

Adaptive Learning to Speed-Up Control of Prosthetic Hands: a Few Things Everybody Should Know

Valentina Gregori¹, Arjan Gijsberts¹ and Barbara Caputo¹

Abstract—A number of studies have proposed to use domain adaptation to reduce the training efforts needed to control an upper-limb prosthesis exploiting pre-trained models from prior subjects. These studies generally reported impressive reductions in the required number of training samples to achieve a certain level of accuracy for intact subjects. We further investigate two popular methods in this field to verify whether this result equally applies to amputees. Our findings show instead that this improvement can largely be attributed to a suboptimal hyperparameter configuration. When hyperparameters are appropriately tuned, the standard approach that does not exploit prior information performs on par with the more complicated transfer learning algorithms. Additionally, earlier studies erroneously assumed that the number of training samples relates proportionally to the efforts required from the subject. However, a repetition of a movement is the atomic unit for subjects and the total number of repetitions should therefore be used as reliable measure for training efforts. Also when correcting for this mistake, we do not find any performance increase due to the use of prior models.

I. INTRODUCTION

A majority of upper-limb amputees is interested in prostheses controlled via surface electromyography (sEMG), but they perceive the difficult control as a great concern [1]. Machine learning has opened a new path to tackle this problem by allowing the prosthesis to adapt to the myoelectric signals of a specific user. Although these methods have been applied with success in an academic setting (e.g., [2] and references therein), they require a long and painful training procedure to learn models with satisfactory performance.

Several studies have proposed to reduce the amount of required training data by leveraging over previous models from different subjects [3, 4, 5]. The underlying idea is that a model for a new target user can be bootstrapped from a set of prior source models. Though the idea is appealing and initial studies have shown remarkable improvements, there is no conclusive evidence that these strategies lead to tangible benefits in the real world. An obvious limitation in the earlier studies is that they only considered intact subjects. This is relevant since myoelectric signals are user-dependent and this holds in particular for amputees, as the amputation and subsequent muscular use have a considerable impact on the quality on the myoelectric signals [6].

Other limitations relate to the technical and conceptual execution of the experimental validation. First, the hyperparameters of the algorithms were not optimized for the method

at hand, but rather chosen based on how well they performed on average when applied on the data of other subjects. Individual hyperparameter optimization for the methods and number of training samples is crucial for the successful application of machine learning algorithms and omission of this procedure may skew results. For instance, this tuning procedure may give an unfair disadvantage to the baseline that does not use prior information from pre-trained users.

On the conceptual level, the previous studies used the number of training samples (i.e., windows of the myoelectric signals) as measure for the required training effort. A consequence of this strategy is that not all available training samples were used to classify the movements, since subjects cannot produce individual samples. Instead, a repetition of a movement consisting of multiple windows is the atomic unit for subjects. This artificial reduction of training information is likely to be disadvantageous for the baseline that only relies on target training data. In our evaluation, we instead use the number of movement repetitions as realistic measure of the training effort. Furthermore, we consider all possible data, subjects and combinations of training repetitions, thereby removing the possible effects of random selection from the evaluation.

In this paper, we provide more insight into the benefits of domain adaptation for prosthetic control by augmenting the experiments by Patricia et al. [5] with amputated subjects while also addressing other limitations. This results in three experimental settings, namely (1) the original experiments according to the setup common in literature [3, 4, 5] extended with amputated subjects, (2) the same setup with hyperparameter optimization and finally (3) a realistic setup where we also address the conceptual issues. In each setting, we perform three experiments, where intact and amputated subjects make up the groups of target and source subjects.

This paper is structured as follows. In Section II we present the related work on domain adaptation in the context of myoelectric prosthetics. The algorithms that will be considered in our experiments will then be explained in detail in Section III. We continue with our experimental setup in Section IV, after which we will present the results in Section V. Finally, we conclude the paper in Section VI.

II. RELATED WORK

One of the first attempts to classify myoelectric signals of three volunteers was by Graupe and Cline [7]. In the following years, studies on prosthetic control led to many advances in the analysis and understanding of sEMG. Castellini et al. [2] noted that myoelectric signals differ significantly

*This work was supported by the Swiss National Science Foundation Sinergia project #160837 “Megane Pro”

¹Department of Computer, Control, and Management Engineering, University of Rome La Sapienza, via Ariosto 25, 00185 Roma, Italy
surname@dis.uniroma1.it

from person to person and that models trained for different subjects are therefore not automatically reusable. However, they showed that a pre-trained model could be used to classify samples from similar subjects.

Several studies continued in this direction with different strategies to build more robust models that take advantage of past information from source subjects or, in the context of repeatability, from the target itself. Matsubara et al. [8] proposed to separate myoelectric data in user-dependent and motion-dependent components, and to reuse models by quickly learning just the user-dependent component for new subjects. Sensinger et al. [9] presented different methods based on an appropriate concatenation of target and source data. Others still approached the problem by searching for a mapping to project data from different subjects into a common domain [10, 11]; a similar strategy was also used to reduce the recalibration time for a target that attempts to use the prosthesis on different days [12].

Other studies proposed to leverage over prior models from already trained source subjects without requiring direct access to their data [3, 4, 5, 13]. They tested different types of so-called Hypothesis Transfer Learning (HTL) algorithms showing a gain in performance with respect to non-adaptive baselines. We are particularly interested in the findings of Tommasi et al. [4], Patricia et al. [5], who worked with a significant number of classes and intact subjects from the public Non-Invasive Adaptive Prosthetics (NinaPro) database [14]. They report that the number of training samples required to obtain a given level of performance can be reduced by an order of magnitude as compared to learning from scratch.

III. ALGORITHMS

We first describe the mathematical background by means of a base learning algorithm in Section III-A, then we proceed with the domain adaptation methods included in our evaluations in Section III-B.

A. Background

Let us define a training dataset $D = \{\mathbf{x}_i, y_i\}_{i=1}^N$ of N input samples $\mathbf{x}_i \in \mathcal{X} \subseteq \mathbb{R}^d$ and corresponding labels $y_i \in \mathcal{Y} = \{1, \dots, G\}$. In the context of myoelectric classification, the inputs are the myoelectric signals and the labels are the movements chosen from a set of G possible classes. The goal of a classification algorithm is to find a function $h(\mathbf{x})$ that, for any future input vector \mathbf{x} , can determine the corresponding output y . Among the algorithms that construct such a model, Support Vector Machines (SVMs) are some of the most popular.

The base of the domain adaptation algorithms described later on is the Least-Squares Support Vector Machine (LS-SVM) [15], a variant of SVM with a squared loss and equality constraint. It writes the output hypothesis as $h(\mathbf{x}) = \langle \mathbf{w}, \phi(\mathbf{x}) \rangle + b$, where \mathbf{w} and b are the parameters of the separating hyperplane between positive and negative

samples. The optimal solution is thus given by

$$\min_{\mathbf{w}, b} \left\{ \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{2} \sum_{i=1}^N \xi_i^2 \right\} \quad (1)$$

$$\text{s.t. } y_i = \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle + b + \xi_i, \quad \forall i \in \{1, \dots, N\},$$

where C is a regularization parameter and ξ denotes the prediction errors. We approach our multi-class classification problem via a one-vs-all scheme to discriminate each class from all others. To obtain a better solution we mapped the input vectors \mathbf{x}_i into a higher dimensional feature space using $\phi(\mathbf{x}_i)$. Usually, this mapping $\phi(\cdot)$ is unknown and we work directly with the kernel function $K(\mathbf{x}', \mathbf{x}) = \langle \phi(\mathbf{x}'), \phi(\mathbf{x}) \rangle$ [15]. In the following, we use a Radial Basis Function (RBF) kernel

$$K(\mathbf{x}', \mathbf{x}) = e^{-\gamma \|\mathbf{x}' - \mathbf{x}\|^2} \quad \text{with } \gamma > 0. \quad (2)$$

B. Adaptive Learning

Domain adaptation algorithms construct a classification model for a new target using past experience from the sources. More specifically, let us assume that we have K different sources, where each source is a classification model for the same set of movements. The used HTL algorithms can then be described as follows.

1) *Multi Model Knowledge Transfer*: This method aims to find a new separating hyperplane \mathbf{w} that is close to a linear combination of the pre-trained source hypotheses $\hat{\mathbf{w}}^k$ [4, 16]. We solve the optimization problem

$$\min_{\mathbf{w}, b} \left\{ \frac{1}{2} \left\| \mathbf{w} - \sum_{k=1}^K \beta^k \hat{\mathbf{w}}^k \right\|^2 + \frac{C}{2} \sum_{i=1}^N \xi_i^2 \right\} \quad (3)$$

$$\text{s.t. } y_i = \langle \mathbf{w}, \phi(\mathbf{x}_i) \rangle + b + \xi_i.$$

The vector $\beta = [\beta_1, \dots, \beta_K]^T$ with $\beta_k \geq 0$ and $\|\beta\|^2 \leq 1$ represents the contribution of each source in the target problem and is obtained by optimizing a convex upper bound of the leave-one-out misclassification loss [16]. A more general case consists of different weights for different classes of the same source, such that $\beta_{k,g}$ is the weight associated to class g of source k . In this work we used this latter version of the algorithm.

2) *Multi Kernel Adaptive Learning*: This algorithm combines source and target information via a linear combination of kernels [17, 18]. Let us define

$$\bar{\mathbf{w}} = [\mathbf{w}^0, \mathbf{w}^1, \dots, \mathbf{w}^K] \quad \text{and} \quad (4)$$

$$\bar{\phi}(\mathbf{x}, y) = [\phi^0(\mathbf{x}, y), \phi^1(\mathbf{x}, y), \dots, \phi^K(\mathbf{x}, y)] \quad (5)$$

respectively as the concatenation of the target and source hyperplanes and the mapping functions into the corresponding feature spaces. These are both composed of $(K + 1)$ elements: the first refers to the target and the remaining ones to the sources. The optimization problem becomes

$$\min_{\bar{\mathbf{w}}} \left\{ \frac{\lambda}{2} \|\bar{\mathbf{w}}\|_{2,p}^2 + \frac{1}{N} \sum_{i=1}^N \xi_i \right\}$$

$$\text{s.t. } \langle \bar{\mathbf{w}}, (\bar{\phi}(\mathbf{x}_i, y_i) - \bar{\phi}(\mathbf{x}_i, y)) \rangle \geq 1 - \xi_i, \quad \forall i, y \neq y_i. \quad (6)$$

The element p regulates the level of sparsity in the solution \bar{w} and can vary in the range $(1, 2]$. The solution is obtained via stochastic gradient descent during T epochs over the shuffled training samples.

IV. EXPERIMENTAL SETUP

The experimental evaluation is subdivided in three settings. The first one is modeled after related literature for this kind of experiments with sEMG data [3, 4, 5]. The second is identical but adds hyperparameter optimization for the target models. The third and final one is a novel framework in which we fixed the shortcomings of the previous settings to make the experiments as realistic as possible. We will refer to the settings as original, optimized and realistic. In the following, we first explain the used data and classifiers, and subsequently elaborate on the experimental settings.

A. Data

The data used in our work are from the NinaPro database¹ [14], the largest publicly available database for prosthetic movement classification with 40 intact subjects and 11 amputees. Each subject executed 40 movements for 6 times, such that each repetition was alternated with a rest posture. While performing the movements, twelve electrodes acquired sEMG data from the arm of the subject. The standardized data were used according to the control scheme by Englehart and Hudgins [19], where we extracted features from a sliding window of 200 ms and an increment of 10 ms. The resulting set of windows was subsequently split in train and test sets for the classifier; data from repetitions $\{1, 3, 4, 6\}$ were dedicated to training while data from repetitions 2 and 5 were used as test. To reduce the computational requirements, we subsampled the training data by a factor of 10 at regular intervals.

B. Classifiers

The algorithms used to build the classification models were the two mentioned HTL algorithms together with two baselines:

- the *no transfer model (NoTransfer)*, which uses an LS-SVM with RBF kernel trained only on the target data. This corresponds to learning without the help of prior knowledge.
- the *prior model (Prior)*, which learns an LS-SVM with linear kernel on top of the raw predictions of the source models. This measures the relevance of the source hypotheses by using them as feature extractors for the target data.
- *Multi Model Knowledge Transfer (MultiKT)*, as explained in Section III-B.1, which learns a model on the target data that is close to a weighted combination of the source hypotheses.
- *Multi Kernel Adaptive Learning (MKAL)*, as explained in Section III-B.2, which linearly combines an RBF kernel on the target data with the source predictions. Parameters p and T were set to 1.04 and 300.

The classification models for the sources were based on a non-linear SVM or LS-SVM with RBF kernel.

C. Settings

For each of the settings, we ran three experiments with distinct groups of target and source subjects:

- Intact-Intact: intact target subjects exploit prior knowledge of other intact sources;
- Amputees-Amputees: amputated target subjects exploit previous experience of other amputees;
- Amputees-Intact: amputated target subjects exploit prior knowledge of intact subjects.

In the first and second experiment, each subject took the role of target just once, while the remaining subjects were used as sources. In the third, all of the amputees were once the target and the set of intact subjects was used only as sources.

1) *Original Setting*: The purpose of the original and optimized settings is to investigate the isolated impact of hyperparameter optimization on the performance of the target classifiers. We therefore replicated, as closely as possible, the experiments from Patricia et al. [5] with 9 amputees² and a random subset of 20 intact subjects from the NinaPro database. For these subjects we considered 17 movements plus the rest posture, appropriately subsampled to balance it with the other movements. The sEMG representation used in this setting was the average of Mean Absolute Value (MAV), Variance (VAR) and Waveform Length (WL) features [20], as to reduce the dependency on one specific type of representation. The details of this and the subsequent settings are presented schematically in Table I.

The source models were created by training an SVM with RBF kernel using all training repetitions of the respective subject. For the target models we trained the classifiers on an increasing number of random samples, from 120 to 2160 in steps of 120, from the training repetitions. The hyperparameters for both the source and target models were chosen from $C, \gamma \in \{0.01, 0.1, 1, 10, 100, 1000\}$ and kept constant regardless of the number of training samples. For each parameter configuration, we evaluated the average balanced classification accuracy of each source subject when tested on the target subjects. For the target subject and its source models, we then chose the configuration that maximizes this average, making sure to exclude the data from the target subject. The motivation for this procedure is that biased regularization in MultiKT requires the source and target models to “live” in the same space.

2) *Optimized Setting*: Strictly speaking, the above assumption only requires that all the sources have the same RBF bandwidth γ as the related target. Moreover, MultiKT can also be interpreted as predicting the difference between the source predictions and the true labels [21]. In this alternative interpretation, there is no need for source and target models to use the same kernel. We therefore tuned the hyperparameters in the optimized setting for each individual

¹<http://ninapro.hevs.ch/>

²We omitted two amputees from the database that had only 10 electrodes due to insufficient space on their stump.

TABLE I
EXPERIMENTAL SETTINGS.

Setting	Movements	# Int./Amp.	Features	Source model	Source hyperparameters	Target hyperparameters
Original	9 wrist, 8 finger, subsampled rest	20/9	avg. MAV/VAR/WL	SVM	balanced accuracy on other subjects	same as source
Optimized	9 wrist, 8 finger, subsampled rest	20/9	avg. MAV/VAR/WL	SVM	balanced accuracy on other subjects	balanced accuracy of 5-fold CV over shuffled training samples of target
Realistic	9 wrist, 8 finger, 23 grasp	40/8	mDWT	LS-SVM	accuracy of k-fold CV over repetitions of source	accuracy of k-fold CV over train repetitions of target

target and for each training set size based on 5-fold cross validation (CV) on the target training set. Note, however, that we still use the original method to determine the parameters for the source models.

3) *Realistic Setting*: In the final setting we extended the hyperparameter optimization procedure and attempted to address all issues to make the experiments as realistic as possible. First, we considered all available movements and subjects in the NinaPro database³. As sEMG representation we used marginal Discrete Wavelet Transform (mDWT) features, which have previously shown excellent performance in related work on this database [22].

The main conceptual innovation with respect to the previous settings is that we trained target models on an increasing number of repetitions. The motivation is that the effort of the subject during data acquisition is given by the required number of repetitions of each movements, so we analyze the accuracy as a function of this atomic unit. Given the set of training repetitions $\{1, 3, 4, 6\}$, we considered all possible subsets of length between 1 and 4 repetitions. For all these cases, we optimized the target model using k -fold CV, where each fold corresponded to samples belonging to one repetition. In the exceptional case of only a single training repetition, we instead used 5-fold CV over the samples. The parameter grid was extended to $C \in \{2^{-6}, 2^{-4}, \dots, 2^{12}, 2^{14}\}$ and $\gamma \in \{2^{-20}, 2^{-18}, \dots, 2^{-2}, 2^0\}$. Models for the source subjects on the other hand were trained using all repetitions and the hyperparameters were optimized specifically for the individual subject using 6-fold CV, where the folds again corresponded to the repetitions. The source models were built with LS-SVM instead of SVM to be more coherent with the other classifiers, which are all derived from LS-SVM. Due to the much larger number of samples in this realistic setting⁴ we further subsampled the data used for hyperparameter optimization by a factor of 4. For the same reason we decided to omit MKAL from the analysis.

V. EXPERIMENTS

In this section we first investigate the isolated impact of hyperparameter optimization when applied to the original setting. Then we verify whether the findings also apply to

the realistic setting described in Section IV-C.3. An in-depth discussion follows on the explanations of the results.

A. Results

In Figure 1 we report the balanced classification accuracy as a function of the number of training samples averaged over all target subjects. The dotted lines indicate the results obtained in the original experimental framework usually employed in literature (see Section IV-C.1). As in the related studies, MKAL and MultiKT outperform the baselines NoTransfer and Prior by a significant margin for all training set sizes. This has led to the claim that the adaptive algorithms can achieve similar performance as NoTransfer using an order of magnitude less training samples. Since this improvement is observed whether the target and source subjects are intact or amputated, it is also assumed that amputees can equally exploit prior information from intact as well as other amputated subjects.

When looking at the solid lines in Figure 1, which show results with hyperparameter optimization, we observe that the discrepancies between the algorithms disappear. In other words, the NoTransfer baseline performs just as well as or even slightly better than the adaptive algorithms. Furthermore, with hyperparameter optimization all methods now outperform the results in the original setting. The only exception to this observation is Prior, which has lower accuracy in the Amputee-Intact experiment. Contrary to the earlier statements, this demonstrates that prior models from intact subjects are not as useful as those from other amputees. Together with the observation that MKAL and MultiKT perform nearly identically to NoTransfer, this also allows us to conclude that rather than transferring from prior models, the HTL algorithms rely almost exclusively on target data.

Figure 2 shows the standard classification accuracy for the realistic setting described in Section IV-C.3 averaged over the target subjects and all possible combinations of a given number of training repetitions. Also in this setting the hyperparameters were tuned appropriately and the differences among the methods are again negligible. In addition, we observe significantly lower accuracy among amputees compared to intact subjects, confirming the deterioration of the myoelectric signals due to amputation and subsequent lack of muscular use.

³Data for the first amputated subject was omitted, since the acquisition was interrupted prematurely.

⁴Each repetition consists of approximately 35000 samples.

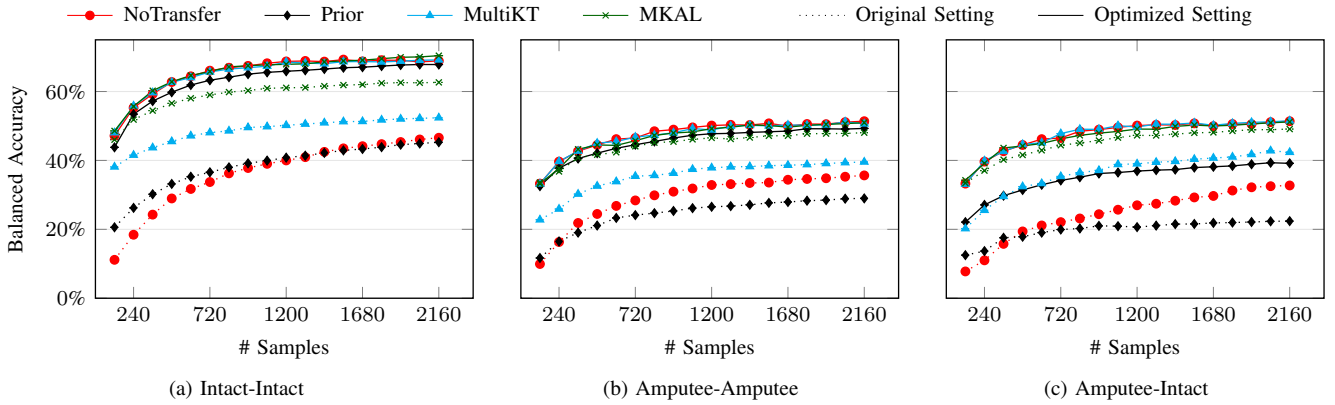


Fig. 1. Balanced classification accuracy for MultiKT, MKAL, NoTransfer and Prior in the original and hyperparameter optimized settings.

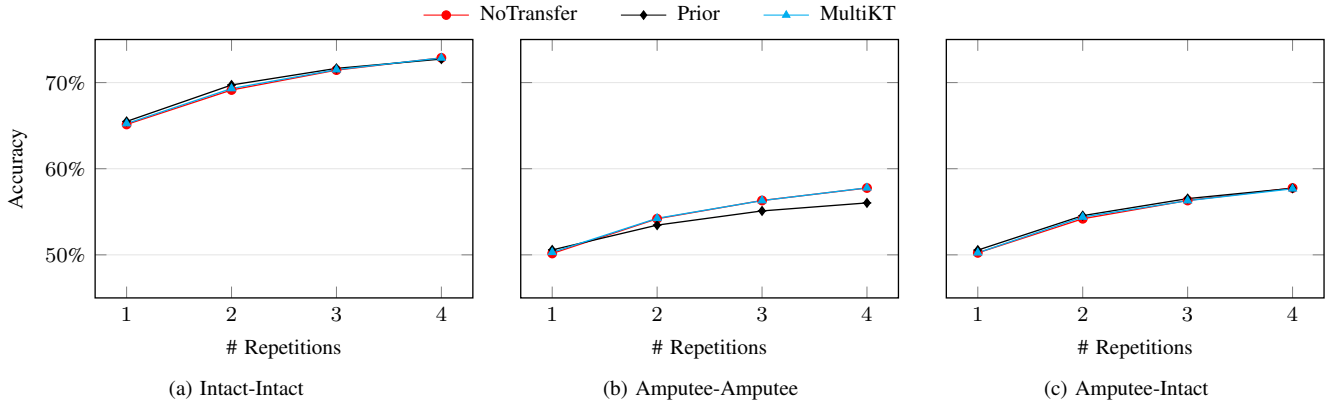


Fig. 2. Standard classification accuracy for MultiKT, NoTransfer and Prior in the realistic setting.

B. Discussion

The results clearly show that the improvements usually attributed to prior knowledge can instead be explained by sub-optimal hyperparameter optimization. With properly tuned hyperparameters, the NoTransfer baseline that completely ignores source information performs as well as the more complicated domain adaptation methods.

There are multiple explanations for the observed differences in performance in the original setting. First, the hyperparameters were chosen based on the performance of an SVM when transferring from the source subjects to the target subjects. This gives a disadvantage to NoTransfer, which does not exploit prior knowledge to train the classifier. Furthermore, this parameter setting is also problematic since all methods are based on LS-SVM, which uses a different loss function than SVM. As can be seen in the objective function in Equation 1, the regularization parameter C is multiplied with the absolute magnitude of all training losses, so an optimal setting for SVM does not necessarily work well for LS-SVM.

A further problem is that the value of C was determined using the total set of training samples. The same value was subsequently used when training on much smaller subsets, leading to a different tradeoff between minimizing training errors and regularizing the solution. This affects all methods except MKAL, for which the specific implementation mul-

tiplied the given value of C with the number of training samples. MKAL therefore effectively used a much larger value of C (i.e., much less regularization), explaining why it performed superior to the other methods.

A similar, though slightly more complicated, argument holds for MultiKT. Recall the formulation of biased regularization in Equation 3; the linear combination of source hypotheses allows to reduce the effect of the regularization term by moving the bias in the direction of the optimal solution. In other words, for the same value of C this allows to concentrate more on minimizing the training errors on the target data.

VI. CONCLUSIONS

In this paper, we have tested two popular domain adaptation algorithms that were proposed to reduce the training time needed to control a prosthesis. We found that the improvements in earlier studies can in fact be attributed to suboptimal hyperparameter optimization, which penalized in particular the NoTransfer reference method. When the hyperparameters are appropriately tuned on the training data of the target subject, the previously reported differences vanish.

This result also holds when correcting for other technical and conceptual mistakes in the original experimental framework. The accuracy of the classification methods in our

updated setting was evaluated with respect to the number of repetitions of each movement, which represents the real effort for the user during the training phase, and for all subjects in the NinaPro database. Also in this case, we do not observe any differences between the HTL algorithms and the NoTransfer baseline.

Intuitively, it should be possible to improve performance on a specific task by using prior information from related tasks. Our findings show, however, that in the context of prosthetic control MultiKT and MKAL, which transfer just source hypotheses rather than source data, do not lead to improved performance. In future work, we will therefore continue to investigate how to successfully leverage over prior information to reduce the training effort for an amputee. Among the directions we consider are unsupervised domain adaptation via distribution alignment [23, 24] and subject invariant data representations using deep learning methods.

REFERENCES

- [1] D. J. Atkins, D. C. Y. Heard, and W. H. Donovan, "Epidemiologic overview of individuals with upper-limb loss and their reported research priorities," *Journal Of Prosthetics And Orthotics*, vol. 8, no. 1, pp. 2–11, 1996.
- [2] C. Castellini, A. E. Fiorilla, and G. Sandini, "Multi-subject / daily-life activity EMG-based control of mechanical hands," *Journal of Neuroengineering and Rehabilitation*, vol. 6, no. 41, 2009.
- [3] F. Orabona, C. Castellini, B. Caputo, A. E. Fiorilla, and G. Sandini, "Model adaptation with least-square svm for adaptive hand prosthetics," in *IEEE International conference on Robotics and Automation*, 2009.
- [4] T. Tommasi, F. Orabona, C. Castellini, and B. Caputo, "Improving control of dexterous hand prostheses using adaptive learning," *IEEE Transactions on Robotics*, vol. 29, no. 1, pp. 207–219, 2013.
- [5] N. Patricia, T. Tommasi, and B. Caputo, "Multi-source adaptive learning for fast control of prosthetics hand," in *2014 IEEE Conference on Computer Vision and Pattern Recognition CVPR*, 6 2014, pp. 2769–2774.
- [6] D. Farina, C. Cescon, and R. Merletti, "Influence of anatomical, physical, and detection-system parameters on surface EMG," *Biological Cybernetics*, vol. 86, no. 6, pp. 445–456, 2002.
- [7] D. Graupe and W. K. Cline, "Functional separation of EMG signals via ARMA identification methods for prosthesis control purposes," *IEEE Transactions on Systems, Man and Cybernetics*, no. 2, pp. 252–259, 1975.
- [8] T. Matsubara, S. Hyon, and J. Morimoto, "Learning and adaptation of a stylistic myoelectric interface: EMG-based robotic control with individual user differences," in *ROBIO*. IEEE, 2011, pp. 390–395.
- [9] J. W. Sensinger, B. A. Lock, and T. A. Kuiken, "Adaptive pattern recognition of myoelectric signals: Exploration of conceptual framework and practical algorithms," *IEEE Transactions on Neural systems and rehabilitation engineering*, vol. 17, no. 3, 2009.
- [10] R. Chattopadhyay, N. Krishnan, and S. Panchanathan, *Topology preserving domain adaptation for addressing subject based variability in SEMG signal*, 2011, vol. SS-11-04, pp. 4–9.
- [11] R. N. Khushaba, "Correlation analysis of electromyogram signals for multiuser myoelectric interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 4, pp. 745–755, 2014.
- [12] J. Liu, X. Sheng, D. Zhang, N. Jiang, and X. Zhu, "Towards zero retraining for myoelectric control based on common model component analysis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, no. 4, pp. 444–454, 2016.
- [13] J. Liu, X. Sheng, D. Zhang, J. He, and X. Zhu, "Reduced daily recalibration of myoelectric prosthesis classifiers based on domain adaptation," *IEEE journal of biomedical and health informatics*, vol. 20, no. 1, pp. 166–176, 2016.
- [14] M. Atzori, A. Gijsberts, C. Castellini, B. Caputo, A.-G. M. Hager, S. Elsig, G. Giatsidis, F. Bassetto, and H. Müller, "Electromyography data for non-invasive naturally-controlled robotic hand prostheses," *Scientific data*, vol. 1, 2014.
- [15] J. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle, *Least Squares Support Vector Machines*. World Scientific, 2002.
- [16] T. Tommasi, F. Orabona, and B. Caputo, "Learning categories from few examples with multi model knowledge transfer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, pp. 928–941, 2014.
- [17] F. Orabona, J. Luo, and B. Caputo, "Online-batch strongly convex multi kernel learning," in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 787–794.
- [18] —, "Multi kernel learning with online-batch optimization," *Journal of Machine Learning Research*, vol. 13, pp. 227–253, 2012.
- [19] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 7, pp. 848–854, 2003.
- [20] I. Kuzborskij, A. Gijsberts, and B. Caputo, "On the challenge of classifying 52 hand movements from surface electromyography," in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2012, pp. 4931–4937.
- [21] I. Kuzborskij and F. Orabona, "Fast rates by transferring from auxiliary hypotheses," *Machine Learning*, pp. 1–25, 2016.
- [22] A. Gijsberts, M. Atzori, C. Castellini, H. Müller, and B. Caputo, "Movement error rate for evaluation of machine learning methods for sEMG-based hand movement classification," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 4, pp. 735–744, 2014.
- [23] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, Feb 2011.
- [24] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *2013 IEEE International Conference on Computer Vision*, Dec 2013, pp. 2960–2967.