Transfer Learning of Complex Motor Skills on the Humanoid Robot Affetto

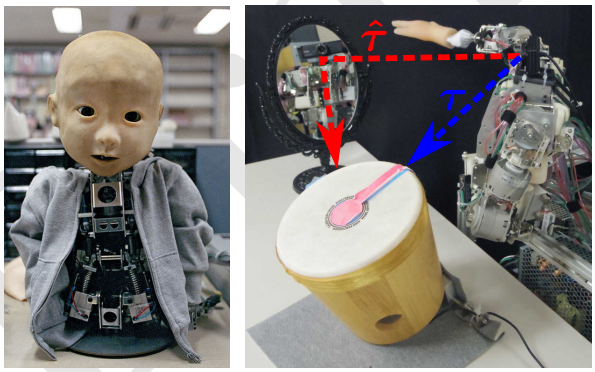
Alexander Schulz*, Jeffrey Frederic Queißer*†, Hisashi Ishihara‡ and Minoru Asada‡
*Center of Excellence Cognitive Interaction Technology (CITEC), Machine Learning Group
†Research Institute for Cognition and Robotics (CoR-Lab), Machine Learning Group
Bielefeld University, 33615 Bielefeld, Germany
Email: [aschulz|jqueisse]@techfak.uni-bielefeld.de
‡Graduate School of Engineering
Osaka University, Suita, Osaka 565-0871, Japan
Email: [ishihara|asada]@ams.eng.osaka-u.ac.jp

***Abstract*— Although autonomous robots can perform particularly well at highly specific tasks, learning each task in isolation is a very costly process, not only in terms of time but also in terms of hardware wearout and energy usage. Hence, robotic systems need to be able to adapt quickly to new situations in order to be useful in everyday tasks. One way to address this issue is transfer learning, which aims at reusing knowledge obtained in one situation, in a new related one. In this contribution, we develop a drumming scenario with the child robot Affetto where the environment changes such that the scene can only be observed through a mirror. In order to address such domain adaptation problems, we propose a novel transfer learning algorithm that aims at mapping data from the new domain in such a way that the original model is applicable again. We demonstrate this method on an artificial data set as well as in the robot setting.**

I. Introduction

Advanced robotic systems face non-static environmental conditions that require context-dependent adaptation of motor skills. Approaches that optimize parameters for a given task, like trajectory estimation [1] or walking gait exploration [2], are only able to deal with static tasks. Although in many cases a low-dimensional parametrization that covers the variance of a task exists. As in our case, consider a drumming task that has to incorporate kinematic constraints of the robot and different target positions. A full optimization for each new task from a reasonable initialization, that was acquired by e.g. kinesthetic teaching, means that many computations and trials need to be performed before the task can be executed.

Previous work successfully achieved throwing of objects at parametrized target positions [5] or playing table tennis using motion primitives that are parameterized with respect to the current ball trajectory [6], [7]. Matsubara [8] focused on parametrized motion primitives by introducing style parameters that determine a linear mixture of policy parameters. The retrieval of parameterized motion primitives by regression techniques was also proposed in [9], [10], [11]. In [12], parameterized motion primitives have been stored within a neural associative memory using non-linear embeddings of policy parameters. While learning in [8], [10], [12] was conducted offline on precollected data sets, e.g. collected from human demonstrations, the work in [5] integrates an active learning criterion in order to select new task parameterizations in unexplored areas of the memory to enhance efficiency of skill memory acquisition. However, the initialization is done as in the previous examples by kinesthetic demonstration of a whole-body overhand throw. But generation of training data for skill memories requires a set of optimized samples, which is costly since each action has to be performed on the robot. Each training sample is based on a full optimization with a fixed initialization or an initialization gathered by demonstration e.g. kinesthetic teaching.



(a) (b)

Fig. 1. Affetto robot, (a) upper body and internal structure as presented in [3], [4]. (b) Experimental setup for transfer learning, the robot has to generate drumming actions based on the perceived reflection of the drum in a mirror.

Besides learning a completely new task, real world situations sometimes require the ability to adapt an already learned task to changing conditions without learning the acquired motion repertoire from scratch. In our case, the humanoid robot Affetto (Fig. 1a) learns to solve a drumming task with varying positions of the drum. Then, the environment changes in a way such that the drum cannot be observed directly and the robot has to perceive its position through a mirror, located at the side of the scenario (Fig. 1b). Other examples for changes in the scenario include the replacement of the original (possibly faulty) sensor by a newer/intact one,

Final Draft

a changed position of the robot which would be otherwise static, or another modified point of view on the scenery. Relearning the complete task in the high-dimensional space of actions would be highly ineffective if instead the already acquired knowledge could be adapted and reused.

The field investigating such principles is called transfer learning [13], [14], where the main goal is to reuse as much as possible of the previous knowledge for the new situation. Recently, a promising transfer learning approach has been proposed for classification in myoelectric prosthesis control under electrode shift [15]. This approach allows to transfer the classification model between two settings, without assuming a continuous drift, by optimizing a mapping of the input features directly for the target task.

In our contribution, we generalize this approach for a regression model and apply it to adapt a previously learned skill of a humanoid robot towards changing task conditions.

The remaining of this paper is structured as follows. After a related work section, which details the important strategies for robot control and depicts the relevant related work on transfer learning, section III introduces the proposed transfer learning algorithm. Section IV illustrates this method for an artificial example while section V describes the main experiment where, first, a drumming task is learned with the humanoid robot Affetto (Fig. 1a) and, then, the proposed transfer learning algorithm is employed to adapt towards a change in the environment.

II. Related Work

A. Parameterized Skills

As in our previously published work [16], parameterization $\boldsymbol{\theta} \in \mathbb{R}^F$ of policies $\pi_{\boldsymbol{\theta}}$ are considered. Additionally, it is assumed that tasks are parameterized by $\boldsymbol{\tau} \in \mathbb{R}^E$. Task instances that are defined by $\boldsymbol{\tau}$ are distributed according to the probability density function $P(\boldsymbol{\tau})$. The variability of the task is represented by the task parameterization $\boldsymbol{\tau}$, e.g target positions, positions of obstacles, or variable loads that are attached to an end-effector. The notion of a parameterized skill is introduced with reference to [5], it is given by a function PS : $\mathbb{R}^E \rightarrow \mathbb{R}^F$ that maps task parameters $\boldsymbol{\tau}$ to policy parameters $\boldsymbol{\theta}$. The aim is to learn a parameterized skill PS($\boldsymbol{\tau}$) that maximizes $\int P(\boldsymbol{\tau}) J(\pi_{\text{PS}(\boldsymbol{\tau})}, \boldsymbol{\tau}) d\boldsymbol{\tau}$ with $J(\pi, \boldsymbol{\tau}) = E\{R(\pi_{\boldsymbol{\theta}}, \boldsymbol{\tau}) | \pi, \boldsymbol{\tau}\}$, the expected reward for using policy $\pi_{\boldsymbol{\theta}}$ to solve a task with parameterization $\boldsymbol{\tau}$. The reward function $R(\pi_{\boldsymbol{\theta}}, \boldsymbol{\tau})$ assesses each action of the robot defined by the policy $\pi_{\boldsymbol{\theta}}$ with respect to the current task parameterization $\boldsymbol{\tau}$. Optimization of PS($\boldsymbol{\tau}$) is usually performed by learning from demonstration, policy optimization or a combination of both.

Definition of $\pi_{\boldsymbol{\theta}}$: For the policy representation $\pi_{\boldsymbol{\theta}}$, that encodes the joint angle trajectories of the drumming action into $\boldsymbol{\theta}$, we refer to Dynamic Movement Primitives (DMPs, [17]). DMPs are widely used in the field of robot motion generation and show good generalization in the parameter space as well as invariances to scaling. The combination of a linear point attractor in combination with a non-linear perturbation results in a stable dynamical system

$$\ddot{q} = k_S(g - q) - k_D\dot{q} + f(x, \boldsymbol{\theta}), \tag{1}$$

that defines velocity and acceleration profiles and thus the output trajectory. Typically, the canonical system of the DMPs is defined as $\dot{x} = -\alpha x$. However, in this work we refer to a linear decay $\dot{x} = -\alpha$ as in [18]. The shape of the primitive is defined by perturbation

$$f(x, \boldsymbol{\theta}) = \frac{\sum_{k=1}^{K} \exp(-\boldsymbol{V}_k(x - \boldsymbol{C}_k))\boldsymbol{\theta}_k}{\sum_{k=1}^{K} \exp(-\boldsymbol{V}_k(x - \boldsymbol{C}_k))}, \tag{2}$$

with a fixed number of K Gaussians that are centered around $\boldsymbol{C}_k$ and the variance of the Gaussians $\boldsymbol{V}_k$. Further, $\boldsymbol{C}_k$ and $\boldsymbol{V}_k$ are assumed to be fixed as in [12]. The remaining parameterization of the DMP are the coefficients $\boldsymbol{\theta}_k$.

Definition of PS($\boldsymbol{\tau}$): An incremental variant of the Extreme Learning Machine (ELM, [19]) is used for implementation of the parameterized skill PS($\boldsymbol{\tau}$). ELMs are feed-forward neural networks with one hidden layer,

$$\boldsymbol{\theta}_i(\boldsymbol{\tau}) = \sum_{j=1}^{H} \mathbf{W}_{ij}^{out} \sigma\left(\sum_{k=1}^{E} \mathbf{W}_{jk}^{inp} \boldsymbol{\tau}_k + \boldsymbol{b}_j\right) \quad \forall i = 1, \ldots, F. \tag{3}$$

E denotes the input dimensionality, H the hidden layer size and F the output dimensionality. Regression is performed on a random projection of the input $\mathbf{W}^{inp} \in \mathbb{R}^{H \times E}$ and a non-linear transformation $\sigma(x) = (1 + e^{-x})^{-1}$. The linear output transformation $\mathbf{W}^{out} \in \mathbb{R}^{F \times H}$ can be updated by incremental least squares algorithms. The incremental update scheme of the ELM was introduced as Online Sequential Extreme Learning Machine (OSELM) [20], [21] that incorporates the ability to perform an additional regularization on the weights [22] or exponential forgetting of previous samples [23]. Since we expect to deal with a small number of training samples, regularization of the network can help to prevent over-fitting and foster reasonable extrapolation.

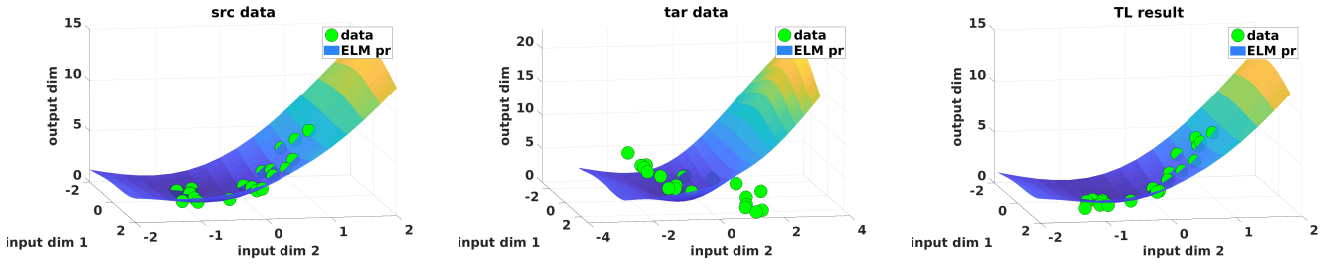
Alternative approaches exist for learning DMPs such as [24]. In this work, however, we focus on the transfer between related tasks and, hence, restrict our learning setup to one DMP learning scheme.

B. Transfer Learning

In the following, we briefly recap the existing literature on transfer learning and describe differences to our proposal.

The literature differentiates between different types of changing conditions [13]: Changes in the task and changes in the data domain. In this work, we consider the latter case where the task to be performed stays the same, while the data domain changes. In particular, the general assumption is that enough data are available in an old scenario, the so called *source domain*, but the goal is to solve the task in the new *target domain*, where only very few data are available. These types of problems are also referred to as *transductive transfer learning* [13] or as *domain adaptation* [25]. More formally, we will refer to data instances from the source domain as $\boldsymbol{\tau} \in \mathcal{T} = \mathbb{R}^E$ and to instances from the target domain as $\hat{\boldsymbol{\tau}} \in \hat{\mathcal{T}} = \mathbb{R}^{\hat{E}}$.

Final Draft



(a) (b) (c)

Fig. 2. Illustration of the proposed transfer learning approach on a toy data: the figures always show data (green circles) and the predictive function of the ELM (trained on the source data). (a) source data; (b) target data; (c) target data after transfer learning.

A popular set of methods in this area are related to the concept of importance sampling, one example being the kernel mean matching algorithm [26]. Here, weights are introduced for the data points in the source space and then utilized for learning a new supervised model to improve its performance in the target space. A central assumption here is that the conditional distributions in both data spaces are the same: $p_{\hat{\mathcal{T}}}(\boldsymbol{\theta}|\boldsymbol{\tau}) = p_{\mathcal{T}}(\boldsymbol{\theta}|\boldsymbol{\tau})$ [13]. This strong assumption, however, does not hold in our scenario where the input space is changed strongly and thus the conditional distribution changes as well.

Another set of transfer learning methods aims to solve the transfer problem by finding a common latent space for the source and target domain [13], [27]. However, these methods assume the availability of only unlabeled data in the target space and, thus, do not make use of any supervised information if existing. Procrustes Analysis [28] requires correspondence information between some samples from both domains which is unavailable in our setting. The work [29] utilizes labels but assumes alignment of the features in source and target space which is unavailable in our setting.

Transfer learning has been applied in robotic settings, like reinforcement learning [30], [31], [32], multi-robot transfer learning [33], [34], (learning a skill for a robot from another robot). A further application is inter-task learning, e.g. transfer knowledge of multiple acquired tasks to solve more complex new tasks [35]. Those settings are, however, different from ours because we consider only changes in the input but not in the output for learning based on kinesthetic teaching to adapt for changing task configurations.

III. Transfer Learning for Non-Linear Regression with the ELM

For formalizing transfer learning, we follow the main idea from [15], [36], which is to learn a mapping that transforms the novel target data in such a way, that the original model is applicable again. In contrast to [15], [36], we implement this idea for a regression model and evaluate it in a robotic scenario.

While in principle this technique is applicable to any supervised machine learning model with a differentiable cost function, we demonstrate it here for the regression model ELM. Given a training data set $\mathcal{D} = \{(\boldsymbol{\tau}^j, \boldsymbol{\theta}^j)|j = 1, \ldots, N\}$ in the source domain, the ELM optimizes the cost

$$\sum_{j=1}^{N}\sum_{i=1}^{F}\left(\boldsymbol{\theta}_i^j - \boldsymbol{\theta}_i(\boldsymbol{\tau}^j)\right)^2 \tag{4}$$

with respect to the parameters $\mathbf{W}^{out}$, where $\boldsymbol{\theta}_i(.)$ is defined in equation (3). This results in a learned function PS($\boldsymbol{\tau}$), applicable to instances from the source domain $\boldsymbol{\tau}$.

For our transfer learning approach, we utilize the same cost function but this time, it takes instances from the target domain as input $\hat{\mathcal{D}} = \{(\hat{\boldsymbol{\tau}}^j, \boldsymbol{\theta}^j)|j = 1, \ldots, \hat{N}\}$, where $\hat{N} \ll N$. Furthermore, we introduce a transfer mapping $h(\hat{\boldsymbol{\tau}})$ which is applied to the input $\hat{\boldsymbol{\tau}}$. Thereby, $h(.)$ realizes a mapping from the target to the source domain and learning its parameters comprises the main part of the transfer learning step. In many application, it is reasonable to assume a linear transformation of the form $h(\hat{\boldsymbol{\tau}}) = \boldsymbol{H}\hat{\boldsymbol{\tau}} + \boldsymbol{c}$, where $\boldsymbol{H} \in \mathbb{R}^{E \times \hat{E}}$ and $\boldsymbol{c} \in \mathbb{R}^E$. The transfer learning problem finally is

$$\min_{\boldsymbol{H},\boldsymbol{c}} \sum_{j=1}^{\hat{N}}\sum_{i=1}^{F}\left(\boldsymbol{\theta}_i^j - \boldsymbol{\theta}_i(h(\hat{\boldsymbol{\tau}}^j))\right)^2 + \lambda\|\tilde{\boldsymbol{H}}\|^2. \tag{5}$$

Thereby, $\tilde{\boldsymbol{H}}$ constitutes the matrix $\boldsymbol{H}$ augmented by an additional column containing the values of $\boldsymbol{c}$ and λ is a weighting for the L2 regularization.

Then, finding a minimum of this problem with respect to the parameters of $h(.)$ constitutes the transfer learning step and we employ the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [37] for optimization.

IV. Experiments I: A Toy Data Example

We demonstrate our proposed transfer learning scheme for a toy data set first, before we apply it to a robotic scenario in the next section.

We sample 20 data points from the function

$$\mathbb{R}^2 \mapsto \mathbb{R} : \boldsymbol{x} \mapsto (x_1 + 1)^3 + (2(x_2 + 1)^3)/10, \tag{6}$$

where we utilize 14 randomly selected points for training a ELM. The mean squared error (MSE) is 0.00 on the training and 0.06 on the remaining testing data. The trained model together with the data is shown in Fig. 2a.

Draft

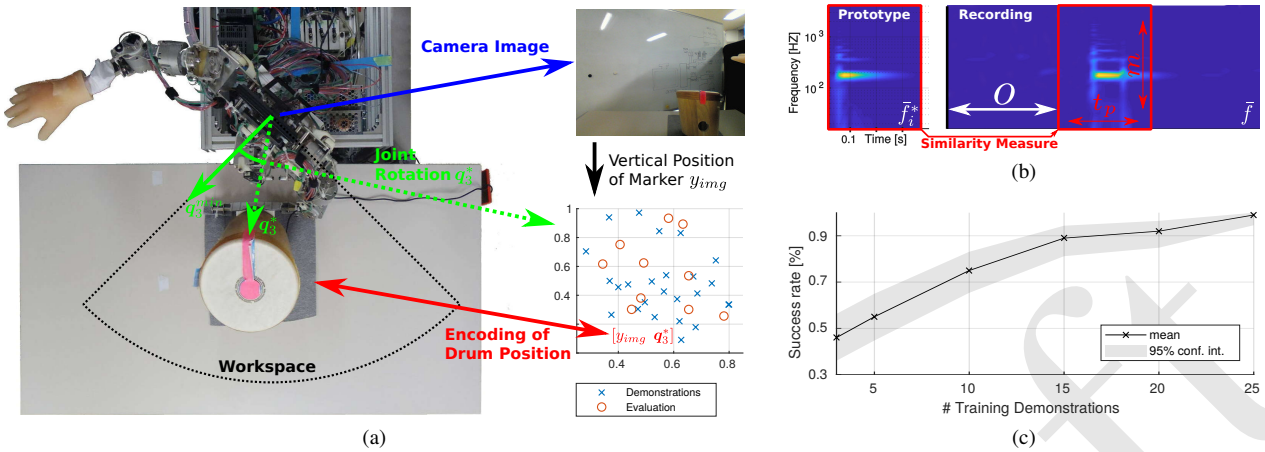


Fig. 3. (a) Experimental setup of drumming scenario. Extraction of low dimensional task parameterization and relation to drum position can be seen. (b) Visualization of similarity measure $\bar{f} \circledast \bar{f}_i^*$. (c) Success rate of drumming actions in relation of number of consolidated successful demonstrations by the parameterized skill.

In order to simulate a systematic disturbance on the data we generate 20 new data points and apply a rotation of 180° to them. The resulting target data together with the original ELM is shown in Fig. 2b. The prediction MSE is 41.71.

Finally, we select five target data points randomly and use them for training a transfer mapping with our proposed transfer learning algorithm. Using these transfered target data we can employ the original ELM to evaluate the quality of the transfer by calculating the MSE. Repeating this transfer step 100 times with different random training points yields the averaged MSE $0.001(\pm 0.001)$ for the points used to train the transfer and $0.354(\pm 0.661)$ for the other points (standard deviations in brackets). An example run is shown in Fig. 2c.

V. Experiments II: Drumming Through Mirror on Humanoid Robot

This section aims at the evaluation of transfer learning for complex robot skills. The upper body of the humanoid robot Affetto has to play a drum positioned on a table in front of the robot, as shown in Fig. 3a. For training, the robot is able to observe the drum position directly which results in the task parameterization and training samples for the parameterized skill are gathered by kinesthetic teaching. After successful task acquisition, the scenario is modified and the robot is not allowed to observe the drum position directly, as shown in Fig. 1b. The robot has to learn to utilize a mirror, to transfer the previously learned skill of drumming to this new situation.

The camera attached to the upper body of the robot performs a simple visual search and blob detection of the marker attached to the drum, giving the horizontal $x_{\text{img}} \in [0, 1]$ and vertical $y_{\text{img}} \in [0, 1]$ position of the center, normalized for drum positions in the workspace. To estimate the task parameterization, the robot moves to a fixed starting configuration $\boldsymbol{q}^{\text{start}}$ (shown in Fig. 4) and centers the marker of the drum in the image of the camera by only rotating the upper body orientation q_3. The task parameterization $\boldsymbol{\tau} = [y_{\text{img}}\ q_3^*]$ includes the final rotation of the upper body q_3^* as well as the height of the marker in the visual image of the camera, resulting in a 2D coordinate that represents the position of the drum relative to the robot. The estimation of the task parametrization is illustrated in Fig. 3a.

For the second condition of our experiments, the Affetto robot is not allowed to observe the drum directly and has to learn a new parameterized skill $\hat{PS}$. As shown in Fig. 1b, the robot is commanded to rotate its upper body into the direction of a mirror. As before, the marker position of the drum is extracted by blob detection. The rotation angle of the upper body is fixed, the task parameterization $\hat{\boldsymbol{\tau}} = [x_{\text{img}}\ y_{\text{img}}] \neq \boldsymbol{\tau}$ is given by the perceived location of the reflection of the marker in the mirror. Accordingly, there is a considerable difference in the mapping $\hat{PS}(\hat{\boldsymbol{\tau}}) \neq PS(\hat{\boldsymbol{\tau}})$, so that relearning of $\hat{PS}(\hat{\boldsymbol{\tau}})$ becomes necessary.

Robot Platform: The experiments are carried out on the humanoid robot platform Affetto, a pneumatically-actuated highly compliant robot with a 22DOF upper body structure. For the experiments we utilize 8DOF, including 3DOF of the abdomen and the right arm and an unactuated soft rubber hand. Policies define joint angle trajectories that are forwarded to the low-level joint controller. To enhance the quality of the tracking performance, we refer to the PIDF controller [38] for the pneumatically driven joints of the robot and optimize the controller parameter by automatic optimization and hand tuning on a test trajectory that includes sine waves and step responses. According to [38], the valve opening is controlled by:

$$v_j^+ = k_F(u_j^{PID} - p_j^{PD}) \tag{7}$$

and vise versa $v_j^- = -v_j^+$ for the antagonistic chamber.

Kinesthetic Teaching Mode: To initiate the teaching mode, the joint PIDF controller are commanded to move the joints of the robot to a predefined initial posture $\boldsymbol{q}^{start}$. After convergence of the robot to the initial posture, the control signals u_j^{PID} of the equilibrium states of the joints j are collected as u_j^{eq} and used as an offset for the feedback

Final Draft

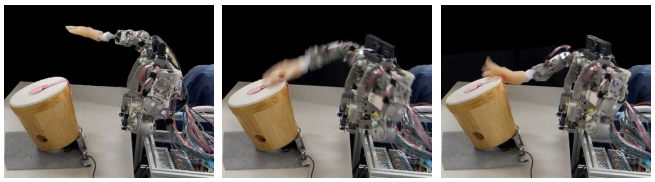


Fig. 4. Snapshots of generalized drumming action. Starting configuration $\boldsymbol{q}^{\text{start}}$ is shown in the leftmost picture.

controller:

$$v_j^+ = k_F(u_j^{PID} + u_j^{eq} - p_j^{PD}) \tag{8}$$

Additionally, we deactivate the integration of error by setting the integral component I of the controller to zero. It can be expected that u_j^{eq} reflects the integral part of the controller as the proportional and derivative components are zero for equilibrium states. A deflection of the robot joint configuration $\boldsymbol{q}^{start}$ during the demonstration phase results in a counter force given by the feedback controller proportional gains that aim to move the robot back to its initial configuration. Each recording is run for 3 sec. and the resulting trajectory is encoded into $\boldsymbol{\theta}$ by the DMP.

A. Learning to Drum

In comparison to our previous work [16], no policy optimization is performed, we collect successful human demonstrations for $N_{\text{tr}} = 25$ drum positions randomly distributed in the workspace of the robot. A demonstration can be considered successful in case the execution of the recorded trajectory by the robot results in a drumming sound. Resulting in the training set $\mathcal{D} = \{(\boldsymbol{\tau}^k, \boldsymbol{\theta}^k)|k = 1, \ldots, N_{\text{tr}}\}$, that is presented in a random order for an incremental update of the parameterized skill PS, as introduced in Sec. II-A. All demonstrations are encoded as a $K = 15$ dimensional DMP for each of the 8 DOF of the robot, resulting in a $F = 120$ dimensional parameterization of $\boldsymbol{\theta}$. For an objective evaluation of the success rate of generalization to unseen drum positions, we define a reward function based on a distance measure of the recorded audio spectrum to the prototypes gathered by the execution of training demonstrations. The similarity measure of a recorded spectrum to one prototype is given by the operator $\circledast : \mathbb{R}^{m \times t_s} \times \mathbb{R}^{m \times t_p} \to \mathbb{R}, \quad (S, P) \mapsto d = S \circledast P$ for input spectrum S, prototype P, m extracted frequency bands and time-steps $t_p \geq t_t$:

$$s \circledast p \stackrel{\text{def}}{=} \min_{0 \geq o \geq t_s - t_p} \left(\sum_{i=1}^{m} \sum_{j=1}^{t_p} \left(s(i, j+o) - p(i,j)\right)^2 \right)^{1/2}, \tag{9}$$

as visualized in Fig. 3b. The reward function for a recorded spectrum $\bar{f}(\omega, t)$ is given by:

$$R(\bar{f}) = \max_{1 \geq i \geq N_{tr}} \frac{\|\bar{f}_i^*\| - \bar{f} \circledast \bar{f}_i^*}{\|\bar{f}_i^*\|}, \tag{10}$$

with $\|\bar{f}_i^*\|$ acting as normalization of different prototype activation strengths to a maximum reachable reward of one. Hidden Layer size of the ELM was set to $H = 50$ with a regularization $\lambda = 10^{-4}$ for online learning as in [22]. We estimate the generalization performance in terms of success rate on a fixed set of $N_{\text{te}} = 10$ positions of the drum that are not part of the training set, as shown in Fig. 3a

The success rate is estimated by a simple threshold operation on the reward function and counted as successful if $R(\bar{f}) > 0.15$. Fig. 3c shows the results of the evaluation, it can be seen that the Affetto robot acquires the skill of drumming for all evaluation positions after presentation of all 25 human demonstrations.

B. Transfer Learning with Mirror

To solve this modified task we evaluate four learning schemes: i) We ignore the modification of the parameter space and evaluate the previously acquired parameterized skill PS as in Sec. V-A; ii) We relearn the task from scratch, in the same way as in Sec. V-A; iii) We reuse the parameterized skill obtained in Sec. V-A and continue training with new human demonstration samples by incremental learning. thereby ignoring the modification of the parameter space; iv) We apply Transfer Learning as proposed in Sec. III. Human demonstrations are utilized to estimate $\tilde{\boldsymbol{H}}$ by application of Eq. 5.

Let $\hat{\mathcal{D}} = \{(\hat{\boldsymbol{\tau}}^k, \boldsymbol{\theta}^k)|k = 1, \ldots, \hat{N}_{\text{tr}}\}$ be the new dataset for transfer learning. We select $\hat{N}_{\text{tr}} = 6$ human demonstrations for drum positions distributed in the workspace of the robot. Each learner is incrementally trained until convergence with 3-5 randomly selected samples of $\hat{\mathcal{D}}$ and generalization performance is evaluated for 6 randomly selected unseen drum positions. The experiment is repeated ten times and the results of the evaluation can be seen in Fig. 5.

Thereby, a baseline is given by the evaluation of the previously learned skill $PS(\hat{\boldsymbol{\tau}})$ (i) resulting in a low performance due to the modifications of the task. Continued training of $PS(\hat{\boldsymbol{\tau}})$ (iii) with new samples is also not able to adapt to the new task situation. A significantly better performance can be reached by transfer learning (iv) in comparison to relearning from scratch (ii).

VI. Discussion

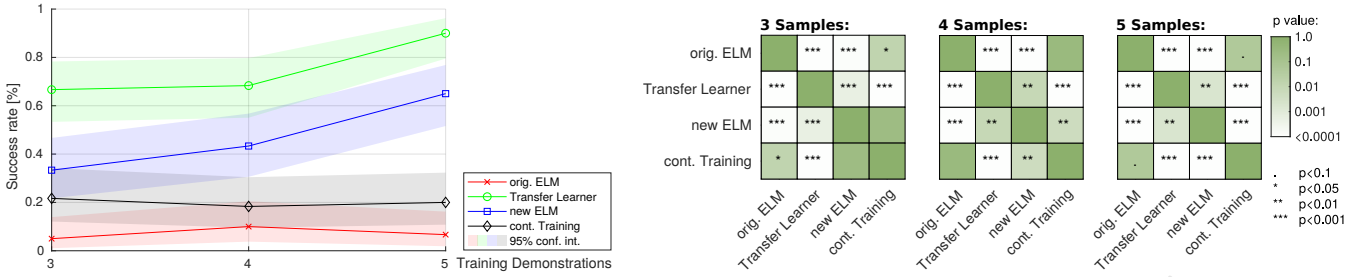
In this contribution, we presented a novel transfer learning algorithm aiming at domain adaptation problems with few labeled instances from the target domain and without correspondence information between the source and target space. We have applied this method to a toy data set for illustration and to a real world robot scenario in order to transfer complex motor skills. The approach significantly outperformed two baselines and a retrained model.

Future work includes testing the applicability of very new approaches [39], [40] in the current setting and comparing their performance to the proposed technique.

Acknowledgment

Funding from the Cluster of Excellence 277 Cognitive Interaction Technology, the CODEFROR project (FP7-PIRSES-2013-612555) - `https://www.codefror.eu/` and by PRESTO, JST Grant Number JPMJPR1652 is gratefully acknowledged.

Final Draft



(a) (b)

Fig. 5. (a) Evaluation of the Transfer Learning approach against three test conditions: No modification to ELM for new situation, train new ELM and continue training of old ELM. (b) Significance analysis of results for 3,4 and 5 presented training samples. Confidence interval based on evaluation of 10 repetitions with 6 random unseen drum positions.

REFERENCES

[1] F. Günter, "Using reinforcement learning for optimizing the reproduction of tasks in robot programming by demonstration," Ph.D. dissertation, STI, Lausanne, 2009.

[2] C. Cai and H. Jiang, "Performance Comparisons of Evolutionary Algorithms for Walking Gait Optimization," in *ISCC-C*, 2013, pp. 129–134.

[3] H. Ishihara, Y. Yoshikawa, and M. Asada, "Realistic child robot *Affetto* for understanding the caregiver-child attachment relationship that guides the child development," in *ICDL*, vol. 2, 2011, pp. 1–5.

[4] H. Ishihara and M. Asada, "Design of 22-dof pneumatically actuated upper body for child android "affetto"," *Advanced Robotics*, vol. 29, no. 18, pp. 1151–1163, 2015.

[5] B. C. da Silva, G. Baldassarre, G. Konidaris, and A. Barto, "Learning parameterized motor skills on a humanoid robot," in *IEEE Intern. Conf. Robotics and Automation*, 2014, pp. 5239–5244.

[6] J. Kober, A. Wilhelm, E. Oztop, and J. Peters, "Reinforcement learning to adjust parametrized motor primitives to new situations," *Autonomous Robots*, vol. 33, pp. 361–379, 2012.

[7] J. Peters, J. Kober, K. Mülling, O. Krämer, and G. Neumann, *ECML PKDD 2013, 2013, Proceedings, Part III*, 2013, ch. Towards Robot Skill Learning: From Simple Skills to Table Tennis, pp. 627–631.

[8] T. Matsubara, S.-H. Hyon, and J. Morimoto, "Learning parametric dynamic movement primitives from multiple demonstrations." *Neural Networks*, vol. 24, no. 5, pp. 493–500, 2011.

[9] A. Ude, M. Riley, B. Nemec, A. Kos, T. Asfour, and G. Cheng, "Synthesizing goal-directed actions from a library of example movements," in *IEEE-RAS Intern. Conf. on Humanoid Robots*, 2007, pp. 115–121.

[10] F. Stulp, G. Raiola, A. Hoarau, S. Ivaldi, and O. Sigaud, "Learning compact parameterized skills with a single regression," in *IEEE-RAS Intern. Conf. on Humanoid Robots*, 2013, pp. 417–422.

[11] A. Baranes and P. Oudeyer, "Active learning of inverse models with intrinsically motivated goal exploration in robots," *Robotics and Autonomous Systems*, vol. 61, no. 1, pp. 49–73, 2013.

[12] R. F. Reinhart and J. J. Steil, "Efficient policy search in low-dimensional embedding spaces by generalizing motion primitives with a parameterized skill memory," *Autonomous Robots*, vol. 38, no. 4, pp. 331–348, 2015.

[13] S. J. Pan and Q. Yang, "A survey on transfer learning," *TKDE*, vol. 22, no. 10, pp. 1345–1359, Oct 2010.

[14] S. M. Salaken, A. Khosravi, T. Nguyen, and S. Nahavandi, "Extreme learning machine based transfer learning algorithms," *Neurocomput.*, vol. 267, no. C, pp. 516–524, Dec. 2017.

[15] B. Paaßen, A. Schulz, J. Hahne, and B. Hammer, "Expectation maximization transfer learning and its application for bionic hand prostheses," *Neurocomputing*, 2018.

[16] J. F. Queißer, R. F. Reinhart, and J. J. Steil, "Incremental bootstrapping of parameterized motor skills," in *Humanoids*, Nov 2016, pp. 223–229.

[17] A. J. Ijspeert, J. Nakanishi, H. Hoffmann, P. Pastor, and S. Schaal, "Dynamical movement primitives: Learning attractor models for motor behaviors," *Neural Computation*, vol. 25, no. 2, pp. 328–373, 2013.

[18] T. Kulvicius, K. Ning, M. Tamosiunaite, and F. Wörgötter, "Joining movement sequences: Modified dynamic movement primitives for robotics applications exemplified on handwriting." *IEEE Trans. Robotics*, vol. 28, no. 1, pp. 145–157, 2012.

[19] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006.

[20] N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "A fast and accurate online sequential learning algorithm for feedforward networks," *T. on Neural Networks*, vol. 17, no. 6, pp. 1411–1423, 06.

[21] K. Neumann and J. Steil, "Optimizing extreme learning machines via ridge regression and batch intrinsic plasticity," vol. 102, p. 2330, 2013.

[22] H. T. Huynh and Y. Won, "Online training for single hidden-layer feedforward neural networks using RLS-ELM," in *CIRA*, 2009, pp. 469–473.

[23] J. Zhao, Z. Wang, and D. S. Park, "Online sequential extreme learning machine with forgetting mechanism," *Neurocomputing*, vol. 87, pp. 79–89, 2012.

[24] B. Nemec, R. Vuga, and A. Ude, "Efficient sensorimotor learning from multiple demonstrations," *Advanced Robotics*, vol. 27, no. 13, pp. 1023–1031, 2013.

[25] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, "Analysis of representations for domain adaptation," in *NIPS*, 2006, pp. 137–144.

[26] J. Huang, A. Gretton, K. M. Borgwardt, B. Schölkopf, and A. J. Smola, "Correcting sample selection bias by unlabeled data," in *NIPS*. MIT Press, 2007, pp. 601–608.

[27] P. Blöbaum, A. Schulz, and B. Hammer, "Unsupervised Dimensionality Reduction for Transfer Learning," in *ESANN*, 2015, pp. 507–512.

[28] C. Wang and S. Mahadevan, "Manifold alignment using procrustes analysis," in *ICML*. ACM, 2008, pp. 1120–1127.

[29] B. Bcsi, L. Csat, and J. Peters, "Alignment-based transfer learning for robot models," in *The 2013 International Joint Conference on Neural Networks (IJCNN)*, Aug 2013, pp. 1–7.

[30] M. E. Taylor and P. Stone, "Transfer learning for reinforcement learning domains: A survey," *JMLR*, vol. 10, pp. 1633–1685, 2009.

[31] T. Killian, S. Daulton, G. Konidaris, and F. Doshi-Velez, "Robust and Efficient Transfer Learning with Hidden-Parameter Markov Decision Processes," *ArXiv e-prints*, Jun. 2017.

[32] M. Wulfmeier, I. Posner, and P. Abbeel, "Mutual alignment transfer learning," *CoRR*, vol. abs/1707.07907, 2017.

[33] M. K. Helwa and A. P. Schoellig, "Multi-robot transfer learning: A dynamical system perspective," *CoRR*, vol. abs/1707.08689, 2017.

[34] M. S. Malekzadeh, S. Calinon, D. Bruno, and D. G. Caldwell, "A skill transfer approach for continuum robots-imitation of octopus reaching motion with the stiff-flop robot," in *AAAI TROB*, 2014, pp. 49–52.

[35] A. Fachantidis, I. Partalas, M. E. Taylor, and I. Vlahavas, "Transfer learning via multiple inter-task mappings," in *EWRL*, S. Sanner and M. Hutter, Eds., 2012, pp. 225–236.

[36] B. Paaßen, A. Schulz, and B. Hammer, "Linear Supervised Transfer Learning for Generalized Matrix LVQ," in *NC*^{2}, 2016, pp. 11–18.

[37] R. Fletcher, *Practical Methods of Optimization; (2Nd Ed.)*. New York, NY, USA: Wiley-Interscience, 1987.

[38] E. Todorov, C. Hu, A. Simpkins, and J. Movellan, "Identification and control of a pneumatic robot," in *BioRob*, 2010, pp. 373–380.

[39] C. Finn, T. Yu, T. Zhang, P. Abbeel, and S. Levine, "One-shot visual imitation learning via meta-learning," *CoRL*, 2017.

[40] T. Yu, C. Finn, A. Xie, S. Dasari, T. Zhang, P. Abbeel, and S. Levine, "One-shot imitation from observing humans via domain-adaptive meta-learning," *CoRR*, vol. abs/1802.01557, 2018.

Final Draft