Multi-Source Adaptive Learning for Fast Control of Prosthetics Hand

Novi Patricia*, Tatiana Tommasi† and Barbara Caputo‡
*Idiap Research Institute, 1920 Martigny, Switzerland
Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland
†K.U. Leuven, ESAT-PSI IMinds, 3001 Leuven, Belgium
‡University of Rome La Sapienza, 00185 Rome, Italy

Abstract—**We present a benchmark of several existing multi-source adaptive methods on the largest publicly available database of surface electromyography signals for polyarticulated self-powered hand prostheses. By exploiting the information collected over numerous subjects, these methods allow to reduce significantly the training time needed by any new prosthesis user. Our findings provide the biorobotics community with a deeper understanding of adaptive learning solutions for user-machine control and pave the way for further improvements in hand-prosthetics.**

I. INTRODUCTION

One of the main goals of the biorobotics community is to develop hardware and software tools for providing amputees with dexterous, easy to control prosthetic hands. While today's hardware of robotics hands has reached impressive levels, control over a satisfactory range of hand postures and force is still coarse. Moreover, the training process needed by a user to alleviate the inconsistencies between the desired and performed movements can take up to several days and it is generally perceived as very tiring, sometimes painful. As a consequence, often the amputees give up, and settle eventually for a cosmetic hand. This issue calls for machine learning techniques able to boost the learning process of each user. Adaptive methods are suitable for this task [1]–[3]: they allow to leverage the experience gained over numerous source subjects to reduce the training time of a new target user. In this way the learning process does not start every time from scratch, but it reduces to a faster refinement of prior knowledge.

Several researchers have already recognized the importance of adaptive approaches for the control of prosthetic hands (we refer to section II for a review of previous work). Still, methods proposed so far have been evaluated in different settings and on different data. This makes it difficult to compare fairly these techniques and understand which is the most promising solution to accelerate the training process. The NINAPRO dataset is the best existing testbed for this purpose: the surface electromyography (sEMG) signals have been acquired from 27 subjects (w.r.t. a maximum of 10 subjects in other collections [4]) performing 52 among finger, hand and wrist movements. This allows a thorough analysis of the cross-subject information transfer.

The main contribution of this paper is the first benchmark evaluation among adaptive learning algorithms presented so far in the literature. We consider methods that have already been tested for hand posture classification, as well as techniques originally presented in the visual learning domain for object categorization. We evaluate the performance of such methods in terms of recognition rate, over an increasing number of subjects and hand posture classes. The rest of the paper is organized as follows: after a review of relevant literature (section II), section III introduces the notation and presents the methods that are compared in our benchmark. In section IV we describe the experimental setup and present the obtained results. Finally section V concludes the paper and indicates possible directions for future research.

II. RELATED WORK

sEMG measures the signals conveyed as motor commands from the brain to the muscles through non invasive electrodes on the skin surface. One general issue pointed out by previous work is the time- and user-dependent nature of sEMG signals [1], [5]. The first is mainly due to fatigue or electrode displacement, while causes of the second are personal quantity of sub-cutaneous fat, skin impedance and differences in muscle synergies. Variations among the probability distribution of sEMG signals across different subjects make the experience gained on one person not naïvely re-usable [6].

Adaptive learning methods focus on transferring information between a source and a target domain despite the existence of a distribution mismatch among them [7], [8]. This fits perfectly on the problem of prosthetics hands control. Consider the ideal case where an amputee wears his new prosthetic hand for the first time and becomes proficient in using it after only few basic exercises. This would dramatically reduce the number of cumbersome training sessions and make the user much more comfortable. To reach this goal, the prosthetic hand should be endowed with an adaptive system which is already informed on the possible basic hand movements and refines this source knowledge through few signals collected from the specific target user. In [1] the authors suggest to extract from the sEMG data a user-independent component that can be transferred across subjects. The source and target data coming from different persons can also be combined together after re-weighting as proposed in [2]. In [3] the transfer process is formulated as a max-margin learning method and relies on pre-trained models. All these algorithms have been tested on proprietary data of limited dimension, with respect to the number of subjects and the number of hand postures considered. Thus, it is not clear how their performance compare against each other, nor how they would perform on the more realistic scenario of larger numbers of subjects and postures.

Many more adaptive techniques have been developed in machine learning for natural language processing [9], sentiment analysis [10] and computer vision [11]. In particular, in the last research area, the state of the art is obtained in different settings by two methods respectively presented in [12] and [13]: they both define new feature representations which allow to share information easily across source and target. To the best of our knowledge, none of these methods, nor similar ones, have been tested before for prosthetic hands control.

III. ADAPTIVE LEARNING METHODS

All the adaptive learning strategies focus on identifying which part of the source knowledge can be leveraged for the target task at hand, and how to formulate algorithms able to exploit this information. One possible solution is to rely directly on the *source data*, combining them with the few available target labeled samples after reweighting or subselection. Alternatively the source knowledge can be already formalized as a set of *models*, or specific model parameters, that are then used as starting point for the target learning stage. Finally a third solution is to look for a new *feature space* where the source and target data appear similar despite the domain-shift. The definition of such a domain-invariant representation can be obtained either through dimensionality reduction, by identifying a low-dimensional subspace shared by the two domains, or conversely by enlarging the original space in such a way that the new dimensions are able to capture the domain similarities.

In our analysis we consider four different adaptive techniques, and we briefly review them in this section. In the following we indicate with $\{\boldsymbol{x}_i, y_i\}_{i=1}^N$ a set of N labeled samples. Here $\boldsymbol{x}_i \in \mathbb{R}^d$ is a vector of sEMG data and $y_i \in \mathcal{Y} : \{1, \ldots, M\}$ refers to the hand movement performed while measuring the myoelectric signals, where $M > 2$ is the total number of considered postures. We suppose to have S auxiliary source subjects with plenty (N_s) of labeled samples, and one target subject for which only a limited number N_t of labeled samples is available as training set, and the test set has N_{tu} unlabeled data.

Source Data Transfer: Two-Stage Weighting Framework. The method introduced in [14] defines a two-stage weighting procedure to rectify the distribution difference between the source and the target data. In the first stage each source subject is considered separately and compared with the target subject. The source samples are weighted by the factor $\boldsymbol{\alpha}^s$ which is chosen by minimizing the Maximum Mean Discrepancy through the following optimization problem [14]:

$$\min_{\boldsymbol{\alpha}^s: \, \alpha_i^s \geq 0} \left\| \frac{1}{N_s} \sum_{i=1}^{N_s} \alpha_i^s \phi(\boldsymbol{x}_i^s) - \frac{1}{N_t} \sum_{i=1}^{N_t} \phi(\boldsymbol{x}_i^t) \right\|^2 , \quad (1)$$

where $\phi(\boldsymbol{x})$ is a feature map onto a reproducing Kernel Hilbert space. In the second stage all the source subjects are considered together and a weight β^s is assigned to each of them. Let's indicate with $H_i^S = [h_i^1, \ldots, h_i^S]$ the vector of labels predicted by the S sources for the $i-th$ sample of the target domain data. The method aims at minimizing the difference in classification

over two target nearby points as follows [14]

$$\min_{\boldsymbol{\beta}: \, \boldsymbol{\beta}^\top e=1, \, \beta^s \geq 0} \sum_{i,j=1}^{N_{tu}} (H_i^S \boldsymbol{\beta} - H_j^S \boldsymbol{\beta})^2 W_{ij} , \quad (2)$$

where the vector $\boldsymbol{\beta} = [\beta^1, \ldots, \beta^S]^\top$ combines linearly all the source predictions, while the matrix $\mathbf{W}$ contains in each position ij a measure of the similarity between the two target domain samples. Finally a Support Vector Machine (SVM) classifier is learned on the combination of re-weighted source data and the available target training samples. A further parameter μ is used to balance the contribution of the two domains, but its value is either chosen at the beginning or estimated empirically a posteriori from the results without a real optimization procedure. We indicate this approach with *2SW-MDA* as originally named in [14] (see Figure 1, top left frame).

Source Model Transfer: Multi-Adapt. In [3] the authors propose an adaptive method that relies on a linear combination of source models. The idea is to build an SVM classifier over each source subject data defining the vector $\boldsymbol{w}^s$, and to tackle the target problem by solving [3]

$$\min_{\boldsymbol{w},b} \frac{1}{2} \left\| \boldsymbol{w} - \sum_{s=1}^S \beta^s \boldsymbol{w}^s \right\|^2 + \frac{C}{2} \sum_{i=1}^{N_t} \xi_i^2$$
$$\text{subject to} \quad y_i = \boldsymbol{w}^\top \phi(\boldsymbol{x}_i) + b + \xi_i . \quad (3)$$

This runs for each class (e.g. hand posture), in a one-vs-all setting, and the optimal weight vector $\boldsymbol{\beta}$ is obtained by minimizing a convex function over the leave-one-out error [15] (see Figure 1, top right frame). This approach was also extended by substituting $\boldsymbol{\beta}$ with the matrix $\mathbf{B} \in \mathbb{R}^{S \times M}$ which contains different weights for each source and each class [3]. This allows to evaluate in a more precise way the relation among the domains, while all the sources collaborate at once to improve the target performance. In the following we will refer to this version of the method with the name *Multi-Adapt*.

Feature Representation and Source Data Transfer: Geodesic Flow Kernel. Two domains can be considered as two points on a low dimensional manifold [12]. Formally, a dimensionality reduction step is applied on the data through PCA and the function $\boldsymbol{\Phi}(t)$ is used to indicate the geodesic flow path between the source ($\boldsymbol{\Phi}(0)$) and the target ($\boldsymbol{\Phi}(1)$) in the reduced space. Any original feature vector $\boldsymbol{x}$ is projected onto the geodesic flow by $\boldsymbol{\Phi}(t)^\top \boldsymbol{x}$ and the inner product between two vectors is obtained by the positive definite *Geodesic Flow Kernel* (GFK) [12]:

$$G(\boldsymbol{x}_i, \boldsymbol{x}_j) = \int_0^1 (\boldsymbol{\Phi}(t)^\top \boldsymbol{x}_i)^\top (\boldsymbol{\Phi}(t)^\top \boldsymbol{x}_j) \, dt = \boldsymbol{x}_i^\top \mathbf{G} \boldsymbol{x}_j , \quad (4)$$

where the matrix $\mathbf{G}$ can be computed in closed form. If the similarity of two samples across domains is evaluated in this way, a classifier (e.g. linear SVM) learned on the source is expected to perform well on the target despite the original domain shift. A schematic representation of this approach is given in Figure 1, bottom left frame.

Feature Representation Transfer: Multi-Kernel Adaptive Learning. Similarly to Multi-Adapt, the approach presented in [13] proposes to leverage over the source models, but instead
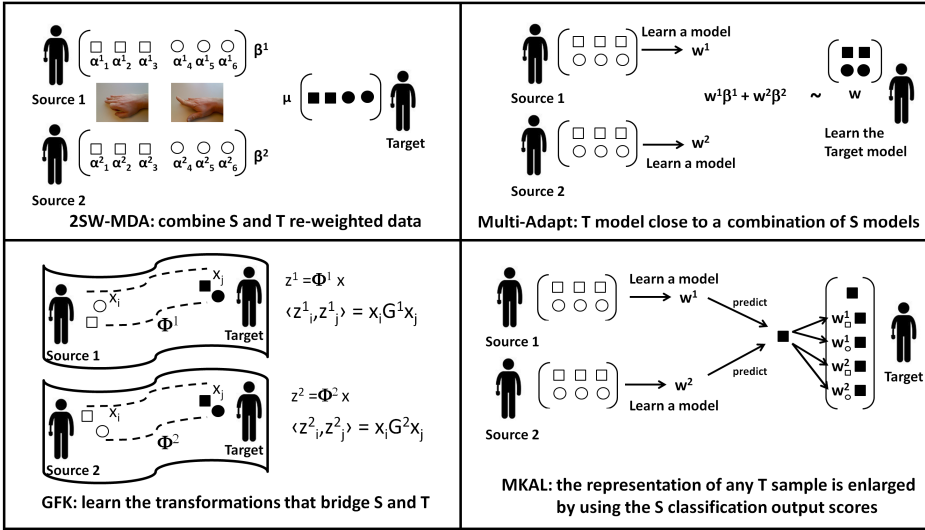
Fig. 1: A scheme of the four considered adaptive methods, when two subjects are used as sources and perform two finger movements. In 2SW-MDA all the data of each source are weighted and combined with the target subject samples. In Multi-Adapt, a model is learned from each source and used as reference when learning on the target. For GFK the source and target data are embedded in a low dimensional manifold and the geodesic flow is used to reduce the domain shift when evaluating the cross-domain sample similarity. In MKAL each source predicts on the target samples and the scores are used as extra features.

of using them directly, the output of their classification is considered as a descriptor, defining a feature transfer method. The score values obtained from different sources on each target sample are concatenated to the original descriptor. Thus, the learning process on this new representation will implicitly and automatically choose how to weight the source knowledge. The *Multi-Kernel Adaptive Learning* (MKAL) method (schematized in Figure 1, bottom right frame) solves the following optimization problem [13]

$$\min_{\bar{\boldsymbol{w}}} \frac{1}{2}\|\bar{\boldsymbol{w}}\|_{2,p}^2 + C\sum_{i=1}^{N_t} \ell(\bar{\boldsymbol{w}}, \boldsymbol{x}_i, y_i). \qquad (5)$$

Here $\bar{\boldsymbol{w}} = \left[\boldsymbol{w}^{(0)}, \boldsymbol{w}^{(1)}, \dots, \boldsymbol{w}^{(V)}\right]$ where $\boldsymbol{w}^{(0)}$ deals with the original sample descriptor $\boldsymbol{x}$, while $v = \{1, \dots, V\}$ runs over the $(source, class)$ classification score pairs, with $V = S \times M$. The optimization problem considers a multiclass loss ℓ [16] and it is regularized through the $(2, p)$ group norm of $\bar{\boldsymbol{w}}$ [17]; tuning the p parameter allows to choose the sparsity level over the source knowledge. This approach is indicated as *Multi-Kernel Adaptive Learning* (MKAL) and it is schematized in Figure 1, bottom right frame.

By comparing MKAL with GFK we can identify two main conceptual differences. The first is that the definition of a new feature representation is obtained by augmentation for MKAL and dimensionality reduction for GFK. The second is that for MKAL it is not necessary to keep the source samples when solving the target problem: the source classifiers are considered as experts which provide a confidence output on the target data. This confidence is the only value necessary to run MKAL independently from the original source data.

IV. EXPERIMENTS

In this section we present our experimental setup and we describe the obtained results (sections IV-A, IV-B).

Experimental Setup. We ran our experiments on the NI-NAPRO database, where the signals are measured with 10 active double-differential OttoBock MyoBock 13E200 surface EMG electrodes, 8 placed beneath the elbow and 2 on the flexor and extensor muscles. More details about the data

collection can be found in [4], [18]. Following [18] each sEMG sample is described by a feature vector of $d = 10$ elements, corresponding to the mean absolute value of the sEMG signal measured by each electrode. During the data acquisition, the subjects were instructed to repeat each movement ten times, alternated with an intermediate rest posture. We ignore the sample time order by simply shuffling the data and selecting a maximum of 1080 samples as training, and about 110000 samples for testing.

We define two experimental settings. In the first we reproduce the standard *small setup* presented in previous papers by randomly selecting 10 classes (9 movements and the rest position) and 10 subjects. In the second we *fully exploit the* NINAPRO *database* considering all the available classes and subjects. For both settings each subject acts in turn as the target while all the others are used as sources. Moreover we organized the classes in four different groups

Exercise 1: finger movements;
Exercise 2: hand postures and wrist movements;
Exercise 3: grasping and functional movements;
Mix: the combination of Exercise 1, 2 and 3.

We run the experiments separately on each exercise and we analyze their recognition rate performance when the number of available target training samples increases. The comparison among the exercises results provides an insight on the difficulty of sEMG automatic signal discrimination for each group of movements and sheds light on how much adaptive learning is beneficial for each specific sub-problem. We benchmark the performance of the approaches presented in the previous section. For all the considered methods we used the code provided by the authors, specific details on their implementation are given below.

2SW-MDA. Solving (1) and (2) is highly computational expensive and it becomes infeasible with more than 10^4 samples[1]. For this reason we run the 2SW-MDA method over a sub-part of the described small setting, considering only 10

---

[1]This number indicates the combination of source and target samples and derives from preliminary experiments.

subjects and a maximum of 600 samples per each of them. We use $\mu = 1$ and linear SVM as proposed in [14] .

GFK. The first step of this method is a PCA projection which reduces the dimensionality of the original feature vectors. For our sEMG signals the descriptors have few (only 10) elements, thus we considered as final dimensionality $d/2 = 5$: this is the minimal reduction suggested in [12] and induces a loss of 5% in the descriptor total variance. Moreover this approach does not manage the case of multiple source sets, hence the target can rely only on each source subject separately and we consider the best linear SVM results obtained in this way on the test set.

Multi-Adapt and **MKAL** need respectively the pre-defined source models and their prediction output on the target. For both methods we follow the strategy described in [3], [13] by training a non-linear SVM model on the samples of each subject separately. The parameter p of MKAL is fixed to $\frac{2 \log(V+1)}{2 \log(V+1)-1}$ according to the automatic setting in [13].

For all the methods the SVM parameter C was chosen in $\{0.01, 0.1, 1, 10, 100, 1000\}$ by cross validation over the sources. The best parameter value was then kept for the target experiments.

We consider as reference the following two baselines:

No-Transfer. This corresponds to standard supervised learning applied on the target task. We use non-linear SVM and we show the best performance obtained by tuning C: these results indicate the upper bound on learning from scratch without exploiting the source knowledge. If the transfer process works properly, we expect any adaptive learning method to perform better or at least equally to this approach.

Prior-Features. The source predictions used as input for MKAL can be concatenated together and used as a new target samples descriptor. We run a linear SVM classifier on them.

In all the experiments, for the non-linear SVM classifier, we used the Gaussian kernel $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = -\exp(\gamma^{-1}||\boldsymbol{x}_i - \boldsymbol{x}_j||^2)$ with γ being fixed to the mean of the pairwise distances among the involved samples. All the methods that rely on a linear SVM have an initial non-linear stage that makes fair the comparison: for 2SW-MDA, $\phi(\boldsymbol{x})$ in (1) is a feature map based on the Gaussian kernel, while the Geodesic Flow Kernel in (4) is non-linear by definition.

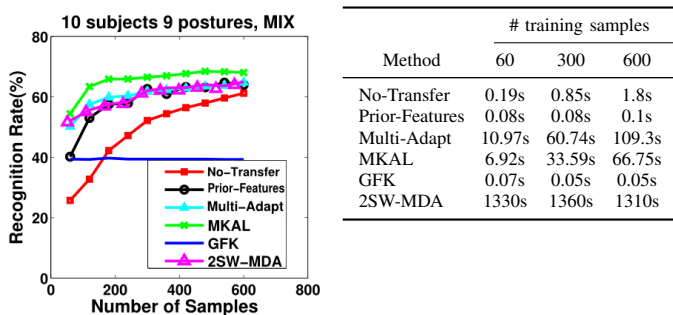| Method	# training samples		
	60	300	600
No-Transfer	0.19s	0.85s	1.8s
Prior-Features	0.08s	0.08s	0.1s
Multi-Adapt	10.97s	60.74s	109.3s
MKAL	6.92s	33.59s	66.75s
GFK	0.07s	0.05s	0.05s
2SW-MDA	1330s	1360s	1310s

Fig. 3: Analysis on the 2SW-MDA method. Left: classification results on the Mix exercise. Right: running time for 60, 300 and 600 target training samples.

A. Small setup results: 10 subjects, 10 classes

The results obtained with the first experimental setting are shown in Figure 2, top line. By comparing the recognition rate of No-Transfer over the four plots we can state that *Exercise 1* appears the most difficult classification tasks i.e. the one with the worst results for equal number of training samples. On the other extreme, *Exercise 3* corresponds instead to the easiest task. Most probably different functional movements induce a larger variety in the sEMG signals w.r.t. the finger movements. Moreover, *Exercise 1* is the case with the highest recognition rate advantage obtained with adaptive learning methods: from 15% for No-Transfer to 57% for MKAL with only 60 target training samples, and 71% (MKAL) against 61% (No-Transfer) at the last training step (1080 samples). This indicates that the information on finger movements is also the easiest to be shared across subjects.

Overall MKAL shows the best performance: its advantage over the other adaptive learning methods is evident in *Exercise 1* and *Mix*. For *Exercise 2* MKAL results are statistically equivalent to that of Prior-Features, but still better than Multi-Adapt. Finally in *Exercise 3*, MKAL, Prior-Features and Multi-Adapt shows analogous performance.

The GFK method presents instead the worst recognition results. Although it performs better or equally than No-Transfer in the very first training steps, the recognition rate does not increase with the number of available training samples. To properly analyze this behaviour it is important to remember that the geodesic flow kernel allows to evaluate the similarity between samples of the source and target data as if they belonged to the same domain. Thus all the source data are kept in training together with the new target labeled samples. As a consequence the training set is already rich in data at the very first step of our plots and, in the new dimensionality reduced space, adding new samples does not bring any extra information. A possible explanation might be in the use of PCA. While this choice has proved effective in the visual domain, it does not seem to be adequate over sEMG signals, which might require higher order approximations for being modeled correctly and handle properly the noise level in the signal. Moreover, GFK deals with a single source subject at a time, while the other adaptive methods rely on a combination of all the available sources. For a fair comparison, the bottom line of Figure 2 shows the result obtained when each adaptive method relies only on the best source chosen a posteriori. The general trend in the results does not change, but it can be noticed that Prior-Features and Multi-Adapt show a decrease in performance in passing from many to one source. This is not the case for MKAL: this behavior suggests that MKAL is able to identify autonomously the most useful source among the others.

Till here our discussion did not include the 2SW-MDA method. We repeated the experiment on the Mix exercise with a maximum of 600 target training samples and we show the results obtained by all adaptive methods in Figure 3. From the plot, it is clear that 2SW-MDA performs almost as Multi-Adapt, however the two methods differ in the learning procedure. 2SW-MDA needs that all the source samples are stored to allow a comparison with the target data, while Multi-Adapt relies only on pre-trained models which usually request a much smaller memory cost than the data themselves. The
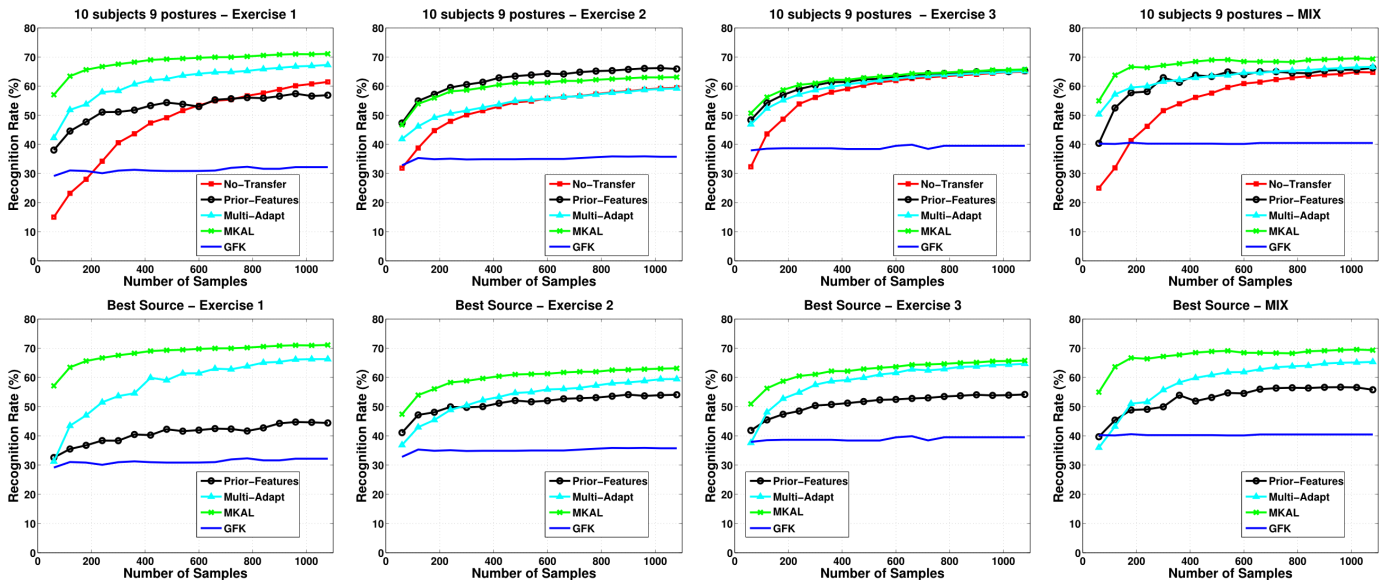
Fig. 2: Average results over 10 subjects, each considered in turn as target. The classification tasks involve 9 hand postures plus rest for a total of 10 classes. The title of the plots indicates the considered exercise. Top line: Prior-Features, Multi-Adapt and MKAL rely on a combination of all the source subjects, while for GFK we show the performance of the best source. Bottom line: all the adaptive methods rely only on the most relevant source subject chosen a posteriori from the results.

table on the right in Figure 3 allows a general evaluation of the computational load for the different adaptive learning methods. The 2SW-MDA approach presents a running (training plus test) time which is 10^1-10^3 times longer than what needed by all the other approaches. Overall we can state that MKAL presents the best trade-off between computational load and recognition results among all the considered adaptive methods.

B. Full NINAPRO dataset results

Passing from the small setting to the full NINAPRO dataset corresponds to increase both the number of classes and the number of available sources: this setup is at the same time the more realistic and the more challenging. Figure 4 shows the performance of all the adaptive learning methods over the three separate exercises and their combination which now covers all the 52 available hand movements, plus the rest position. The global decreasing trend in recognition rate when passing from *Exercise 1* to *Exercise 3* is due to the growing number of classes in the different experiments. Here MKAL shows the best performance, always followed in the same order by Multi-Adapt and Prior-Features. For the *Mix* experiment MKAL and Prior-Features appear equivalent and outperform Multi-Adapt with a major advantage when more than 300 target training samples are available. The main difference between Prior-Features and MKAL is in the fine choice of the weights assigned to each source subject and class knowledge. Our results indicate that, when a very high number of sources and classes are available, the weight tuning is not really needed.

GFK presents the same behavior already discussed before for *Exercise 1-3* with recognition rate better than No-Transfer only in the first four steps of the plot, from 60 to 240 target training samples. For *Mix*, GFK has a constant performance, always equivalent to what obtained by No-Transfer with 60 target training samples.

In this full setting we can also analyze separately the results

on each target subjects instead of considering only the average performance over them. Specifically, Figure 4 shows in the middle and bottom lines respectively the best and worst cases for each experiment. With best/worst we indicate the target subject for which adaptation gives the maximum/minimum advantage with respect to learning from scratch. Moreover, it is interesting to notice that in the *Mix* experiment for the worst case, using Prior-Features is the optimal adaptive strategy for more than 500 training samples. This indicates that in such a challenging setting with a high number of classes it is still possible to obtain a significant gain in classification by transferring even from weakly related sources. In general, all the methods except GFK shows classification results better than, or at least equal to, No-Transfer.

V. SUMMARY AND FUTURE WORK

In this paper we argued for the importance of adaptive learning methods in reducing the training time of amputees when learning to control an sEMG based prosthetic hand. To this end we considered four adaptive learning algorithms, two of which already tested in the prosthetic domain. We compared their performance when dealing with various types and number of hand movements, increasing also the set of available source subjects. Overall our analysis indicates that adaptive learning has a strong potential in this field: given a fixed recognition result they allow for a reduction of one order of magnitude in the number of training samples needed with respect to learning from scratch. However, the mere plug and play of algorithms developed in other reasearch domains may be not enough. This calls for the definition of new adaptive approaches addressing the specific nature of the hand prosthetic problem, namely the small feature dimensionality derived from the sEMG signals and a high number of sources which are conditions not generally expected in other fields such as visual recognition. Feature transfer methods that increase the descriptor dimensionality
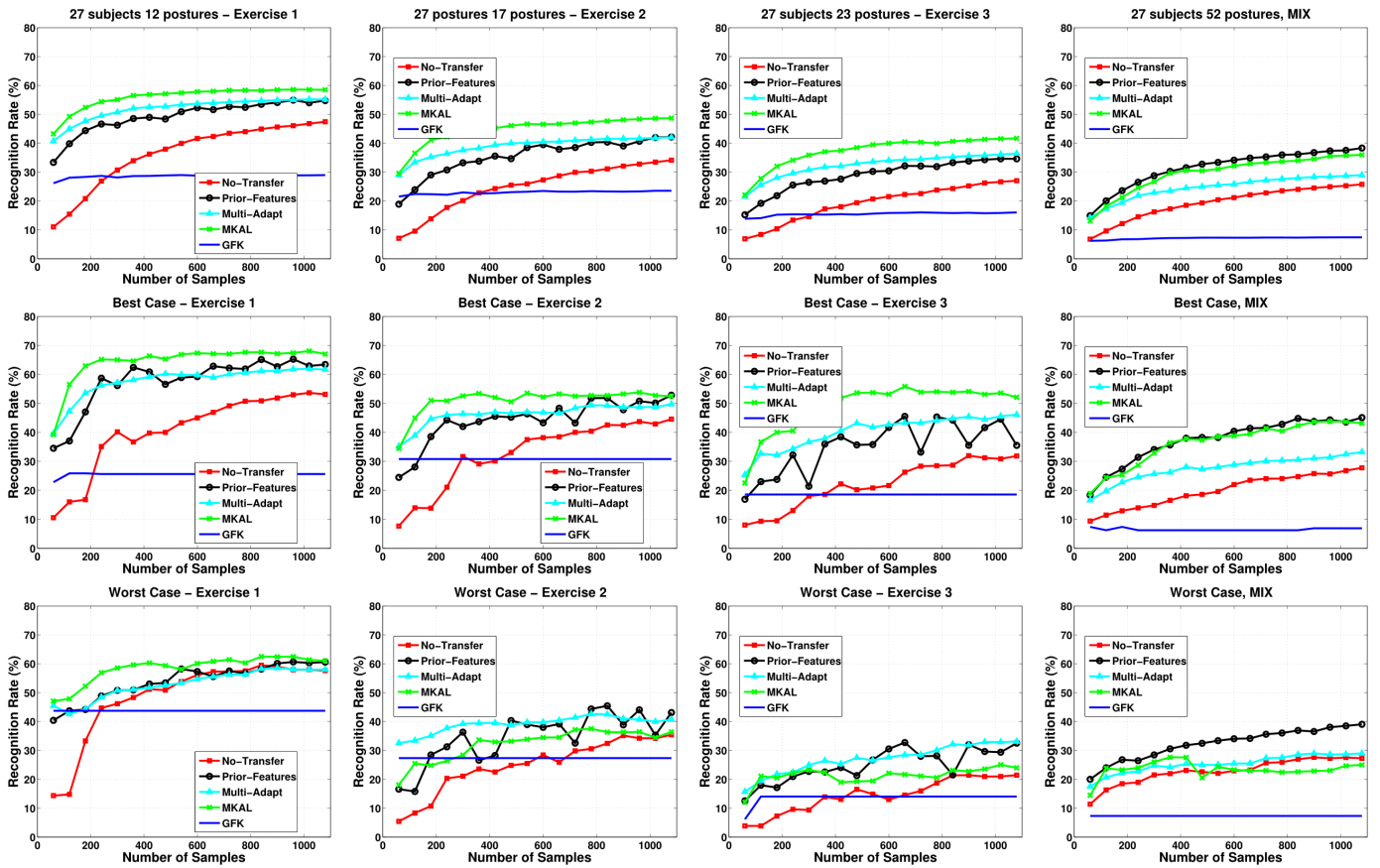
Fig. 4: Average results over 27 subjects (top line), each considered in turn as target. Classification results of the target subject which gets the the highest (middle line) and lowest (bottom line) advantage from adaptive learning. The classification tasks involve a different number of hand postures depending on the exercise indicated in the title of each plot.

and the definition of a hierarchical structure over the sources may be good directions for future work.

ACKNOWLEDGMENT

REFERENCES

[1] T. Matsubara, S.-H. Hyon, and J. Morimoto, "Learning and adaptation of a stylistic myoelectric interface: Emg-based robotic control with individual user differences." in *Proc ROBIO*, 2011.

[2] R. Chattopadhyay, N. C. Krishnan, and S. Panchanathan, "Topology preserving domain adaptation for addressing subject based variability in semg signal," in *AAAI Spring Symposium: Computational Physiology*, 2011.

[3] 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.

[4] M. Atzori, A. Gijsberts, S. Heynen, A.-G. Mittaz Hager, O. Deriaz, P. van der Smagt, C. Castellini, B. Caputo, and H. Müller, "Building the ninapro database: a resource for the biorobotics community," in *Proc. BioRob*, 2012.

[5] J. W. Sensinger, B. A. Lock, and T. A. Kuiken, "Adaptive pattern recognition of myoelectric signals: exploration of conceptual framework and practical algorithms." *IEEE Trans Neural Syst Rehabil Eng*, vol. 17, pp. 270–278, 2009.

[6] 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.

[7] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Machine Learning*, vol. 79, pp. 151–175, 2010.

[8] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.

[9] J. Blitzer, R. McDonald, and F. Pereira, "Domain adaptation with structural correspondence learning," in *Proc EMNLP*, 2006.

[10] J. Blitzer, M. Dredze, and F. Pereira, "Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification," in *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 2007, pp. 440–447.

[11] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *Proc ECCV*, 2010.

[12] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *CVPR*, 2012, pp. 2066–2073.

[13] J. Luo, T. Tommasi, and B. Caputo, "Multiclass transfer learning from unconstrained priors," in *ICCV*, 2011, pp. 1863–1870.

[14] Q. Sun, R. Chattopadhyay, S. Panchanathan, and J. Ye, "A two-stage weighting framework for multi-source domain adaptation," in *NIPS*, 2011, pp. 505–513.

[15] I. Kuzborskij and F. Orabona, "Stability and hypothesis transfer learning," in *Proc. ICML*, 2013.

[16] I. Tsochantaridis, T. Hofmann, T. Joachims, and Y. Altun, "Support vector machine learning for interdependent and structured output spaces," in *ICML*, 2004.

[17] F. Orabona, J. Luo, and B. Caputo, "Online-batch strongly convex multi kernel learning," in *CVPR*, 2010, pp. 787–794.

[18] I. Kuzborskij, A. Gijsberts, and B. Caputo, "On the challenge of classifying 52 hand movements from surface electromyography," in *Proc. EMBC*, 2012.