Labor Statistics, in 2002, there were more than 345,000 on the job back injuries, due to fatigue, which required time off from work. The annual direct cost of occupational injuries due to slip and fall, caused due to muscle fatigue ,is expected to exceed $43.8 billion by the year 2020 in the US (Bureau of Labor Statistics, 2004).

These accidents and the consequential loss in work hours and life, besides the high medical cost, can be avoided if one can intervene at an early stage by intelligent devices having the capability for monitoring and detecting different stages of fatigue. Technologies for detecting muscle fatigue at an early stage can also be used to remove the cause of fatigue by altering the environmental ergonomics where possible [2].

Researchers have observed that certain aspects of Surface Electromyogram (SEMG) signals such as shift in the power spectral density, root mean square (rms), instantaneous frequency, median-frequency change as muscles become fatigued. Changes in the SEMG power spectrum and their shifts vary significantly across subjects. These wide and unpredictable variations make the task of modeling and classification of SEMG signals challenging.

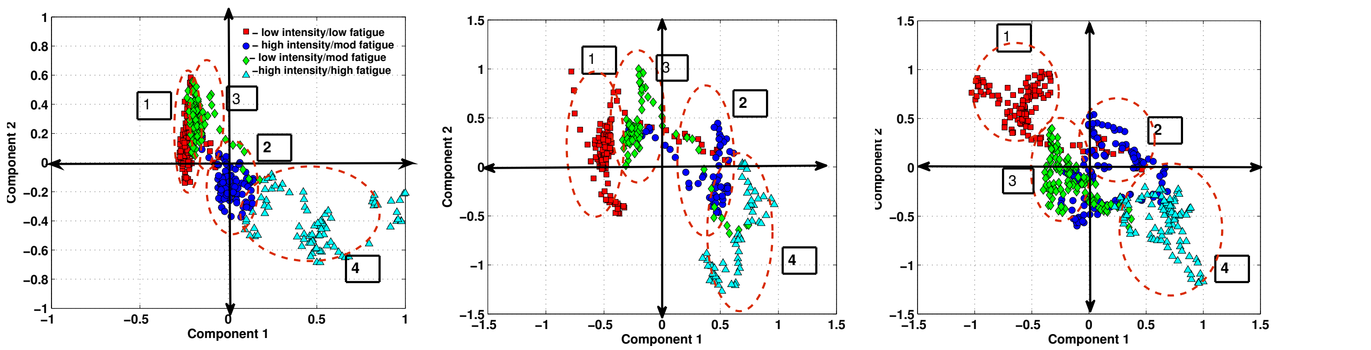
Figure 1 shows the distribution of SEMG data over four stages of a fatiguing activity for three different subjects. The data distribution shown in Figure 1 is a two dimensional projection obtained through factor analysis on the twelve dimensional feature vectors derived from raw SEMG signals. The four stages of fatigue with varying intensity of activity, corresponding to four classes, shown in the figure, are (l) *low intensity of activity and low fatigue*, (2) *high intensity of activity and moderate fatigue*, (3) *low intensity of activity and moderate fatigue* and (4) *high intensity of activity and high fatigue*. We observe that the data distribution during each stage or class varies from subject to subject. This variation leads to both marginal and conditional probability differences across subjects.

In this paper we present a successful case study of application of domain adaptation techniques for detecting different stages of fatigue in the SEMG signal of a test subject using the available knowledge from multiple subject data. for detecting different stages of fatigue on a test or target subject. The proposed methodology can also be used to develop a generalized classification framework for other physiological signals such as ECG, EEG, respiratory rate and pulse rate besides applications related to emotion and speech analysis which have significant subject based variabilities.

II. RELATED WORK

Most of the past work on fatigue analysis has been subject specific. Currently there are no reliable and roboust indicators of fatigue across subjects [3], consequently most of the work has focussed on feature analysis at individual subject level. There has been some effort on developing classification framework [4], [5] for detecting fatigue, but these frameworks were mostly confined to distinguishing non-fatigue and fatigue state SEMG signals. Classifying intermediate stages of fatigue is still an under explored research area.

Domain adaptation methodologies have been commonly used to address distribution differences in multiple research areas such as Text classification, Video Concept Detection,



(a) Subject 1 (b) Subject 2 (c) Subject 4

Fig. 1. Three sample subjects (subjects 1, 2, 4) with four classes (four physiological stages) in our SEMG data set: Differences in marginal and conditional probability distribution across subjects.

Sentiment Analysis, WiFi Localization [6], [7], [8], [9]. In this paper we propose a novel application of domain adaptation methodology i.e. subject based variability in SEMG signal.

Most of the existing domain adaptation methodologies are focused on reducing marginal probability differences between two data distributions [10], [11], [12], [13]. Huang et al [12] re-weight the samples in source domain so as to minimize the marginal probability difference, referred as *Kernel Mean Matching (KMM)*, while Pan et al [13] suggests feature mapping for reducing the marginal probability differences between the source and target distribution referred as *Transfer Component Analysis (TCA)*.

However there has been little effort in addressing conditional probability distribution differences. Gao et al [14] reduces the conditional probability difference by comparing the clusters of the target domain with respect to source domain data (*Locally Weighted Ensemble (LWE)*). Zhong et al [15] proposed a two stage approach, feature mapping followed by instance selection, for addressing marginal and conditional probability differences (*KMapEnsemble (KE)*). We have previously presented an abstract of a multi source domain adaptation methodology [16], based on learning a target classifier using labels generated by an unsupervised scheme.

We compare the proposed methodology with four different domain adaptation techniques: KMM and TCA that address marginal probability differences; LWE, that addresses only conditional probability differences and KE which addresses both marginal and conditional probability differences, besides comparing with other baseline methods.

III. Proposed Framework

A. Problem Formulation

Assume that there are K subjects in the source domain with M classes. The k-th subject in the source domain is characterized by a sample set $D^k = (x_i^k, y_i^k)|_{i=1}^{N_k}$, where x_i^k is the feature vector, y_i^k is the corresponding label, and N_k is the total number of samples for the subject k. The target domain consists of a few labeled data $D_l^T = (x_i^T, y_i^T)|_{i=1}^{N_l}$ and plenty of unlabeled data $D_u^T = x_i^T|_{i=1}^{N_u}$ where N_l and N_u are numbers of labeled and unlabeled target domain samples respectively, $D^T = D_l^T \bigcup D_u^T$, and $N_T = N_l + N_u$. The goal is to develop a target classifier f^T that can predict the labels of the unlabeled data in the target domain, using the multi-source domain data and a few labeled target domain data.

One simple approach is to learn a classifier on data from multiple subjects available in the source domain. This method has a drawback as the classifier is learned to minimize loss on source domain data and not on target domain. Another approach is to measure the distribution difference between each subject data in source domain with the target subject data and combine the hypothesis generated by each subject based on the measure. The challenge in this approach is to select the right measure. In addition, we observe from Figure 1 that in SEMG signal different classes vary differently over subjects. For example, subjects 1 and 2 have similar data distribution for classes 1 and 3, and subjects 2 and 4 have similar data distribution for classes 4. Hence computing a single similarity measure for a source domain subject data with respect to target data does not capture the differences at the class level. The proposed approach presented below addresses all these challenges by following a hierarchical confidence weighted sample selection strategy that considers similarities at all levels.

B. Hierarchical Confidence Weighted Sample Selection

In the proposed approach we measure the similarities between the source domain subject data and the target subject data at three different levels, each with increasing granularity. We call this framework *Hierarchical Confidence Weighted Multi Source Domain Adaptation* (HC-MDA) which is outlined in Algorithm 1.

As a first step to the proposed approach we learn a model M^{T_l} from the labeled target subject data D_l^T. We compute the classification accuracy obtained on each subject data $\{D^k\}_{k=1}^K$, by this model. This classification accuracy given by w^k is used to estimate the similarity of a particular source domain subject k with respect to the target subject. The classification accuracy reflects the differences in both marginal and conditional probability distributions. The classification accuracies are normalised across subjects to obtain a relative similarity measure between a source domain and a target domain subject. While this measure encapsulates the overall similarity of a source domain subject k with respect to target subject, it does not address the distribution difference or similarities at individual class level.

The next level involves computing similarity between the individual classes of source domain subject and the target subject. This similarity measure is computed by determining the average true positive rate, w_c^k, for a class c belonging to a source domain subject k. Normalised w_c^k, classwise across subjects, reflects conditional probability differences between source and target subject classes.

However this measure still overlooks the dissimilarities between the instances of source subject with respect to target subject data e.g. even if a particular class has a true positive rate as high as 80%, there are still 20% of the instances which are not similar to the target domain distribution. In order to avoid selecting these instances for domain adaptation, the proposed framework advocates computation of similarity at a still finer level of granularity. This selection is achieved by concentrating only on the correctly classified instances of a class and selecting instances with higher confidence of prediction, $D_{sel}^{K,C}$. A classifier is learned these and labeled target samples D_l^T and the learned model is used to classify and label the unlabeled target domain data D_u^T.

Algorithm 1 The Hierarchical Confidence Weighted-Multi-Domain Adaptation

Input Source domain subject samples $\{D^k\}_{k=1}^K$ and small amount of target domain subject training examples D_l^T
Output $D_{sel}^{K,C}$
Learn a model M^{T_l} using D_l^T
for $k = 1, \ldots, K$ **do**
    Weight for source D^k : w^k = Classification accuracy for D^k using $M^T l$
    for $c = 1 \ldots, C$ **do**
        Weight for class c of source D^k: w_c^k = True positive rate for class c on the samples from D^k using $M^T l$
    end for
end for
Normalise the w^k and w_c^k for each c, over all K subjects
Compute the number of samples N_c^k to be selected from a source subject k for a class c

$$N_c^k = (w_c^k \times w^k) \times |\{x_i : M^{T_l}(x_i) = y_i and y_i = c\}| \tag{1}$$

$D_{sel}^{k,c}$ = first N_c^k samples $\in \{x_i : M^{T_l}(x_i) = y_i and y_i = c\}$ with the highest confidence of classification using M^{T_l}
Output $D_{sel}^{K,C}$

IV. Experiments

A. Data Collection and Feature Extraction

The SEMG data was collected during a repetitive gripping action performed by the forearm. Figure 2 shows the subject with surface EMG differential electrodes on the extensor carpi radialis muscle to record the SEMG signal. The subject performs a cycle of flexion-extension of forearm as shown in Figure 2 at two different speeds, i.e., low speed (1 cycles/sec) and high speed (2 cycles/sec) repetitively for about 4 minutes. The cycles of low and high speed are alternated after every minute to form four phases (or classes) as described in Section I

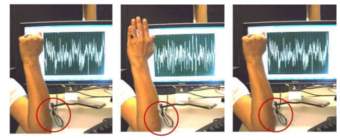


Fig. 2. SEMG data collection during a repetitive gripping activity

The raw SEMG activity was recorded by Grass Model 8-16C at 1000Hz and passed through a band pass filter of 20Hz to 500Hz. The data was collected and saved by the LabView software (from National Instruments) running on a PC. Data was collected from 8 subjects including male and female of the age group of 25 years to 45 years.

A set of twelve amplitude and frequency domain features including mean frequency, median frequency, spectral energy, spectral entropy, root mean square, number of zero crossings, to mention a few are derived from running windows of 1000 time samples with 50% overlap.

B. Experimental Procedure

Effectiveness of the proposed method is evaluated against the baseline methods SVM-C and SVM-T and against state-of-the-art domain adaptation methods namely Kernel Mean Matching (KMM), Transfer Component Analysis (TCA), Kernel Ensemble (KE) and Locally Weighted Ensemble (LWE). The description of these domain adaptation methods is presented in Section II. SVM-C refers to *all but one* method where the training data comprises of data from seven subjects and the trained model is tested on unseen test subject and SVM-T refers to a classifier learnt only on the labeled samples from the test subject. Results are obtained from a leave one subject out cross validation process. We randomly select a set of 4 labeled samples from the target subject to constitute D_l^T. These labeled test data is added to all the methods for fair comparison. The classification accuracies are averaged over ten folds of execution to remove any bias due to selection of any specific labeled data set from test subject.

Model parameters for different techniques were obtained through a cross validation process on a set aside validation data set. SVM-C was trained with Gaussian Kernel with $\sigma = 0.5$ and high C. For KMM, Gaussian Kernel with $\sigma = 10$ gave the best results on validation set. For KMM, weighted SVM, part of LibSVM package, was used to learn a model with weights associated with every data sample. TCA was implemented with linear kernel and feature mapping was obtained with dimension value of 10. As suggested by Zhong et al, ‘bisectKmeans’ was used for clustering, in KE and in LWE ‘Kmeans’ was used for clustering the test data.

V. Results and Discussion

Average cross validated accuracy for each of the methods is summarized in Table I. The first column of the table indicates the test subject and the training data consists of the data from the remaining seven subjects. We observe that the proposed method significantly improves the performance over other methods by an average gain of 15% to 20%.

TABLE I

COMPARATIVE PERFORMANCE OF DIFFERENT METHODS ON SEMG DATA - ACCURACY (%)

Test Sub	SVM-C	SVM-T	TCA	KE	KMM	LWE	HC-MDA
1	72.76	62.12	55.45	65.45	72.42	67.44	82.61
2	53.69	67.50	59.94	60.98	63.63	77.54	80.06
3	55.11	62.58	72.57	63.16	68.69	75.55	81.45
4	59.65	64.42	69.89	59.68	72.38	81.22	87.05
5	60.37	71.87	64.06	61.33	62.5	52.48	87.97
6	59.21	49.09	59.02	54.54	70.62	65.77	78.86
7	57.13	51.09	62.42	60.17	61.13	60.32	80.43
8	64.85	70.79	62.48	83.41	74.79	68.55	85.73
Average	60.34	62.43	63.22	64.41	68.27	69.14	**83.02**

LWE, that addresses conditional probability distribution difference, performs better than other methods that address only marginal probability differences. However LWE, performs poorer than HC-MDA since it computes the weights of each subject depending upon a overall similarity factor, overlooking the similarities at different levels.

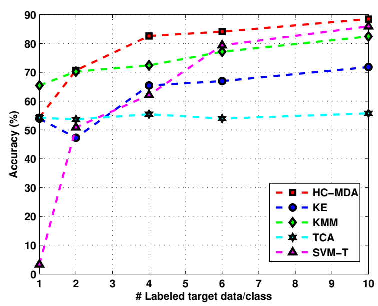


Fig. 3. Accuracy (%) Vs No of labeled target data

Figure 3 presents the variation in classification accuracies for different methods with varying number of labeled data available from the target subject. The performance of HC-MDA is moderate when the nukmber of labeled samples from target subjects is either 1 or 2. This is bacause the small number of samples is insufficient to learn the target model M_l^T. However beyond 2 labeled samples from target subject, HCMDA performs better than other techniques. Performace of TCA is low because of conditional probability differences in the data, that negatively affects the feature mapping process of TCA [17]. Even though KE too involves a feature mapping step, the sample selection strategy helps it to perform better than TCA.

SVM-T performs very poorly upto 4 samples per class. After 6 samples per class, SVM-T performs better than most of the approaches as the number of labeled samples from target subject is sufficient to learn a reliable model.

VI. CONCLUSIONS AND FUTURE WORKS

We consider the characterization of muscle fatigue through noninvasive sensing mechanism such as surface electromyography . The variation in Surface Electromyogram (SEMG) signal parameters from subject to subject creates differences in the data distribution making traditional data mining algorithms ineffective. In this paper, we propose a hierarchical confidence weighted domain adaptation methodology for detecting different stages of fatigue for multiple subjects. We have validated our framework on real-world SEMG data collected from eight different subjects during a fatiguing exercise. Our comprehensive experiments demonstrate the effectiveness of the proposed framework and suggest that it is possible to develop a generalized framework for SEMG data. We plan to extend the proposed framework to applications involving other types of physiological signals for emotion and health monitoring in everyday life, industrial work and geriatric care.

REFERENCES

[1] V. L. Young, M. K. Seaton, C. A. Feely, C. Arfken, D. F. Edwards, C. M. Baum, and S. Logan, "Detecting cumulative trauma disorders in workers performing repetitive tasks," *American Journal of Industrial Medicine*, vol. 27, pp. 419–431, 1995.

[2] http://www.environmental-ergonomics.org/.

[3] P. Contessa, A. Adam, and C. J. D. Luca, "Motor unit control and force fluctuation during fatigue," *Journal of Applied Physiology*, April 2009.

[4] N. F. Gler and S. Koer, "Classification of EMG signals using PCA and FFT," *Journal of Medical Systems*, vol. 29, no. 3, pp. 241–250, June 2005.

[5] A. Subasi and M. Kiymik, "Muscle fatigue detection in emg using time-frequency methods, ica and neural and networks," *J Medical Systems*, April 2009.

[6] S. J. Pan, J. T. Kwok, and Q. Yang, "Transfer learning via dimensionality reduction," in *Proceedings of AAAI 08*.

[7] L. Duan, I. W. Tsang, D. Xu, and T. S. Chua, "Domain adaptation from multiple sources via auxiliary classifiers," in *ICML*, 2009, pp. 289–296.

[8] J. Blitzer, M.Dredze, and F. Pereira, "Biographies, bollywood, boomboxes and blenders: Domain adaptation for sentiment classification," in *ACL*, 2007.

[9] H. D. III, "Frustratingly easy domain adaptation," in *ACL*, 2007.

[10] M. Sugiyama, S. Nakajima, H. Kashima, P. V. Buenau, and M. Kawanabe, "Direct importance estimation with model selection and its application to covariate shift adaptation," in *NIPS*, 2008.

[11] S. Bickel, M. Brückner, and T. Scheffer, "Discriminative learning under covariate shift," in *JMLR*, vol. 10, 2009, pp. 2137–2155.

[12] J. Huang, A. J. Smola, A. Gretton, K. M. Borgwardt, and B. Scholkopf, "Correcting sample selection bias by unlabeled data," in *NIPS*, 2007, p. 601.

[13] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," in *IJCAI*, 2009.

[14] J. Gao, W. Fan, J. Jiang, and J. Han, "Knowledge transfer via multiple model local structure mapping," in *KDD*, 2008, pp. 283–291.

[15] E. Zhong, W. Fan, J. Peng, K. Zhang, J. Ren, D. Turaga, and O. Verscheure, "Cross domain distribution adaptation via kernel mapping," in *KDD*. ACM, 2009.

[16] R. Chattopadhyay, J. Ye, and S. Panchanathan, "Transfer learning framework for early detection of fatigue using non-invasive surface electromyogram signals (semg)," *Extended Abstract: AAAI Spring Symposium Series*, 2011.

[17] J. Jiang, *A literature survey on domain adaptation of statistical classifiers*. Citeseer, 2008.