Multi-Source Domain Adaptation and Its Application to Early Detection of Fatigue

Rita Chattopadhyay [1], Jieping Ye [1,2], Sethuraman Panchanathan [1], Wei Fan [3], Ian Davidson [4]

[1]Computer Science and Engineering, Arizona State University, AZ 85287
[2]Center for Evolutionary Medicine and Informatics, The Biodesign Institute, Arizona State University, AZ 85287
[3]IBM T.J.Watson Research, Hawthorne, NY 10532
[4]Department of Computer Science, University of California, Davis, CA 95616

ABSTRACT

We consider the characterization of muscle fatigue through noninvasive sensing mechanism such as surface electromyography (SEMG). While changes in the properties of SEMG signals with respect to muscle fatigue have been reported in the literature, the large variation in these signals across different individuals makes the task of modeling and classification of SEMG signals challenging. Indeed, the variation in SEMG parameters from subject to subject creates differences in the data distribution. In this paper, we propose a transfer learning framework based on the multi-source domain adaptation methodology for detecting different stages of fatigue using SEMG signals, that addresses the distribution differences. In the proposed framework, the SEMG data of a subject represent a domain; data from multiple subjects in the training set form the multiple source domains and the test subject data form the target domain. SEMG signals are predominantly different in conditional probability distribution across subjects. The key feature of the proposed framework is a novel weighting scheme that addresses the conditional probability distribution differences across multiple domains (subjects). We have validated the proposed framework on Surface Electromyogram signals collected from 8 people during a fatigue-causing repetitive gripping activity. Comprehensive experiments on the SEMG data set demonstrate that the proposed method improves the classification accuracy by 20% to 30% over the cases without any domain adaptation method and by 13% to 30% over the existing state-of-the-art domain adaptation methods.

Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications - Data Mining

General Terms

Algorithm

Keywords

Multi-source domain adaption, transfer learning, surface electromyogram, subject based variability

1. INTRODUCTION

Daily life activities such as typing on the keyboard, dusting, brooming, ironing, as well as the use of hand tools such as scissors and knives, repetitive work in assembly lines, repetitive lifting, involve repetitive movements of the different parts of the body. It has been proved that repetitive task makes work particularly hazardous, as it is the primary cause of muscle fatigue [28, 13, 30]. According to the US Bureau of Labor Statistics, in 2002, there were more than 345,000 on the job back injuries, due to muscle fatigue, which required time off from work. According to the Bureau of Labor Statistics (2004), 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.

These accidents and the consequential loss in work hours and lives, besides the high medical cost, can be avoided if one can intervene such fatigue inducing repetitive activities at an early stage by intelligent devices having the capability for 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 [1].

There are a number of techniques that can be used to objectively determine the level of fatigue in a subject. Electromyography (EMG) is a method for biosignal recording of skeletal muscle activity. Surface Electromyography (SEMG) allows for noninvasive recording of these biosignals. Researchers have observed that certain aspects of SEMG signals change as a muscle becomes fatigued. Localized muscle fatigue has been correlated with a shift in the power spectral density of SEMG signals, root mean square (rms), instantaneous frequency, zero crossing rate, mean-frequency, median-frequency, etc. However, there is a large variation in the values of these measures across different subjects, due to variances in their SEMG power spectrum and their shifts. These generally unpredictable and wide variations make the task of modeling SEMG difficult, and the task of automating the process of signal classification as a generalized tool challenging. The variation in SEMG parameters from subject to subject creates differences in the data distribution. Figure 1 shows the distribution of the data over four stages of a fatigue-causing activity, done with varying speed, for three different subjects (subjects 1, 2, 4). The data distribution shown in Figure 1 is of factor scores obtained as a
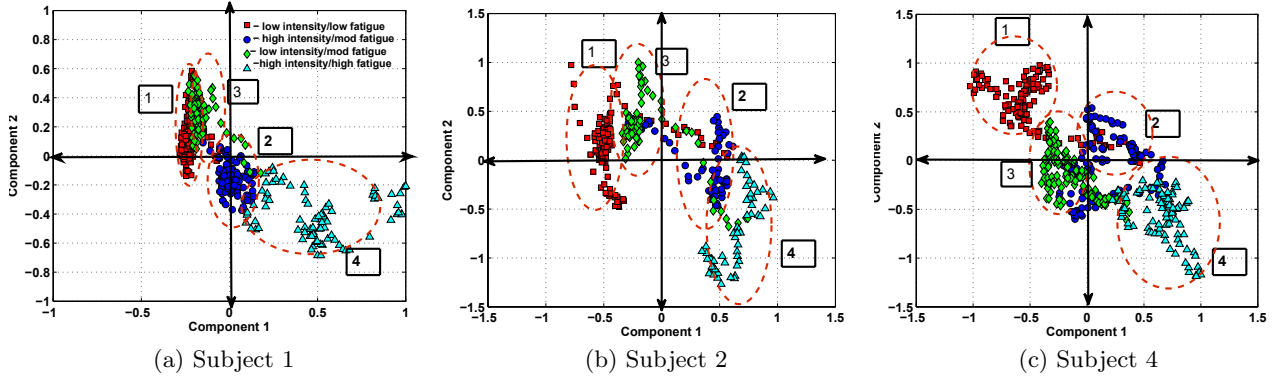
Figure 1: Three sample subjects (subjects 1, 2, 4) with four classes (four physiological stages) in our SEMG data set: SEMG signals are predominantly different in conditional probability distribution across subjects.

result of factor analysis [2] applied on the twelve dimensional feature vectors derived from raw SEMG signals[1]. The four physiological stages corresponding to four classes, shown in the figure, are (1) *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 predominantly conditional probability differences across subjects.

Traditional data mining algorithms assume that training data and test data are drawn from the same distribution, and they may not be effective if the assumption is violated as in the case of SEMG data over multiple subjects. One effective approach is domain adaptation which enables transfer of knowledge between the source and target domains [22]. It has been applied successfully in various applications [6, 11, 10, 20] including text classification (parts of speech tagging, webpage tagging, etc), video concept detection across different TV channels, sentiment analysis (identifying positive and negative reviews across domains), WiFi Localization (locating device location depending upon the signal strengths from various access points.

In this paper we present a successful case study of application of multi-source domain adaptation techniques for detecting different stages of fatigue based on the Surface Electromyogram signals across multiple subjects. The proposed framework addresses the subject based variability, predominantly the distribution differences in conditional probabilities in Surface Electromyogram signals. Specifically, a classifier is learnt to distinguish the four classes as shown in Figure 1 on the basis of some labeled and unlabeled data from the target domain (or subject). The unlabeled data are labeled using a weighting scheme that measures the similarities in conditional probabilities between the source and target domain data; the key of the proposed weighting scheme is a joint optimization framework based on smoothness assumption on the probability distribution of the target domain data. To the best of our knowledge, this is the first systematic analysis of subject based variability in SEMG signals. We have applied the proposed algorithm to Surface Electromyogram signals collected from 8 people during

---

[1] More details on the twelve features derived, the factor analysis results and a real time deployment of fatigue grading framework can be found in our earlier papers [8, 9].

a fatigue-causing repetitive gripping activity. Our extensive experiments on the SEMG data set demonstrate that the proposed method improves the subject independent classification accuracy by 20% to 30% over the cases without any domain adaptation method and by 13% to 30% over the existing state-of-the-art domain adaptation methods.

2. PROPOSED FRAMEWORK

The proposed domain adaptation framework focuses on learning from multiple auxiliary sources related to the target data, e.g., multiple subject data having different distributions, collected under similar physiological conditions. Specifically, we consider the problem of detecting different stages of fatigue in a subject for whom we have very few labeled samples available in the training data. The training data also includes data from many other subjects collected under similar physiological conditions. The test subject data forms target domain data and the multiple subject data in the training domain form multiple auxiliary sources.

2.1 Problem Setting and Motivation

Assume that there are k subjects in the source domain. The s-th subject in the source domain is characterized by a sample set $D^s = (x_i^s, y_i^s)|_{i=1}^{n_s}$, where x_i^s is the feature vector, y_i^s is the corresponding label, and n_s is the total number of samples for the subject s. 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 $D_u^T = x_i^T|_{i=n_l+1}^{n_l+n_u}$ 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 for predicting the labels of the target domain data is to combine the training samples from all subjects and build a single classifier based on the pooled training samples. However, this simple approach will not work well in our application as there are significant conditional probability differences across subjects in the SEMG data. A better alternative is to learn individual models for each subject in the source domain, and then combine the hypothesis generated by each of these source models, on the basis of some similarity measures between the source and target domains. The similarity measures are commonly computed by considering each source domain data separately. This pro-

cedure has two potential limitations. First, it minimizes the loss with respect to the probability distribution $P_s(x, y|D^s)$ on the source domain which will not generally coincide with the minimal loss on the distribution $P_T(x, y|D^T)$ on the target domain. Second, it assumes all sources are independent when computing the similarity measures, thus it does not fully exploit the interaction among multiple sources.

Thus one of the important issues is to choose the right similarity measure between the source and target domains depending upon the nature of differences in the distribution. There are predominantly conditional probability differences in our multi-subject SEMG data as shown in Figure 1. In addition, we observe from the figure that different classes vary differently over subjects. For example, subjects 1 and 2 have similar data distributions of classes 1 and 3, and subjects 1 and 4 have similar data distributions of class 4. Hence computing a single similarity measure for an auxiliary source with respect to target data does not capture the differences at the class level. The proposed approach presented below addresses all these challenges.

2.2 Proposed Approach

We learn a classifier f^T for the target domain data, using a few labeled samples and a large number of unlabeled samples from the target domain. The key of our proposed approach is a novel weighting scheme that integrates multiple source domain data using a set of weights, one for each source domain. We use these weights to compute the labels of the unlabeled target domain data, called "pseudo labels". The target domain prediction model is then learnt from both labeled and pseudo labeled target domain samples in a regularized framework. Specifically, the proposed multi-source domain adaptation framework is given as follows:

$$\min_{f^T \in H_K} \gamma_A \|f^T\|_K^2 + \frac{1}{n_l} \sum_{i=1}^{n_l} V(x_i, y_i, f^T) + \Omega_r(f_u^T) + \Omega_m(f^T)$$

$$(1)$$

The first term controls the complexity of the classifier f^T in the Reproducing Kernel Hilbert Space (RKHS) H_K, γ_A controls the penalty factor, the second term is the empirical error of the target classifier f^T on the few labeled target domain data D_l^T, and n_l is the number of labeled target domain data. The empirical error on the unlabeled target data, labeled using a conditional probability based weighting scheme, forms the third term. This regularizer enforces the target classifier f^T to have similar decision values to the auxiliary source which has similar conditional probability distribution, explained in detail in Subsection 2.3. The fourth term is a manifold based regularizer based on the smoothness assumption [4] on target domain data: if two points x_i and x_j are close to each other in the intrinsic geometry of marginal distribution then they are most likely to have similar conditional probabilities, i.e., $f^T(x_i)$ should be similar to $f^T(x_j)$. The *manifold based regularizer* is defined as in [4]:

$$\Omega_m(f^T) = \frac{\gamma_I}{n_T^2} f^{T'} L f^T. \qquad (2)$$

where L is the graph Laplacian matrix constructed on D^T, $f^T = [f^T(x_1), \cdots, f^T(x_{n_T})]$, γ_I controls the complexity of the function f^T in the intrinsic geometry of the marginal probability of x and the normalizing coefficient $\frac{1}{n_T^2}$ is the natural scale factor for the empirical estimate of the Laplace

operator, and the symbol $'$ is used to represent the matrix or vector *transpose* operation.

2.3 Multi-Source Weighting

Let $f_u^T = [f_{n_{l+1}}^T \cdots f_{n_T}^T]'$ be the decision values of the target classifier f^T for the unlabeled target domain data and let $f_u^s = [f_{n_{l+1}}^s \cdots f_{n_T}^s]'$ be the decision values of the s-th auxiliary classifier for the same unlabeled target domain data. Let γ_s be the measure of relevance or similarity between the distributions of the s-th source and the target data, and let $f_j^T = f^T(x_j)$ be the decision value of target classifier on the target domain data x_j and $f_j^s = f^s(x_j)$ be the decision value of the s-th auxiliary source classifier on x_j. We use a weighted combination of the k source domain classifiers f^s to estimate the target classifier. Specifically, the estimated label ($\hat{y}_j$) of the unlabeled target data x_j based on the k source domain classifiers f^s is given by

$$\hat{y}_j = \sum_{s=1}^{k} \gamma_s f_j^s, \qquad (3)$$

where $\gamma_s > 0$ is the weight for the s-th source. We assume that the weights are normalized, that is, $\sum_s \gamma_s = 1$. The auxiliary classifier f^s for the s-th source is pre-computed based on its respective data. The auxiliary classifiers f^s and the target classifier f^T can be trained using different kernels or even different learning methods. The resulting regularizer, $\Omega_r(f_u^T)$, named as *relevance based regularizer* measures the difference between the target classifier decision value and the estimation based on multiple source data, and is defined as follows:

$$\Omega_r(f_u^T) = \frac{\theta}{2} \sum_{j=n_l+1}^{n_T} \|f_j^T - \sum_{s=1}^{k} \gamma_s f_j^s\|^2, \qquad (4)$$

where $\theta > 0$ is a constant.

The weight γ_s which provides a measure of relevance between the s-th auxiliary source domain and the target domain is computed on the basis of a *Conditional Probability based Weighting Scheme*, which evaluates the similarities in distributions between the source and target domains predominantly based on conditional probability differences.

Next, we show how to estimate the weights γ_s's. The proposed weighting scheme evaluates the similarities between auxiliary source data and the target domain data considering the similarities in their conditional probabilities.

Let $F_i^S = [f_i^1 \cdots f_i^k]$ be the $1 \times k$ vector of predicted labels of k auxiliary source models for the i-th sample of target domain data. Let $\gamma = [\gamma_1 \cdots \gamma_k]^T$ be the $k \times 1$ weight vector, where γ_s is the weight corresponding to the s-th auxiliary source. Following (3), the predicted label for the i-th sample of target domain data is

$$\hat{y}_i = \sum_{s=1}^{k} \gamma_s f_i^s = F_i^S \gamma. \qquad (5)$$

This motivates us to estimate the weight vector γ based on the smoothness assumption on the conditional probability distribution: we compute the optimal weight vector γ by minimizing the difference in predicted labels between two nearby points in the target domain. Specifically, the pro-

posed weighting framework solves the following problem:

$$\min_{\gamma:\gamma^{'}e=1,\gamma\geq 0} \sum_{i,j=n_l+1}^{n_l+n_u} (F_i^S\gamma - F_j^S\gamma)^2 W_{ij} \qquad (6)$$

where $F_i^S\gamma$ and $F_j^S\gamma$ are the predicted labels for i-th and j-th samples of target domain data and W_{ij} is the edge weight between the two samples. We can rewrite the minimization problem as follows:

$$\min_{\gamma:\gamma^{'}e=1,\gamma\geq 0} \gamma^{'}\left(F^S\right)^{'} L_u F^S \gamma \qquad (7)$$

where F^S is an $n_u \times k$ matrix with each row of F^S being the $1 \times k$ vector of k predicted labels for a sample of target domain data and L_u is graph Laplacian associated with the target domain data D_u^T, given by $L_u = D - W$, where W is the adjacency graph defining edge weights between the samples in the target domain data, and D is the diagonal matrix given by $D_{ii} = \sum_{j=1}^n W_{ij}$.

The minimization problem in (7) is a standard quadratic problem and can be solved by applying many existing solvers. We simply use the 'quadprog' function in MATLAB. With the computed weights, the labels for the unlabeled target domain data, called *psuedo labels*, are computed using (3), and are substituted into the regulariser in (4).

Intuitively, by enforcing that nearby points in the marginal distribution of the target data have similar class labels (or conditional probability) via the optimization in (7), the proposed weighting scheme is likely to give higher weights to those sources with the conditional probability distribution similar to the target data. This is verified in our empirical study on both SEMG and synthetic data. If a source has a conflicting conditional distribution as the target, it is likely to get a low or even zero weight. In addition, different from many existing weight schemes which compute the weights by considering each source independently, the proposed weighting scheme computes the optimal value of γ or the optimal weights of all the k sources simultaneously, thus taking the potential interaction among multiple subjects in the source domain into account.

2.4 Proposed Algorithm

Using the least square error and substituting the regularizers we can rewrite (1) as follows:

$$\min_{f^T\in H_K} \gamma_A\|f^T\|_K^2 + \frac{1}{n_l}\sum_{i=1}^{n_l}(f_i^T - y_i^T)^2$$
$$+\frac{\theta}{2}\sum_{j=n_l+1}^{n_T}\|f_j^T - \sum_{s=1}^k \gamma_s f_j^s\|^2 + \frac{\gamma_I}{n_T^2}f^{T'}Lf^T \qquad (8)$$

By the Representer theorem [24], we can find an optimal solution of (8), which is a linear expansion of the kernel function K, over both the labeled D_l^T and the pseudo labeled target domain data D_u^T given as follows:

$$f^T(x) = \sum_{i=1}^{n_l+n_u} \alpha_i K(x_i, x). \qquad (9)$$

Substituting this into (8), we can obtain the optimal $\alpha = [\alpha_1 \cdots\cdots\cdots \alpha_{n_l+n_u}]^T$ by solving the following optimization problem:

$$\min_{\alpha} \frac{1}{n_l + \theta n_u}(Y - K\alpha)^{'} J(Y - K\alpha)$$
$$+\gamma_A\alpha^{'}K\alpha + \frac{\gamma_I}{(n_u + n_l)^2}\alpha^{'}KLK\alpha \qquad (10)$$

where K is the $(n_l+n_u)\times(n_l+n_u)$ kernel Gram matrix over the target domain data, Y is the label vector over labeled and pseudo labeled target domain data points given by:

$$\left[y_1\cdots y_{n_l} \sum_s \gamma_{s(n_l+1)}f_{(n_l+1)}^s \cdots \sum_s \gamma_{s(n_l+n_u)}f_{(n_l+n_u)}^s\right] \qquad (11)$$

L is the graph Laplacian defined over labeled and pseudo labeled target domain data, and J is a diagonal matrix of size $(n_l+n_u)\times(n_l+n_u)$ given by $J = diag(1,\cdots 1, \theta,\cdots\theta)$ with the first n_l diagonal entries as 1 and the rest as θ. θ is assigned a number between 0 and 1, thus the pseudo labels of the target domain data get smaller weights compared to the labels of the labeled target domain data. From (10), the optimal α^* is given by:

$$\alpha^* = \left(JK + \gamma_A(n_l + \theta n_u)I + \frac{\gamma_I(n_l + \theta n_u)}{(n_u + n_l)^2}LK\right)^{-1} JY.$$

With the computed α^*, the prediction of any unseen test data x is given by:

$$f^T(x) = \sum_{i=1}^{n_l+n_u} \alpha_i^* K(x_i, x). \qquad (12)$$

Since the proposed domain adaptation framework is based on multiple sources whose similarities to target domain data or weights are computed based on a conditional probability based weighting scheme, we refer the proposed framework as *Conditional Probability based Multi-Source Domain Adaptation (CP-MDA)*.

3. RELATED WORK

Most of the existing methods measure the similarity between a particular source and the target domain based on the similarity of their marginal probabilities. Shimodaira *et al.* [27] biased the training samples by their test-to-training ratio to match the marginal distribution of the test data. Sugiyama *et al.* [29] proposed to reduce the gap in marginal probabilities by minimizing the KL-divergence between test and weighted training data and Bickel *et al.* [5] discriminated training against test data with a probabilistic model that accounts for the marginal probability difference between training and test distribution. There are several other methods which are also based on marginal probability differences using Maximum Mean Discrepancy [7] as a measure such as Kernel Mean Matching [14] and Transfer Component Analysis [20]. The proposed domain adaptation framework CP-MDA, differs from all these methods in two ways: (1) it is predominantly based on conditional probability differences, and (2) it is based on multiple source domains.

Several algorithms have been developed in past to combine knowledge from multiple sources. Luo *et al.* used consensus maximization as the basis of combining multiple source data [18]. Mansour *et al.* based the transferability of knowledge on a distribution weighted combination of the hypothesis generated by the independent sources [19]. The theoret-

ical proof of both frameworks are based on strong assumptions on the predictive power of the individual source domains on the target domain data. In [26], a clustering based knowledge transfer was proposed for applications with different class labels across source and target domains, unlike the application addressed in this paper.

The proposed framework is related to two multi-source domain adaptation frameworks including Domain Adaptation Machine (DAM) [11] and Locally Weighted Ensemble (LWE) [12]. The proposed framework differs from DAM in the way the weights are computed for different auxiliary sources. In DAM, the weight assigned to each auxiliary source is obtained by measuring the marginal probability distribution difference between the target domain and the particular auxiliary source only, using an empirical estimate of the difference based on the Maximum Mean Discrepancy measure [7]. The proposed framework however computes weights for the auxiliary source data considering predominantly conditional probability distribution of the target data. The weights for all sources are computed in a joint optimization framework, which takes the interaction among multiple auxiliary sources into account.

The proposed framework differs from LWE [12] in that in LWE, the label y of an unlabeled target domain data x is computed using a local weighting ensemble (LWE) scheme:

$$P(y|x) = \sum_{i=1}^{k} w_{M_i,x} P(y|M_i, x) \qquad (13)$$

where $P(y|M_i, x)$ is the prediction made by one of the k models M_i for target data point x and $w_{M_i,x}$ is the weight of the model M_i at point x computed by comparing the similarity graphs of the source and target data around point x. Different from the proposed weighting scheme where we compute all weights in a joint framework, the weight for each auxiliary classifier is computed independently [12].

We also compare our framework with representative single-source domain adaptation algorithms such as Kernel Mean Matching (KMM) proposed by Huang *et al.* [14], Transfer Component Analysis (TCA) proposed by Pan *et al.* [21] and KMapEnsemble (KE) proposed by Zhong *et al.* [31]. KMM re-weights the samples in the source domain so as to minimize the marginal probability difference between the source and target domain using Maximum Mean Discrepancy (MMD) as the measure. TCA is based on feature mapping so as to reduce the marginal probability differences between the source and target distributions again using MMD as the measure. KE differs from the first two algorithms, in which it addresses the conditional probability differences by sample selection after performing a feature mapping step to reduce the marginal probability differences.

There was some classification work dealing with physiological signals using neural networks [17] and linear discriminant analysis [15]; they achieved moderate generalization performance across subjects. To the best of our knowledge we report the first systematic empirical analysis of domain adaptation methods to address the distribution differences due to the subject based variability in physiological signals.

4. EXPERIMENTS

The proposed algorithm has been evaluated on a synthetic dataset and on real-world SEMG data collected from 8 different subjects during a fatigue-causing exercise.
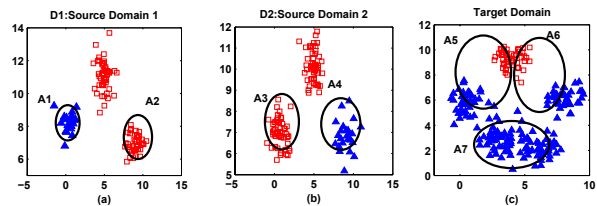
Figure 2: The synthetic dataset includes two source domains (D1 and D2) and one target domain with conflicting conditional probabilities.

4.1 Experimental Setup

4.1.1 Synthetic data

The synthetic data set consisting of two source domains D1 and D2 and a target domain as shown in Figure 2, are generated from several Gaussian distributions with different mean and variances. Source domain D1 has 20 positive and 90 negative examples and source domain D2 has 25 positive and 100 negative examples. There are conflicting conditional probabilities between D1 and D2 as shown in four regions (A1, A2, A3 and A4) in Figure 2. The target domain has 180 positive and 60 negative examples. Region A5 of the target domain is similar to D1, while region A6 is similar to D2. Thus the data exhibits conflicting and varied difference in conditional probability between the source domains and the target domain.

4.1.2 SEMG data

The SEMG data was collected during a repetitive gripping action performed by the forearm. Figure 3 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 3 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 discussed in the introduction.

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 of the order of 1.92 Million samples ($1000 \times 4 \times 60 \times 8$), 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 [16].

Each subject data consists of around 280 to 400 samples of 12 dimensional feature vectors, belonging to four classes with around 70 to 100 samples per class (some subjects who got fatigued sooner and hence could not maintain the required uniform speed for 1 minute the time period was reduced to 30 to 45 secs per phase, hence the number of samples varies between different subjects).

4.1.3 Experimental Procedure

To evaluate the effectiveness of the proposed method, we compare the results with four baseline methods, including
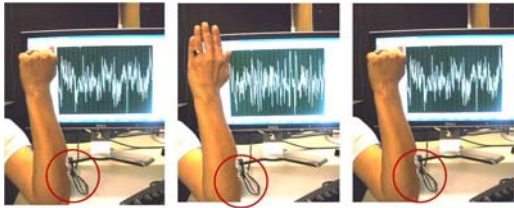
Figure 3: SEMG data collection during a repetitive gripping activity

SVM-C, SVM-M, SMA, and TSVM (Transductive SVM), and two recently proposed multi-source learning methods, including Locally Weighted Ensemble (LWE) [12] and Domain Adaptation Machine (DAM) [11]. More details on LWE and DAM are given in Section 3.

SVM-C refers to *all but one* method where the training data comprises of data from seven subjects and the test data is the data from the remaining subject. SVM-M, refers to the *majority voting* based ensemble framework. The class y assigned to each unlabeled test data x is $\max_y NV(y|x)$ where $NV(y|x)$ is the number of votes given for class y for a particular test sample x by the seven auxiliary sources. SMA refers to *simple model averaging*, which provides equal weight to all the classifiers learned on each auxiliary source domain in an weighted ensemble framework used to generate the label for the target domain data. TSVM refers to Transductive SVM [3] implemented in the svmlight package. It is a semi-supervised method where the training data consists of labeled data from all seven subjects from the source domain and unlabeled data from the target subject.

We vary the number of labeled samples per class in the target domain. DAM(1) and DAM(7) refer to the DAM framework with 1 and 7 target domain labeled samples per class respectively. The proposed CP-MDA method is also implemented using 1 and 7 labeled data from target domain, referred as CP-MDA(1) and CP-MDA(7) respectively. For both cases the unlabeled data from the target domain, is fixed at 10% of the target domain data. The weights of the auxiliary sources computed by the proposed method are also based on this 10% unlabeled target domain data. The rest of the target domain data is treated as unseen target domain data. All the methods are tested on the same pool of *unseen* unlabeled target domain data. The accuracies are computed in a subject independent manner.

We mention here briefly some of the parameters used in implementing the existing and the proposed method. The values of γ_A and γ_I were kept as 0.014 and 0.01 respectively, as suggested in [4]. The Laplacian graph matrix used in calculating the weights was set as 'binary' type based on the N nearest neighbors with $N = 10$. The value of Θ was estimated via 5-fold cross validation on the set $\{i10^{-2}|i = 0, 1, \cdots, 100\}$.

4.2 Results and Analysis

4.2.1 Synthetic data

Figure 4 shows the classification accuracies of different methods on the synthetic dataset. The classification accuracies on the target domain data when the SVM classifier is applied on the source data D1 or D2 without using any transfer learning methodologies are 62.97% and 72.61%, respectively.
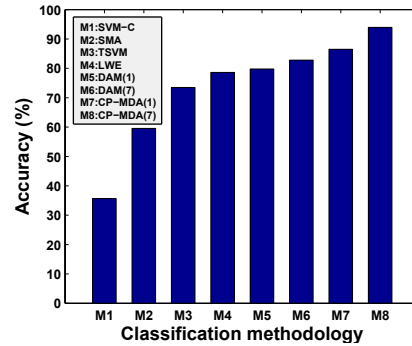

Figure 4: Comparative performance of different methods on the synthetic data - Accuracy (%)

We observe that SVM-C and SMA perform poorly. This may be attributed to the conflicting conditional probabilities between the two training sets, that degrade the performance of the classifier when learned on the combined data. In addition, there is a conditional probability difference with the test data, hence the brute-force use of knowledge without evaluating similarities between the source and the target data degrades the performance of the classifier in the target domain [25] related to *negative transfer* [23].

We also observe that CP-MDA(1) improves the classification accuracy by 20% to 50% over the baseline methods without any transfer learning and by 11% over LWE. We also observe that CP-MDA(1) outperforms DAM(7). Since the target classifier is learned using some labeled and some pseudo labeled data, which is labeled using the proposed weighting scheme, hence higher classification accuracy verifies the effectiveness of the proposed weighting scheme in Section 2.3.

To further evaluate the effectiveness of the proposed weighting scheme, we compared three different weighting schemes using the weighted ensemble methodology of LWE in (13). MMD-WE refers to the method when weights are computed based on MMD, and CP-MDA-WE refers to the method when weights are computed based on the proposed framework in (7). The accuracies using LWE, MMD-WE, and CP-MDA-WE are 78.6%, 72.75%, and 80.54%, respectively. To examine why the proposed method outperforms the other two, we report the predictive power of each of the two source data (D1, D2) as well as the weights obtained by different methods. The classification accuracies obtained using the source data D1 and D2 individually are 62.97% and 72.61%, respectively. The weights for D1 and D2 obtained by MMD-WE are 0.93 and 0.06, respectively; the weights for D1 and D2 obtained by CP-MDA-WE are 0.08 and 0.91, respectively. Thus, the weights obtained by the proposed method match the predictive powers of the source data better than MMD-WE.

4.2.2 SEMG data

We compare different methods including SVM-C, SVM-M, SMA, TSVM, LWE, DAM and the proposed CP-MDA. The results are summarized in Table 1. The first column of the table indicates the subject data under test (target domain). The training data (source domain) consists of the data from the remaining seven subjects. Similar to the results obtained in the case of synthetic data we see that

Table 1: Comparative performance of different methods on SEMG data - Accuracy (%)

Test Sub	SVM-C	SVM-M	SMA	TSVM	LWE	DAM(1)	DAM(7)	CP-MDA(1)	CP-MDA(7)
1	70.76	33.9	44.96	49.09	67.44	74.83	77.43	81.93	85.25
2	43.69	50.76	44.61	55.68	77.54	81.36	83.35	84.73	87.7
3	50.11	56.85	56.84	65.09	75.55	74.77	78.99	82.45	85.06
4	59.65	47.93	49.67	56.98	81.22	80.63	84.32	81.27	86.4
5	40.37	44.79	50.15	62.5	52.48	76.74	81.14	80.74	86.62
6	59.21	61.45	60.33	71.32	65.77	59.21	74.28	83.12	88.09
7	47.13	46.91	45.76	60.73	60.32	74.27	83.31	81.57	86.4
8	69.85	64.53	74.46	68.55	72.81	84.55	86.6	88.5	90.56
Average	55.09	50.85	53.34	61.24	69.14	75.79	81.18	**83.04**	**87.01**

Table 2: Comparison of SVM-T, DAM, and CP-MDA on Subject 6 (top) and Subject 7 (bottom) in terms of accuracy (%) when the number of labeled target domain data per class varies.

Method	Number of labeled data per class					
	1	2	3	4	6	7
SVM-T	4.26	4.26	49.09	73.63	84.69	85.67
DAM	59.12	59.21	59.35	59.59	65.03	74.28
CP-MDA	**83.12**	83.12	85.45	87.58	87.77	88.09
SVM-T	10.59	45.5	77.79	83.25	85.48	84.97
DAM	74.27	75.11	79.10	81.19	82.43	83.3
CP-MDA	**81.57**	83.99	85.81	86.24	86.32	86.4

SVM-C, SVM-M, SMA, and TSVM perform very poorly. We observe significant improvement in classification accuracy when domain adaptation methodologies are employed. The proposed method CP-MDA(1) provides a 20% to 30% improvement over the baseline methods including SVM-C, SVM-M, SMA and TSVM. The classification accuracies of the proposed method are in average 13% higher than LWE. It is also observed as in the case of synthetic data that CP-MDA(1) performs not only better than DAM(1) but also better than DAM(7) in 5 out of 8 cases. These results verify the effectiveness of the proposed method.

Next, we evaluate the performance of CP-MDA when the number of labeled target domain data varies. We compare CP-MDA with DAM and SVM-T. SVM-T refers to an SVM classifier trained on the labeled target domain data. The results for two subjects are summarized in Table 2; we obtain similar results for the other six subjects and the results are omitted. We can observe from the table that when the number of labeled target domain data per class is small, e.g., 1 to 4 samples per class, both domain adaptation methods perform much better than SVM-T. But with an increasing number of labeled data from the target domain the accuracies become comparable. However the proposed method always performs better than the other two methods. This result demonstrates that domain adaption is especially useful when the amount of labeled target domain data is small.

Similar to the synthetic data case, we compare the performance of the weighting schemes used in LWE, DAM and CP-MDA. Table 3 summarizes the results for different test cases. We observe that CP-MDA-WE performs better than the other methods in 6 out of 8 cases, and LWE performs better in the remaining 2 cases. Recall that like CP-MDA-WE, LWE computes weights for the auxiliary source domain based on the conditional probability differences between the source and target domains, while MMD-WE computes the weights based on the marginal probability differences only. Since SEMG data has significant conditional probability differences, CP-MDA-WE and LWE are expected to outperform DAM-WE.

Table 3: Comparison of different weighting schemes for different test subjects - Accuracy (%).

Test Sub	LWE	MMD-WE	CP-MDA-WE
1	67.44	68.27	75.12
2	77.54	69.48	83.23
3	75.55	71.84	75.68
4	81.22	62.65	81.09
5	52.48	68.32	78.16
6	65.77	58.91	76.11
7	60.32	67.75	75.07
8	72.81	66.11	78.71
Average	69.14%	66.66%	**77.89**%

The proposed algorithm CP-MDA computes weights for each class for each of the auxiliary source domain data, thus exploiting the similarities and dissimilarities at the class level. Table 4 shows the weights for four different classes assigned to each training subject in the source domain for test subject 1 in the target domain. We observe that the proposed weighting scheme assigns different weights to different auxiliary source domain data (subject data) for different classes. Subjects 5 and 8 get higher weights for class 1, subject 7 gets a higher weight for class 2, subject 5 gets a higher weight for class 3, and for class 4 subject 4 gets a higher weight. We observe from Figure 1 that the data distribution of class 4 of subject 1 is very similar to that of class 4 of subject 4.

Table 4: Weights computed by CP-MDA for four different classes for each of the source domain subjects 2-8 for test target subject 1.

Class	Target subject						
	2	3	4	5	6	7	8
1	0	0	0.02	0.50	0.48	0	0
2	0	0.01	0.03	0	0.11	0.74	0.11
3	0	0.02	0.12	0.75	0	0.01	0.11
4	0.09	0.02	0.66	0.11	0.11	0.01	0

One of the key advantage of the proposed algorithm is that it exploits the information from multiple source domains for classifying the target data. It will be interesting to study how the number of sources used in the training set affects the classification. Figure 5 presents the error rates
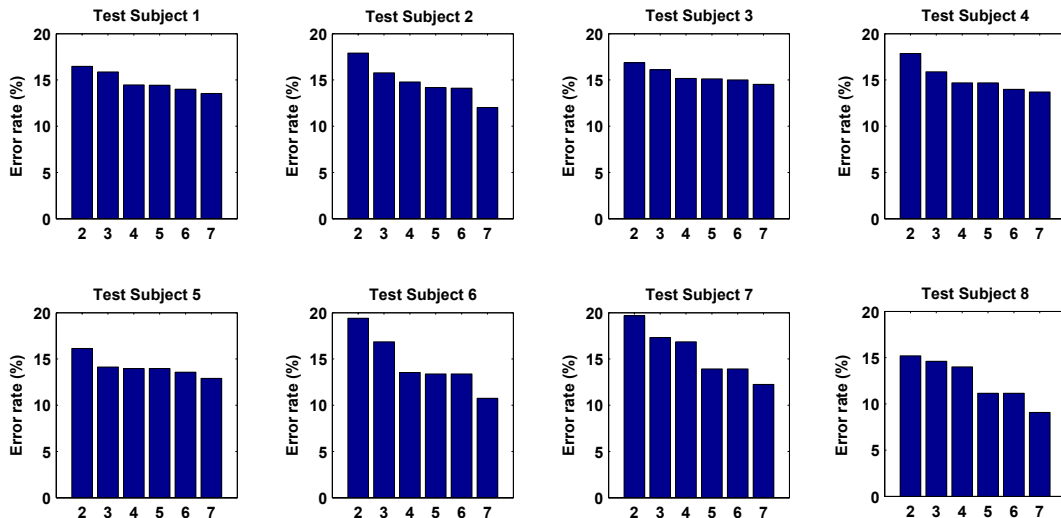
Figure 5: The effect of the number of auxiliary source domains (horizontal axis) in the training set on the proposed CP-MDA algorithm in terms of the classification error rates (%) for all eight subjects.

obtained for each test subject when the number of subjects in the source domain varies (from 1 to 7); we simply keep adding subjects to the training set in increasing order of the subject number. We observe that for all subjects, the error rate decreases monotonically when the number of subjects increases. These results demonstrate the effectiveness of the proposed algorithm for extracting useful information from multiple sources.

Table 5: Comparison of CP-MDA with three single source domain adaptation algorithms (KMM, TCA, and KE) - Accuracy(%).

Test Sub	CP-MDA(7)	KMM	TCA	KE
1	85.25	65.15	45.15	71.85
2	87.7	46.96	68.93	74.62
3	85.06	59.55	56.78	74.79
4	86.4	73.38	52.68	69.35
5	86.62	45.31	60.15	73.44
6	88.09	70.62	76.92	83.92
7	86.4	51.13	55.64	77.97
8	90.56	42.79	67.24	79.48
Average	**87.01**	56.86	53.84	75.67

To evaluate the benefit of a multi-source domain adaption framework for addressing subject based variability, we compare the proposed algorithm with three representative single-source domain adaption algorithms discussed in Section 3. Table 5 summarizes the classification accuracies obtained by different methods for each of the test subjects. The target data is from the subject shown in column 1 and the source data consists of the combined data from the remaining seven subject. Classification results were averaged over 10 runs with different sets of randomly selected 7 labeled samples per class from the target domain data. We can observe from the table that combining all the subject data and forming a single domain degrades the performance. We also observe that among the three single domain adaption algorithms, KMM or TCA which consider the marginal

probability differences only perform worse than KE. These results are expected as SEMG data has significant conditional probability differences. Our results demonstrate the effectiveness of the proposed multi-domain framework for dealing with subject based variability in SEMG data.

5. CONCLUSION

We consider the characterization of muscle fatigue through noninvasive sensing mechanism such as surface electromyography (SEMG). The variation in SEMG parameters from subject to subject creates a difference in the data distribution, thus traditional data mining algorithms may not be effective. In this paper, we propose a transfer learning framework based on the multi-source domain adaptation methodology for detecting different stages of fatigue using SEMG signals. The proposed CP-MDA framework differs from most existing work in domain adaption in that it addresses the conditional probability distribution differences across domains based on a novel weighting scheme. We have validated the proposed framework using a synthetic dataset and 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 with SEMG data. This increases the scope of application of SEMG data for fatigue detection in a larger scale beyond training and rehabilitation settings for monitoring during daily life and industrial work, as it alleviates the need for subject specific calibration for assessing the physiological state across subjects.

We plan to extend the proposed framework to applications involving other types of physiological signals for developing generalized models across subjects for emotion and health monitoring in everyday life, industrial work and geriatric care [17, 15]. In addition, we plan to study the generalization performance of the proposed framework.

6. ACKNOWLEDGEMENTS

7. REFERENCES

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

[2] http://en.wikipedia.org/wiki/Factor_analysis.

[3] http://svmlight.joachims.org/.

[4] M. Belkin, P. Niyogi, and V. Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *JMLR*, 7:2399–2434, 2006.

[5] S. Bickel, M. Brückner, and T. Scheffer. Discriminative learning under covariate shift. *JMLR*, 10:2137–2155, 2009.

[6] J. Blitzer, M.Dredze, and F. Pereira. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL*, 2007.

[7] K. Borgwardt, A. Gretton, M. Rasch, H. Kriegel, Schölkopf, and A. J. Smola. Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, 22(14):49–57, 2006.

[8] R. Chattopadhyay, P. Gaurav, and P. Sethuraman. A generalized machine learning framework for continuous monitoring of physiological conditions based on fatigue and intensity of activity in daily living. In *WIML, NIPS*, 2009.

[9] R. Chattopadhyay, P. Sethuraman, and P. Gaurav. Towards fatigue and intensity measurement framework during continuous repetitive activities. In *I2MTC*, 2010.

[10] H. Daumé III. Frustratingly easy domain adaptation. In *ACL*, 2007.

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

[12] J. Gao, W. Fan, J. Jiang, and J. Han. Knowledge transfer via multiple model local structure mapping. In *KDD*, pages 283–291, 2008.

[13] P. Higgs. Upper extremity impairment in workers performing repetitive tasks. *Plastic and reconstructive surgery*, 90:614, 1992.

[14] J. Huang, A. J. Smola, A. Gretton, K. Borgwardt, and B. Schölkopf. Correcting sample selection bias by unlabeled data. In *NIPS*, 2007.

[15] J. Kim and E. Andre. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(12):2067–2083, 2008.

[16] M. Knaflitz and P. Bonato. Time-frequency methods applied to muscle fatigue assessment during dynamic contractions. 9(5):337–350, Oct. 1999.

[17] E. Leon, G. Clarke, V. Callaghan, and F. Sepulveda. A user independent real time emotion recognition system for software agents in domestic environment. *Engineering Applications of Artificial Intelligence*, 20(3):337–345, 2007.

[18] P. Luo, F. Zhuang, H. Xiong, Y. Xiong, and Q. He. Transfer learning from multiple source domains via consensus regularization. In *CIKM*, 2008.

[19] Y. Mansour, M. Mohri, and A. Rostamizadeh. Domain adaptation with multiple sources. In *NIPS*, 2009.

[20] S. J. Pan, J. T. Kwok, and Q. Yang. Transfer learning via dimensionality reduction. In *AAAI*, 2008.

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

[22] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 2009.

[23] M. Rosenstein, Z. Marx, and L. Kaelbling. To transfer or not to transfer. In *NIPS 2005 Workshop on Inductive Transfer: 10 Years Later*, 2005.

[24] B. Schölkopf and A. J. Smola. *Learning with Kernels Support Vector Machines, Regularization, Optimization and Beyond*. The MIT Press, 2002.

[25] G. Schweikert, C. Widmer, B. Schölkopf, and G. Rätsch. An empirical analysis of domain adaptation algorithms for genomic sequence analysis. In *NIPS*, 2009.

[26] X. Shi, W. Fan, Q. Yang, and J. Ren. Relaxed transfer of different classes via spectral partition. In *KDD*, 2009.

[27] H. Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. In *JSPI*, 2000.

[28] B. Silverstein, L. Fine, and T. Armstrong. Hand wrist cumulative trauma disorders in industry [1351-0711]. *Occupational and Environmental Medicine*, 43:779, 1986.

[29] 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.

[30] 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. In *AJIM*, 1995.

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