Semantic structure for robotic teaching and learning

Sayanti Roy
Robotics Cognition Laboratory
Department of Computer Science
Oklahoma State University
Stillwater, Oklahoma 74078
sayanti.roy@okstate.edu

Emily Kieson
Department of Psychology
Oklahoma State University
Stillwater, Oklahoma 74078
kieson@okstate.edu

Charles Abramson
Department of Psychology
Oklahoma State University
Stillwater, Oklahoma 74078
charles.abramson@okstate.edu

Christopher Crick
Robotics Cognition Laboratory
Department of Computer Science
Oklahoma State University
Stillwater, Oklahoma 74078
chriscrick@cs.okstate.edu

ABSTRACT

Instructing human novices on complex tasks in non-standardized environments are an underexplored potential use for social co-robots, since instruction and skill transfer involving human experts can require an enormous commitment of time and resources. In this paper, we enable a humanoid Baxter robot to build a semantically accessible framework for task learning, teaching and representation via active learning with human experts using hierarchical semantic labels. This process not only helps the robot to learn tasks from expert demonstrations, but later improves the ability of the robot to teach novice human operators. Our results show that the better-understood learning from demonstration (LfD) task is greatly enhanced by the active learning and mutual semantic structure building in a expert-robot partnership, while the robot's ability to teach novices is improved, though the results are suggestive rather than conclusive at this point. We discuss the important aspects and power of learning *and teaching* from demonstration and how both benefit from communication and joint human-robot creation of semantic hierarchies.

1. INTRODUCTION

Learning from demonstration (LfD) [1, 2] enables end-users to teach tasks to robots without traditional programming. This hopefully involves less time and effort and adds more flexibility for non-expert use of robots in a wide array of circumstances. LfD is a well-studied problem, though a great deal of research work remains before its promise is realized. In this work, we not only explore the use of building a human-understandable representation of a complex task through active learning and a conversational interface to improve the LfD process, but we explore the problem of robotic teaching as well. A set of American Sign Language (ASL) motions is taught to the humanoid robot Baxter, which learns to communicate the sentence "Hello, please listen to me" using its left arm. During the learning process, Baxter can request information from the human expert using label and demonstration queries. While learning, the robot is able to segment and hierarchically construct the components of the demonstrated task using expert feedback. This enables the human expert operator to understand the learning activity of the robot and jointly guide the teaching process, depending on the input the robot receives from the end-user.

While learning from demonstration, the robot improves its learned model by interacting with its human teacher and cooperatively building a structure of hierarchical semantic labels [3, 4, 5], as illustrated by our experimental results. The goal of the work described here is to create robots with sufficient human-accessible task understanding so that they can act as successful tutors or coaches for complex skill learning. Compared to experts, novice learners are often unsuccessful at noticing important information and task patterns. They fail to prioritize subtasks in terms of their importance to the overall goal, and hence tend to attend inappropriately to distracting or irrelevant aspects of the task environment. If the entire task can be segmented into well-organized subtasks with meaningful labels, the novice should be better able to distinguish relevant relationships, and should learn better.

In this paper we quantify the performance of Baxter as a learner and a teacher using the root-mean square error (RMSE) of the robot's arm trajectories, compared to a gold-standard task demonstration. When Baxter is employed as a learner, the RMSE is calculated comparing the task success under traditional LFD versus query-driven active learning which jointly builds a semantic structure with a human expert. Participants are recruited to act as experts and novice human operators. The participants allotted in the novice group are taught by Baxter with and without this semantic

ACM ISBN 978-1-4503-2138-9.
DOI: 10.1145/1235

scaffolding. The RMSE is calculated comparing the performances of these two groups to evaluate their performance.

2. RELATED WORK

Many contemporary researchers are now working in the fertile field of LfD, and a few of them are discussed in this section.

2.1 Robot Learning from Demonstration

Konidaris [6] describes robots that can learn from trajectory demonstrations by constructing skill trees (CST). Chains from multiple human expert demonstrations can be merged into a single skill tree with a policy learning algorithm which efficiently increases robot learning rate. Crick [7] and Knox [8] illustrate that human experts directly teaching a robot is usually a better option than the same robot learning virtually, as humans have better understanding of the environment and a better decision making ability than unsophisticated robotic controllers. Similar work has been performed by Grizou [9], where unknown human teaching instructions are utilized by the robot to improve its learning. This work addresses how a robot can use unfamiliar and noisy teaching instructions to acquire knowledge to generate new tasks, and use that knowledge to improve its learning policy in an inverse reinforcement learning domain. While interacting, the robot tends to ask different questions to the end-user. Cakmak [10, 1] describes different platforms where the robot is trained to ask good questions and how their performance is improved via human feedback.

2.2 Robot Teaching via Demonstration

Humans have distinct teaching strategies [11] which can be effectively utilized in human-robot communication to build effective robot learners. A robot learning from human feedback tends develop a mental model of its own which can be later utilized to teach novice human operators. Scassellati [12][13] discusses human-robot collaboration for social good. If robots and humans can interact on an interpersonal level, achieving complex tasks is easier. In this work, feedback-based human-robot interaction demonstrates that if humans are guided by the robot or vice-versa, relevant questions are addressed, and with continuous collaboration the task becomes easier. If robots are able to teach human novice operators, this can improve their social reliability and enhance people's eagerness to interact with them. In this work, we describe human-robot interaction where the robot acts a teacher to guide humans to achieve complex sets of tasks.

3. HIERARCHAL SEMANTIC LABELS IN ACTIVE LEARNING

In comparison to traditional flat labeling [4], hierarchical semantic structures are more successful at capturing important information during a task execution. Within the framework of active learning and with the robot's participation, we generate hierarchal semantic labels for defining and modeling the robot's task. This is built from a set of queries and provides a better understanding of the task. These meaningful labels are important to establish relationships between subtasks, generating a human-accessible cognitive framework which can be later utilized to teach a novice human human operator.

3.1 Learning using Semantic Labels

In this work, we use a segmentation method based on action primitives to identify important change points and low confidence motions, indicating poses where the robot should ask disambiguating questions to the expert. The queries are answered with semantic labels and demonstrations, and these are later utilized for teaching purposes.

3.1.1 *Action primitive based change point identification*

Expert user demonstrations are taken into consideration to generate the learned motion, which is later passed through a segmentation algorithm to initiate the learning process in an action primitive framework. Action primitives are the fundamental building blocks of a complex task [14], and can be used to project user actions from controller inputs. Since the model requires instructor intervention using a joystick, action primitive based segmentation is preferred over trajectory-based methods like dynamic motion primitives (DMP) [15]. Within the segmentation algorithm, the continuous state action spaces are decomposed into discrete state-action pairs which later give the user insight into the number of unique classes derived from the action primitives.

Demonstrations $D_1, D_2, D_3, \dots, D_n$ are given by the expert demonstrator to generate the learned motion L using dynamic time warping with barycenter averaging (DBA) [16, 17]. This learned motion L obtained from the robot consists of the joint positions x_t and joint velocities v_t for J joints over time $T = \{t_1, t_2, t_3, \dots, t_n\}$. Since any joint j provides a single degree of rotational freedom, three broad joint movements are possible: 1) counter-clockwise, 2) stationary (with some assumed noise), and 3) clockwise, $S = \{1, 2, 3\}$. In our model, the action primitives are not differentiated on by the magnitude of the joint velocity (although nothing prevents this in our approach), only the direction. The sign language demonstration uses six degrees of freedom in the left arm, two each of shoulder, elbow and wrist joints denoted s_0, s_1, e_0, e_1, w_0 and w_1. Hence at a particular time t, if the action primitive consists of $\alpha_t = [122123]^T$, then s_0 and e_1 are moving counter-clockwise, s_1, e_0 and w_0 are stationary (possibly with some noise), and w_1 is moving clockwise.

Action primitives generated by considering the joint positions x_τ and the signs of the joint velocities v_τ are used as input to our segmentation algorithm over time $\tau = \{\tau_1, \dots, \tau_n\}$. To define the distribution of each action primitive class, sampled velocities are clustered assuming a Gaussian distribution for each cluster. Let v_{τ_j} be the sampled velocity of joint j; then to determine clusters we use k-Means with $k = 3$ for states $S = \{1, 2, 3\}$. Based on these parameters, we calculate the mean μ_{ji} and the variance σ^2_{ji} to define the probability distribution:

$$p(r_j = i | v_{\tau j}) = N(v_{\tau j} | \mu_{ji}, \sigma^2_{ji}) \tag{1}$$

and assign the variable r_j for each joint j as follows:

$$\begin{aligned} r_j = 2, p(r_j = 2) > \eta \\ r_j = 1, p(r_j = 1) > p(r_j = 3) \\ r_j = 3, p(r_j = 3) < p(r_j = 1) \end{aligned} \tag{2}$$

where η is a noise threshold computed from a hand-labeled trajectory. This procedure is repeated for each time instant to generate the class of each action segment. Now the similar neighbouring segments are merged with each other to find the unique action primitives $A_i = n(a_i)$ over time $\tau = \{\tau_i, \tau_{i+1}, \tau_{i+2}, \ldots, \tau_n\}$. Each segment is considered as a subtask, and the subtask boundaries, or change points, are considered to be profitable points for obtaining expert feedback.

From the learned sign language motion L, the algorithm identified 20 unique classes of change points, which became the basis for establishing a semantic structure based on active learning queries, as described in the next section. Each class is associated with a unique trajectory. Since the complex motion contains a sequence of sub tasks, repetition of the same classes after certain intervals is a common phenomenon. Fig. 1 illustrates the outcome of query based learning; the basic components of the structure are identified using the change point algorithm [2, 8], while the organization and semantic content of the structure are determined from queries.

3.1.2 *Query based learning*

Active learning using queries to leverage human task understanding and effort cooperatively creates an informative task representation useful to both humans and robots. After subtask segmentation and change point generation, the subtask boundaries are used to generate label and demonstration queries for each point.

1. *Is the motion correct?*
2. *What is the name of this motion?*
3. *Please show me the demonstration?*

The first two are the label queries and the third is the demonstration query. If the answer to the first question is yes it proceeds to the second question for the semantic label and saves the motion. If the answer of the first query is no , then it asks the third question and saves the demonstrated motion and the corresponding query.

If the answer to the first query is unimportant, it does not save that particular motion or state and proceeds to the next motion without asking any further questions, as the motion is irrelevant to that particular task.

As an example, in one interaction with an expert, after the segmentation process, the robot asked twice for demonstrations at different subtask boundaries, and two movements were discarded as unimportant. Fig. 1 clearly depicts the entire semantic structure learned by the robot, after both task segmentation and query-driven structure building. The blue boxes denote the primary and the secondary tasks to be performed by the robot. The red boxes denote the discovered motions by the robot at each change point. The red circles denote the endpoints of a task. After the demonstration, the robot returns to the position where the last subtask ended and progresses to the next motion from there. The green boxes designate repeating subtasks, discovered when the robot learned their appropriate semantic labels. When the expert demonstrator provides a label in response to the robot's first question, the system matches with the preconditions or previous learned semantic labels. If the labels already existed the robot announces, "*I know this motion*" and executes the trajectory by itself. Alternately, it communicates, "*Sorry, this is new to me*". Thus the entire movement is executed by the robot with human expert feedback, is saved as a semantic hierarchical structure, and is later utilized to teach novice human operators.

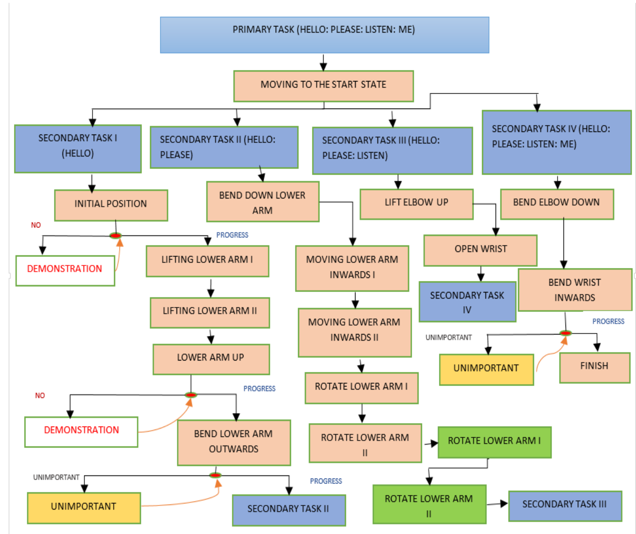

Figure 1: An example of a learned semantic hierarchical structure, obtained by task segmentation, change point detection, and label and demonstration queries obtained in conversation with an expert demonstrator.

3.2 Teaching using Semantic Labels

After the learning process with human interaction, the robot is employed to demonstrate and teach the same task to a novice human operator. Each subtask with its corresponding semantic structure and label is used for the teaching purpose [5]. The novice human operator observes the task demonstration with the labels and gains a better understanding of the motion and the task. In section 5, we demonstrate an experiment which explores the performance difference between novices who are taught merely by demonstration, without the semantic scaffolding developed during the teaching task, and novices who have the advantage of human-accessible semantic structure.

4. EXPERIMENT: ROBOT LEARNING FROM DEMONSTRATION

In this experiment participants are recruited as experts to demonstrate an American Sign Language (ASL) motion to the robot using a joystick in an HRI setting. During the experiment, the complex task given to the participant to perform demonstrates the signs for "Hello, please?". The original experiment design incorporated the whole sentence "Hello, please listen to me" but it was scaled back to the simpler phrase, involving only 9 control inputs, due to time constraints. The test subjects recruited were not well acquainted with robotic motions or with the control inputs, although they knew the intended ASL motion. Hence their interactions with the robot are expected to be naive, and their responses to the robot's asking them for labels or demonstrations were expected to be interesting.

Only one sign language motion was used for the experiment to maintain the reliability and the consistency of the proposed model. The motion was one phrase, but actually consisted of a series of small tasks that were required to be taught and mimicked with specific requirements for each motion and performed in a specific order. So while the motion taken in its entirety consisted of a single phrase, the operation of the system required a complex series of tasks and motions. Introducing more motions and phrases would have required a large increase in the number of subjects and the time commitment contributed by each, which was beyond the scope and resources of the current study.

4.1 Procedure

In this experiment, $n = 18$ teaching trials were involved, 9 using a more traditional LfD technique and 9 in the semantic structure and active learning scenario. During teaching, subjects are provided with the joystick control details and all the necessary hardware information regarding Baxter. Participants in this group are made to learn a series of detailed movements using the controller to transfer information to Baxter. Participants practiced the skill set on the controller under the guidance of the researchers for at least 40 minutes prior to interacting with Baxter to become experts. They performed the series of controls a minimum of four times when interacting with Baxter in order to teach Baxter the motions. These four demonstrations are used to generate the learned motion L discussed in Section 3.1.

Participants in this group were asked to return after three days to interact with Baxter when they are asked to teach Baxter with the semantic structure interaction system. The participants were asked to practice again under supervision of researchers before they started interacting with the robot. The entire interaction was recorded and the motion generated with this interaction is regarded as the semantically learned motion.

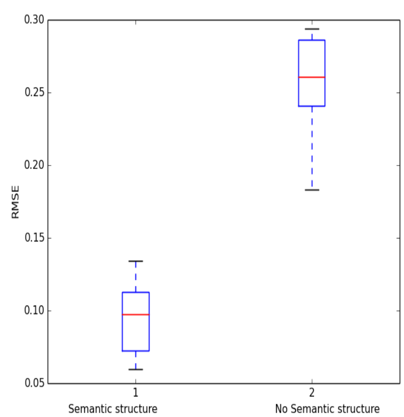

Figure 2: RMSE for motion learned through mutual construction of semantic sturcture is far better than from dynamic time warping and barycenter averaging alone ($p < 0.002$).

4.2 Results

Fig. 2 shows the comparative analysis between the traditional learned motion with the semantic learned motion. Root mean square error (RMSE) of the joint positions are calculated with respect to the ideal motion. Since the test subjects are allowed to practice for a long time in both days and also had expert supervision during the process the result is calculate irrespective of their joystick proficiency. A t-test demonstrates the very high level of statistically significant improvement between the behavior learned through the process of change point detection and cooperative building of a shared semantic representation of the task, and a motion learned simply from raw demonstrations. Semantically guided task demonstrations lead to a mutually comprehensible joint understanding of the task, and motion generated from this process is much more accurate.

5. EXPERIMENT: ROBOT TEACHING VIA DEMONSTRATION

This experiment involves performance analysis of participants when they are taught by the robot acting as an expert. The robot taught them the same sign language motion used for the experiment involving experts. In this experiment, the subjects are asked to imitate the same task they are taught by the robot, and return to demonstrate the task again after an interval of at least three days. The goal of the experiment was to evaluate the teaching action of a robot while interacting with a novice. If the novice human operator successfully manages to imitate the task taught by the robot, then we can infer that the robot has taught that person well, especially if the skill persists over time. Since semantics play a great role in providing useful information, we used our learned semantic model for teaching with semantic labels along with the corresponding gesture. The experiment is a between-subjects study where one group receives the benefit of semantic structure, while the other only receives demonstrations.

5.1 Procedure

In this experiment $n = 38$ subjects were involved. There were 20 people in the semantic labeling group and 18 in the control group. Participants in these groups learned to control the robot's arm using a joystick controller to produce the "Hello, please" sign language phrase which the robot had learned from expert demonstrators. The self-reported joystick proficiencies of the participants were noted at the beginning of the experiment and used as a control. The robot performs a motion and the participant attempts to duplicate that motion using the controller device. During the experiment the participants were provided with necessary information regarding the robot and the functionality of the joystick controller. They were allowed to take notes for their convenience. On the first day of the experiment, both groups of participants were allowed to see Baxter demonstrating the task as many times as they wished, and could practice for half an hour to get acquainted with the robot. Since this group of participants were intended to be novice human operators, they were not given any human guidance from the researchers and were only allowed to learn from the robot.

Subjects in the semantic structure group were "taught" using semantic labels assigned to each movement, which were previously developed through active engagement with an expert as described in Section 4. These labels were broken down to indicate the smaller actions that make up the entire task. Participants followed the robot's instructions to learn the movements necessary for using the joystick controller to perform the same actions, thereby mimicking the motions of an “expert”. The participants in this group can see the labels and task structure on the monitor during the task and are also given out handouts containing necessary information about Baxter. During the experiment, since Baxter is teaching them a new motion, no other human guidance is involved.

Participants in the no semantic structure group were “taught” without semantic labels, so that there are no associations between each smaller movement and any assigned categorization or word. Participants are expected to follow the robot’s demonstrations to learn the movements necessary for using the joystick controller to perform the sign language task. Participants in this group were only provided with the handout containing important information about Baxter and the joystick controller.

All of the participants in both groups were asked to return after at least two days and were asked to perform the same movements on the robot. They were only given one chance to perform the movement and were not allowed to practice or see any demonstration, although they did first have the opportunity to practice random movements with the controller. The subjects in the no semantic structure group were asked to move certain joints to perform their motion, whereas the people in the semantic group were provided the semantic labels with which they were taught, as instructions to execute the corresponding motion.

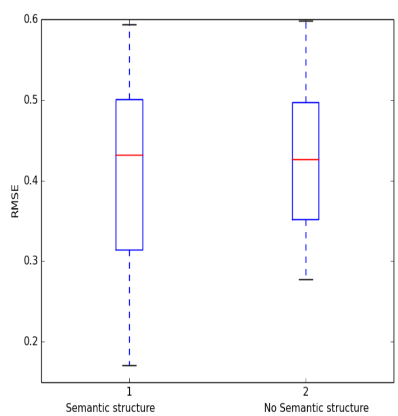

Figure 3: Performance analysis of the participants on the first day of the experiment. No significant difference between groups ($p \approx 0.5$).

5.1.1 Subjective Performance Evaluation

In Fig. 3, which represents the initial training session where the robot demonstrates the task, there is almost no difference between the performances of the two groups. This is because the subjects were allowed to see the task demonstration as many times as they wished and to practice it several times to achieve adequate performance. Subjectively, we observed that subjects in the semantic structure condition asked for fewer demonstrations than those of the group without such structure. Since we determined that joystick proficiency played an important role in subject performance, and subjects were randomly assigned to the two groups, we found that the mean self-reported joystick proficiency of the

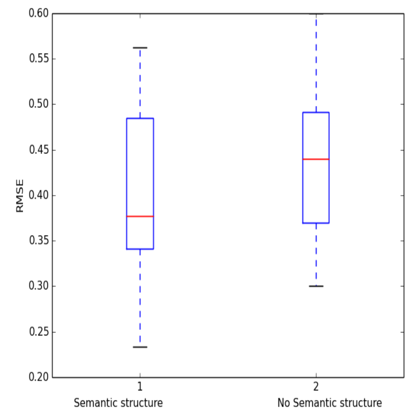

Figure 4: Performance analysis of the of the participants on the second day of the experiment. Difference is suggestive but not significant ($p = 0.18$).

non-semantic structure group (mean of 3.13 on a 5-point Likert scale) was significantly higher than that of the semantic structure group (2.51). Even when controlling for this, however, no significant difference in task performance was found.

Fig. 4 demonstrates that the subjects in the semantic structure group *may* have been able to retain their task expertise over a period of several days, compared to the subjects who were not provided with the same level of semantic assistance by the robot. A t-test shows a p-value of 0.18, suggestive but not a significant difference. Again, this data is after controlling for self-reported joystick proficiency. Subjects in the semantic structure condition may have had a more thorough understanding of the task, but participants in both groups had access to their notes, taken during the initial teaching encounter, and this may have muted the effect. Even so, the non-semantic group did perform less well, and a larger fraction had larger errors than in the semantic group – just not to the level of statistical significance.

5.1.2 Discussion

The subjects provided with semantic structure happened to have lower mean joystick proficiency than the other group. We hypothesized that subjects learning with semantic labels would retain the task better than the other group, although we did not quite confirm this to a statistically significant level, though we can still say that the robot made a decent teacher. Since all of the subjects in both groups were allowed to take copious notes and use them during the experiment, we believe that the task execution challenge became too easy, accounting for the relative success of the non-semantic structure group. In addition, the chosen sign language phrase may not have been sufficiently complex to differentiate the learning process. Even so, the subjects in the no semantic structure group did indeed perform less well than those with access to human-accessible semantic guidance.

Subjectively, participants without semantic labels more often skipped trajectories and chose incorrect motions more often. Since they were novice operators, they were not precise about their movements. For example, in the first two movements, the robot's arm is aligned perpendicularly with its shoulder in the ideal motion. However, the participants sometimes failed to achieve this pose, which caused significant deviation since the shoulder joint influences the subsequent position of all other joints. The subjects had little understanding of how precise various actions needed to be, and this was true for both groups. During the experiment, the labels and structural information were displayed on a monitor placed beside each candidate. It is likely the case that the learning and teaching would have gone better with a more audiovisual interaction. Without any audio, it was difficult for some participants to keep track of the labels on the screen and Baxter's movement at the same time. Especially during fine trajectory adjustments, it was very difficult to look at both the robot and the monitor simultaneously. These are a few of the factors which might have influenced the performance of the participants during the experiment irrespective of their groups.

6. CONCLUSION

Our main contribution is to evaluate how robotic performance can be improved by the collaborative construction of a shared, human-accessible semantic task model, and how that learned model can be of value in the further collaborative teaching of novice human operators. We have demonstrated a very significant improvement in robot learning from demonstration using this framework, and also showed some effect in the reverse problem of robot teaching via demonstration. If humans and robots can jointly develop and share cognitive task models, it should be of immense help to both tasks. If robots can be taught in this fashion by experts, it will assist robots to acquire complex skills, and in addition can be utilized for teaching purposes. These learned labelled models are human accessible and can be a great tool for learning. Our research focuses on both the learning and teaching aspect of the skill transfer problem, where robots, after acquiring necessary skills from experts, can teach the same to novices.

7. ACKNOWLEDGMENT

This work was supported by NSF award #1527828 (NRI: Collaborative Goal and Policy Learning from Human Operators of Construction Co-Robots).

8. REFERENCES

[1] M. Cakmak and A. L.Thomaz, "Designing robot learners that ask good questions," in *Proceedings of the ACM International Conference on Human-Robot Interaction (HRI)*, 2012.

[2] H. Maske, M. Matthews, A. Axelrod, H. Mohamadipanah, G. Chowdhary, C. Crick, and P. Pagilla, "Collaborative goal and policy learning from human ooperator of construction co-robots," in *Neural Information Processing Systems (NIPS)*, 2014.

[3] L. P. Kaelbling and T. Lozano-Perez, "Hierarchical task and motion planning in the now," in *Proceedings of the IEEE Conference on Robotics and Automation (ICRA)*, 2011.

[4] C. Wu, I. Lenz, and A. Saxena, "Hierarchical semantic labeling for task-relevant rgb-d perception," in *Robotics: Science and Systems (RSS)*, 2014.

[5] S. Roy, H. Maske, G. Chowdhary, and C. Crick, "Teaching and learning using semantic labels," in *Proceedings of the 12th ACM/IEEE Conference on Human-Robot Interaction (HRI)*, 2017.

[6] G. Konidaris, S. Kuindersma, R. Grupen, and A. Barto, "Robot learning from demonstration by constructing skill trees," *International Journal of Robotics Research*, vol. 31, no. 3, pp. 360–375, 2012.

[7] C. Crick, S. Osentoski, G. Jay, and O. C. Jenkins, "Human and robot perception in large-scale learning from demonstration," in *Proceedings of the 6th ACM/IEEE Conference on Human-Robot Interaction*, 2011.

[8] W. B. Knox, C. Breazeal, and P. Stone, "Training a robot via human feedback: A case study," in *Proceedings of the International Conference on Social Robotics (ICSR)*, 2013.

[9] J. Grizou, M. Lopes, and P.-Y. Oudeyer, "Robot learning simultaneously a task and how to interpret human instructions," in *IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics (ICDL)*, 2013.

[10] M. Cakmak, C. Chao, and A. L. Thomaz, "Designing interactions for robot active learners," *IEEE Transactions on Autonomous Mental Development*, vol. 2, no. 2, pp. 108–118, 2010.

[11] F. Khan, B. Mutlu, and X. Zhu, "How do humans teach: On curriculum learning and teaching dimension," in *Neural Information Processing Systems (NIPS)*, 2011.

[12] S. Strohkorb and B. Scassellati, "Promoting collaboration with social robots," in *Proceedings of the Eleventh ACM/IEEE Conference on Human-Robot Interaction (HRI)*, 2016.

[13] B. Scassellati, H. Admoni, and M. Mataric, "Robots for use in autism research," *Annual Review of Biomedical Engineering*, vol. 14, pp. 275–294, 2012.

[14] V. Kruger, D. L. Herzog, S. Baby, A. Ude, and D. Kragic, "Learning actions from observations," *IEEE Robotics & Automation Magazine*, vol. 17, no. 2, pp. 30–43, 2010.

[15] A. J. Ijspeert, J. Nakanishi, and S. Schaal, "Movement imitation with nonlinear dynamical systems in humanoid robots," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2002.

[16] F. Petitjean, A. Ketterlin, and P. Gancarski, "A global averaging method for dynamic time warping, with applications to clustering," *Pattern Recognition*, vol. 44, no. 3, pp. 678–693, 2011.

[17] G. A. ten Holt, M. J. T. Reinders, and E. A. Hendricks, "Multi-dimensional dynamic time warping for gesture recognition," in *Thirteenth Annual Conference of the Advanced School for Computing and Imaging (ASCI)*, 2007.