Transparent Learning from Demonstration for Robot-Mediated Therapy

Alexander Tyshka[1] and Wing-Yue Geoffrey Louie[1], *IEEE Member*

***Abstract*— Robot-mediated therapy is an emerging field of research seeking to improve therapy for children with Autism Spectrum Disorder (ASD). Current approaches to autonomous robot-mediated therapy often focus on having a robot teach a single skill to children with ASD and lack a personalized approach to each individual. More recently, Learning from Demonstration (LfD) approaches are being explored to teach socially assistive robots to deliver personalized interventions after they have been deployed but these approaches require large amounts of demonstrations and utilize learning models that cannot be easily interpreted. In this work, we present a LfD system capable of learning the delivery of autism therapies in a data-efficient manner utilizing learning models that are inherently interpretable. The LfD system learns a behavioral model of the task with minimal supervision via hierarchical clustering and then learns an interpretable policy to determine when to execute the learned behaviors. The system is able to learn from less than an hour of demonstrations and for each of its predictions can identify demonstrated instances that contributed to its decision. The system performs well under unsupervised conditions and achieves even better performance with a low-effort human correction process that is enabled by the interpretable model.**

I. Introduction

Autism Spectrum Disorder (ASD) is one of the most prevalent developmental disabilities and studies estimate that 1 in 44 individuals are affected [1]. ASD is characterized by impairments in social interaction and communication, which negatively impact the quality of life for these individuals [2]. Applied Behavioral Analysis (ABA) is the leading therapeutic approach for improving social skills for individuals with ASD [3]. Studies have shown interventions are most effective when administered at least 20 hours per week during early childhood [4], [5]. However, due to the large number of individuals with ASD and the time-intensive nature of ABA there is an insufficient supply of trained professionals to deliver interventions [6]. Recent research is exploring the use of robots to assist in the delivery of ABA therapy to address these supply and demand challenges [7]–[9]. In general, the goal of these works is not to replace human therapists with robots but rather use robots as a tool to reduce therapist workload and enable improved therapeutic outcomes. Recent work has even demonstrated that children with ASD can learn some social skills more easily via robot therapy [9].

Although many therapists and researchers agree that robots have the potential to improve outcomes for ASD therapy, they caution against a one-size-fits all approach [10], [11]. This is because ASD encompasses a wide spectrum of social abilities and, therefore, human judgment is necessary to develop an optimal treatment protocol. Therapists have expressed a desire to supervise the way a robot interacts with children with ASD because current systems are not always capable of adapting a robot's responses to the range of behaviors children with ASD may exhibit [12]. Human supervision provides a "safety net" that ensures optimal care is delivered in all scenarios and improves trust in such systems [10]. For robot-assisted therapy, this presents a need for personalization rather than a universal approach to meet each individual's needs.

Current research has begun to improve the ability to personalize robot-assisted therapy by developing Learning from Demonstration (LfD) approaches [7], [8]. LfD involves teaching new skills to a robot by having the robot imitate a human demonstration of a particular skill. This can improve personalization by allowing rapid re-teaching of a therapeutic robot for different children. Wizard of Oz (WoZ) is a popular technique utilized for teaching a robot through human demonstration and involves a human teleoperating the robot during a human-robot interaction while humans are led to believe the robot to be autonomous. WoZ allows for collecting demonstrations in a human-robot interaction scenario rather than a human-human scenario to minimize covariate shift.

For LfD approaches to deliver optimal patient outcomes in ABA, the state of the art must still be advanced in three key areas: generalization, transparency, and data-efficiency. First, current LfD approaches are unable to learn multiple tasks and often focus on learning a single task (e.g., teaching a greeting or emotion recognition) [7], [8]. However, ABA therapy encompasses a wide range of intervention behaviors due to both the range of skills being taught and the fact that intervention styles can vary from therapist to therapist and child to child. Second, existing LfD approaches utilize learning approaches where the learned model cannot be interpreted, so non-technical experts are unable to interpret how the system works and where it can fail [7], [8]. If robots are to learn a task and be applied to healthcare applications, it is essential that failure cases can be foreseen and mitigated because failures have the potential to negatively impact patient treatment [13]. Lastly, many existing LfD approaches utilize machine learning techniques that require a significant amount of data, but in clinical settings for children it is challenging to gather these datasets due to privacy concerns [14]. Hence, data-efficiency is crucial to make LfD practical to deploy in ABA clinic environments because large amounts of labeled data can incur significant time and monetary costs that will inhibit adoption by ABA practitioners.

In this work, we present a transparent and generalizable LfD-based approach to enable a robot to learn to deliver a

This work was supported by the National Science Foundation grant #1948224

[1]Intelligent Robotics Laboratory, Oakland University, Michigan, USA, 48309 (e-mail: louie@oakland.edu, atyshka@oakland.edu)

variety of ABA therapies by observing low-level features during a human demonstration of a therapy. Herein, we define low-level demonstration features as complex, time-domain sensor data which includes a demonstrator's voice and joint positions. This is contrasted with discrete actions that comprise a predetermined trajectory of low-level features for each discrete action to define a specific utterance and/or gesture (e.g., wave and say "Hi"). Our approach allows a therapist to use a semi-supervised teaching interface to naturally demonstrate an ABA interaction via teleoperation of a robot through upper-body and voice imitation. Clustering then segments and identifies similar discrete therapist actions from low-level demonstration features (voice and joint positions). Our approach then learns a policy of when to deliver each of the learned actions based on a given interaction state, using machine learning algorithms which require few therapist demonstrations. The robot can rationalize what factors led to each decision and point to similar instances in its training data. The resulting system enables teaching a robot to perform novel interactions in an ABA session in a generalized and interpretable manner.

II. Related Works

Recent research has proposed frameworks for learning from demonstration general social tasks [15] and specific social tasks such as robot-mediated therapy [7], [8], [16]. Senft et al. [16], [17] proposed an online learning framework called Supervised Progressively Autonomous Robot Competencies (SPARC) to learn via WoZ demonstrations to deliver robot-child interactions to children with and without ASD. The wizard selected discrete actions from a predetermined set during the interactions and Multilayer Perceptron (MLP) as well as K-Nearest Neighbor (KNN) learning algorithms were investigated for learning a policy to imitate the wizard's choices based on children's actions. The learning framework was not a fully autonomous system but rather proposed actions to a human operator with the goal of reducing workload. Winkle et al. [18] further extended the KNN-based SPARC framework to focus on long-term personalization and enable a robot to interact autonomously after several demonstrations of a fitness coaching scenario. All of these approaches used learning algorithms that could often be trained by a single individual providing relatively few demonstrations. KNN-based models have the additional advantage of transparent results by retrieving the training dataset samples that were used for a given prediction. However, a limitation of these approaches is that they use special interfaces to limit robot actions to a set of discrete choices. This requires manually creating a new action space for each task and impedes the LfD approaches from being generalized to other tasks.

Several works have turned towards data-driven methods to work with lower-level features (e.g. raw sensor data) to develop more general approaches for learning social interactions from demonstration. Such features enable the use of natural and continuous action spaces such as voice and joint positions rather than using a predefined set of discrete actions. Clark-Turner et al. [7] proposed a Deep Q-network (DQN) based approach to learn how to deliver an ABA intervention for teaching greetings. An expert teleoperator selected from three therapy actions (prompt, reward, and end session) based on responses from adult participants who simulated children's behavior during a therapy session. A DQN was then trained to select from the three discrete therapist actions in response to video, audio and gaze cues from the participants. Hijaz et al. [8] also used Deep Neural Networks (DNNs) to learn from therapist demonstrations an emotion recognition ABA therapy based on real-world clinical interactions with children. The approach not only used DNNs to select the appropriate discrete therapist action to take based a child's behavior but also used clustering methods to automatically segment, cluster, and identify therapists' discrete verbal actions from their demonstrations. The ability of these data-driven approaches to learn in more unstructured settings is a significant advantage for generalizing to real-world ABA practices.

Although data-driven methods demonstrate some advantages, they often require large datasets for high accuracy and lack transparency. The DQN approach by Clark-Turner et al. struggled to achieve high accuracy on the test dataset (67.8%) and failed to make the important distinction of when to deliver a reward (32.1% accuracy). The model correctly identified unresponsive participant behaviors with 95% accuracy but had much lower accuracy when the participants were responsive (37.5% accuracy). The authors note that this result was likely due to overfitting and that more data would be necessary to improve the model. Similarly, Hijaz et al. also noted significant overfitting and data imbalance as possible contributors to the DNN's suboptimal accuracy (43.5%). However, collecting large amounts of labeled data can incur significant costs that could inhibit adoption of robotic systems by ABA practitioners. Additionally, interpretability and transparency are essential for clinical applications but DNNs are often infeasible to interpret and considered "black-box" machine learning [19]. Although significant research has been done on post-hoc techniques to interpret black-box models, such interpretations are often criticized for convincing yet inaccurate explanations [20], [21]. This has led a growing number of researchers to caution against the use of black-box DNNs in high-stakes scenarios such as healthcare [22].

Alternative approaches exist that can be more interpretable and data-efficient than DNNs while retaining the ability to work with low-level features. Liu et al. [15] proposed a data-driven model for teaching a robot to serve as a shopkeeper. This system incorporated vectorized speech, gestural, and spatial cues and performed fully unsupervised learning by clustering different shopkeeper actions and finding their relationship to customer actions. Using several hours of recorded customer and shopkeeper verbal and nonverbal interactions, a Naïve Bayes model was able to imitate socially appropriate shopkeeper behavior. The authors noted that using well-designed abstractions of input data via clustering and dimensionality-reduction techniques created a more noise-tolerant and accurate system than using raw sensor features. The Naïve Bayes model used is also more data-efficient because it assumes independence of features and, therefore,

has fewer trainable parameters. This feature-independence also makes the model an inherently interpretable model because a human can easily see what input factors led to the final prediction [20]. Such data-efficient and interpretable learning algorithms provide promising avenues for advancing LfD approaches for robot-mediated ABA therapy.

From the existing body of work, it is clear there is a need for effective, data-efficient, and interpretable models for LfD robot-mediated therapy as well as social robotics in general. An ideal method should be able to learn from general low-level features as in [7], [8], [15], while maintaining the data-efficient aspects of simpler approaches as in [16]–[18]. Additionally, an ideal approach should be transparent to enhance therapist trust and patient outcomes. In this research, we aim to address the two challenges of data efficiency and interpretability while learning a social task from human demonstration. Herein, we present a LfD approach to learn structured social interactions such as ABA in a semi-supervised and data-efficient manner.

III. Data Collection via LfD

A remote operator (i.e., WoZ) teleoperated a NAO robot via upper-body and voice imitation during the delivery of a robot-mediated ABA therapy to gather data for training and evaluating our LfD approach, Figure 1. The procedures for collecting this dataset were reviewed and approved by the Institutional Review Board at Oakland University.

(a) Interaction scenario (b) Wizard controlling the robot

Fig. 1: LfD data collection

A. Interaction Scenario

The ABA therapy used in this study followed a Discrete Trial Training (DTT) protocol which is a highly structured ABA approach for teaching social skills via prompting, rewarding, and errorless teaching [23]. Namely, a discriminative stimulus (S^D) is used to elicit a response from a child such as asking a question or requesting them to perform an action, Figure 1(a). The following S^D's were used in this study:

- Emotion recognition from nonverbal behavior (Robot asks "how am I feeling" and acts sad, happy, or angry)
- Manding (Robot asks child for his/her preferred reward, among options of a game, video or song)
- Imitation (Robot asks child to imitate its gestural actions such as touching its head, raising its arms, or waving)
- Verbal instructions (Robot asks child to perform an action such as "wave your hand" or "touch your head" without modeling the action)
- Wh-question answering (Robot asks a question like "Who flies a plane?" or "What tells time?")

After delivering an S^D, the robot would wait for a child's response. The robot delivers a prompt if a child did not respond at all, an error correction if the child responded incorrectly, and a reward for correct responses. A prompt consists of demonstrating the correct response to teach the child the appropriate way to perform the skill. An error correction is similar to a prompt but the therapist re-delivers the S^D, and then immediately delivers a prompt. If the child again responds incorrectly, the error correction is repeated up to 3 times before moving onto the next S^D. The therapist delivers either verbal praise or a physical reward if the child responds correctly. Herein, we refer to each S^D, prompt/error correction, and reward sequence as a trial. An overall session with a child followed a standard DTT protocol and consisted of presenting nine randomly selected S^D's (i.e., nine trials) from the list presented above.

B. Participants

We do not collect data from real children with ASD but rather adult participants imitating them, because autonomous systems such as ours and [7], [8] are experimental in nature. Hence, they must be extensively validated before trials with children to mitigate risks to a child's treatment program if failures occur. We collected data from 4 English-speaking adult participants that participated in 4 sessions (36 trials per participant). The participant was instructed to respond correctly for half of the trials and incorrectly for the other half. Written informed consent was obtained from all participants prior to the data collection and participants could withdraw from the data collection at any time.

C. Teleoperation and Data Collection System

The teleoperation system allows an operator to remotely teleoperate the NAO via upper-body and voice imitation while data is collected during the demonstration of the ABA therapy, Figure 1(b). The operator's speech is captured and played back via the robot's speakers. Similar to [24], the voice that is played back is pitch-shifted to mask the human identity and make the speech sound more child-like as well as robotic. The robot captures the participant's audio and plays it back for the operator during the interaction. An Orbbec Astra Pro is also used to capture the upper body skeletal position of both the operator and participant. This allows the NAO to replicate the operator's posture and the operator to observe the participant from the robot's perspective. This system allows for bidirectional verbal and gestural communication, while maintaining the impression of an autonomous robot. In addition, the wizard is provided with a GUI interface that allows the wizard to select between 4 different prompting levels, 3 different child rewards (in response to manding) and a toggle to indicate the start of a new trial.

IV. Learning System

Our learning system aims to imitate the wizard's verbal, gestural, and discrete reward selection actions during DTT trials by learning from the wizard's demonstration data. This learning process is accomplished in two steps: 1) learning a

model of discrete wizard actions and 2) learning a policy to execute the appropriate action based on the interaction state.

A. Data

The data from a session can be subdivided into data tuples that each consist of a single trial (τ_k). Each trial is composed of the robot's joint positions ($j_{k,R}$), child joint positions ($j_{k,C}$), robot verbals ($v_{k,R}$), child verbals ($v_{k,C}$), and time-stamped reward presses (r_k):

$$\tau_k = (j_{k,R}, j_{k,C}, v_{k,R}, v_{k,C}, r_k) \tag{1}$$

$j_{k,R}$ and $j_{k,C}$ are times-series 12-dimensional vectors that represent the Cartesian position of 4 key joints (both hands and elbows). $v_{k,R}$ and $v_{k,C}$ are a series of time-stamped phrases obtained by processing the raw audio data through Google Speech-to-Text.

Each trial is an instance of a particular target skill (t_i) (e.g., recognizing the emotion sad or answering "who flies a plane?") among a set of target skills (T) being taught to the child such that $t_i \in T$. Each target skill $t_i = \{B_i, \pi_i\}$ can be modeled as a set of robot behaviors (B_i) that can be used to teach the skill to a child and a policy (π_i) that defines when during an interaction to utilize these behaviors.

For a set of behaviors $B_i = \{b_{i,1}, b_{i,2}, ...b_{i,j}\}$ each individual behavior $b_{i,j} = \{a_1, a_2, ..., a_n\}$ represents a high-level grouping of different verbal and/or gestural robot actions (a_n) with the same communicative goal. Each individual action in $b_{i,j}$ may have minor differences in delivery but serve the same purpose in an interaction (e.g. "good job" and "great job" are different actions in the "verbal praise" behavior). We also define each target skill's policy $\pi(s_{i,j}) = b_{i,j}$ as a mapping of individual interaction states ($s_{i,j}$) to behaviors.

Formally, state at time t is defined as:

$$s_t = (j_{t,C}, v_{t,C}, \{b_1, b_2, ..., b_{t-1}\}) \tag{2}$$

where $j_{t,C}$ and $v_{t,C}$ are the child's joint positions and verbals from the beginning of b_{t-1} to the current time (beginning of b_t). Our approach seeks to model T, $B_i\ \forall t_i \in T$, and π using only $\{\tau_1, \tau_2, ...\tau_k\}$.

B. Segmentation

First, trial τ_k must be segmented into its constituent robot actions A_k where $\forall a \in A_k, \exists b | a \in b \wedge b \in B_i$ for some unknown target skill t_i. The raw joint positions $j_{k,R}$ are sampled at 20Hz and first smoothed via a Savitzky-Golay filter. A scalar gestural activity signal g_k is created by taking the L2 norm of the time-derivative of the smoothed robot joints. We threshold g_k to obtain a binary signal indicating whether the robot is in significant motion. Small gaps in this signal are removed with binary dilation and erosion. A binary signal is also constructed to represent robot voice activity by using the phrase timestamps of $v_{k,R}$ obtained by Speech-to-Text. These continuous and discrete signals are visualized in Figure 2 for a sample trial. Applying the constraint that only one continuous verbal phrase can be present in an action, robot actions A_k can be segmented from τ_k by logical OR'ing the verbal and gestural signals, locating the positive regions, and sampling $j_{k,R}$ and $v_{k,R}$ at these regions.

C. Clustering

After segmenting the set of actions A_k for trial τ_k, we model T and $B_i \forall t_i \in T$ using a two stage clustering process. The two stage clustering approach consists of: 1) clustering similar actions into sets of behaviors B_k and 2) clustering and merging similar behavior sets to form T.

1) Action Distance Function: We define a distance function D_a between two actions a_1 and a_2. The action phrases $v_{1,R}$ and $v_{2,R}$ are first converted to unit vectors $\psi_{1,R}$ and $\psi_{2,R}$ using a Sentence-Transformer model [25]. We then formally define the distance function as:

$$D_a(a_1, a_2) = \alpha * \frac{dtw(j_{1,R}, j_{2,R})}{min(|j_{1,R}|, |j_{2,R}|)} + \beta * ||\psi_{1,R} - \psi_{2,R}|| \tag{3}$$

where:

$$\alpha = c * \frac{G}{G+V}, \beta = d * \frac{V}{G+V} \tag{4}$$

$dtw()$ = Dynamic Time Warping distance of timeseries
$G = avg(g_1) + avg(g_2)$
g_1, g_2 = segment of g_k for time intervals of a_1 & a_2
$V = \begin{cases} 1, & \text{if } a_1 \text{ or } a_2 \text{ has a verbal} \\ 0 & \text{otherwise} \end{cases}$
c = scaling constant for gesture
d = scaling constant for verbals

This distance metric effectively captures both gestural and verbal distance. The dynamic weights α and β serve as a simple "attention" mechanism to focus the distance on gesture or verbals. For example, if the robot is at rest during a question answering trial, the learning system should disregard gesture as it is unimportant and could add noise. Constants c and d tune the sensitivity to gesture and verbals and were determined using a small subset of training data.

2) Clustering Similar Actions: We can cluster the actions into behaviors by using the established distance metric and hierarchical agglomerative clustering. Hierarchical clustering was selected based on its ability to cluster with only distance information. This contrasts alternative methods such as K-Means that compute centroids. This is because we cannot model a centroid of two gestures but can still compute the distance between them. We perform this clustering step to cluster each trial's actions into behaviors. For k trials we obtain k behavior sets $\{B_1, B_2, ..., B_k\}$ that must then be further clustered into target skills.

3) Clustering Similar Behaviors: To cluster the behavior sets into targeted skills, we first define the distance $D_b(b_1, b_2)$ between two behaviors b_1 and b_2 as the average distance of all pairwise comparisons between the actions from each behavior. We define the distance $D_B(B_1, B_2)$ between two sets of behaviors B_1 and B_2 as the minimum-cost assignment of behavior pairs. This solves the problem of how to optimally match the behaviors in two sets so as to minimize the total distance D_b between matched behavior pairs. The min-cost assignment for two dissimilar behavior sets should be high, while for two similar sets there should exist a low-cost pairing

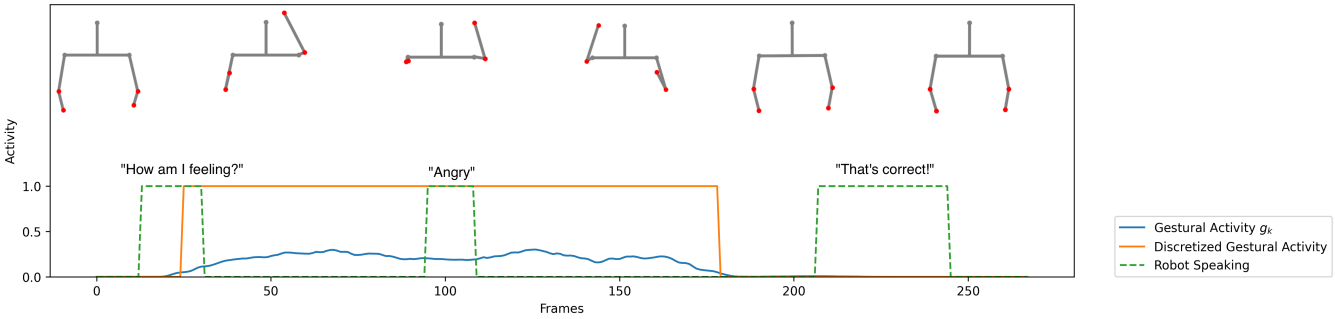

Fig. 2: Action Segmentation

of behaviors. To accomplish this, we compute a matrix of the distance D_b between all possible behavior pairs from B_1 and B_2. We obtain the optimal assignment of rows and columns that minimizes the total cost using the Hungarian algorithm. Dividing that cost by the size of the smaller set $min(|B_1|, |B_2|)$ results in the normalized distance between B_1 and B_2. Applying hierarchical agglomerative clustering to behavior sets $\{B_1, B_2, ..., B_k\}$ with distance metric D_b, we cluster similar behavior sets to form unique target skills t_i. We prune very small clusters as these are likely outliers. The set of unique target skills t_i then provides the high-level model of the target skills T.

For a given target skill t_i consisting of k trials, we could construct B_i by simply merging the behavior sets $\{B_1, B_2, ..., B_k\}$ belonging to it. However, B_i is prone to have duplicate behaviors if constructed this way. This is because the behavior clusters were initially created at the trial level, where very little information is available and noise is high (i.e. only one or two instances of a given behavior). Mismatches in the Hungarian assignment can also accumulate errors and amplify cluster noise. Hence, we merge all the actions from the behavior sets belonging to target skill t_i to obtain a set of all actions A_i for t_i. We re-compute the hierarchical clustering with A_i to obtain the new set of behaviors B_i. This creates clusters with fewer duplicates of behaviors because it uses a globally optimal solution instead of a union of locally optimal trial-level solutions.

D. Cluster Repair

Clinical settings require minimal failures in intervention delivery given their potential to negatively impact outcomes. However, mistakes in clustering actions or behaviors can occur when the process is completely automated. Given the necessity to minimize robot failures in clinical settings, we can use human labeling to amend the learned model and reduce the possibility of intervention delivery failures. This is possible because our model learns an interpretable and easily understandable model of the interaction structure (clusters). This contrasts end-to-end black-box systems which learns a complex latent space. In our learning system, a human can view the learned actions, behaviors, and skills, and correct any errors in the learned model. Herein, we refer to these model corrections as cluster repair. For each action, the human has the ability to reassign it to another behavior cluster within either the same target skill or in a different target skill. Additionally, the human can hide certain actions due to corrupted verbals or gestures (e.g., speech-to-text or wizard skeleton tracking occlusion errors). We hide actions instead of removing them because they are still useful for training the policy but should not be presented to the child when the robot samples an action from that behavior. We also provide the human with the option to delete spurious actions which should not have been recognized in the first place and which add significant noise to the policy training data. Enabling such cluster repair improves robot performance with minimal additional human effort.

E. Policy

We use a KNN model to create a policy that is both data-efficient and interpretable. For each target skill t_i, we extract the behaviors executed and the state immediately preceding them to form (state, behavior) pairs. From there, we construct a table of such pairs. In addition to the learned behaviors, we append an "End" behavior to each trial, allowing the policy to predict when to start the next skill.

To decide which action a robot should take, the policy first looks up its behavior history $\{b_1, b_2, ..., b_{t-1}\}$ in the current trial against all table entries. If one or more behavior-history matches is found in the state-behavior table, the model uses a KNN approach to compare current child states $(j_{t,C}, v_{t,C})$ against those of the table entries. We use the same distance metric used for robot action in equation (3) and select K=1 neighbor. The closest neighbor's behavior is executed if the distance is below an empirically determined threshold. Otherwise, the robot is observing a novel child behavior and lacks a conclusive answer as to how to behave. Here, the policy can soften the history constraint and select a match based on $\{b_{t-1}\}$ only rather than $\{b_1, b_2, ..., b_{t-1}\}$. This behavior can be proposed to the therapist and, if accepted, added to the table.

Similarly, if no matches with the same behavior history are found the robot can again select a behavior based on $\{b_{t-1}\}$. If there are any matches, the robot can select a behavior based on the closest child-state and propose it as a suggestion. For example, if the robot has not previously experienced a state where the S^D is delivered 3 times, this state would be missing from the table. However, other entries may indicate verbal praise is often delivered after an S^D and the robot can suggest verbal praise. The robot adds this new instance to the table if the proposed behavior is accepted.

V. Experiment

We conducted live DTT interactions to compare the performance of a policy that is learned with the raw clusters and a policy learned with the repaired clusters to investigate how the transparent learning approach could enable human refinement to increase performance. This allows us to measure the accuracy of the clustering, the accuracy of the policy, and the end-to-end unsupervised accuracy of the system.

A. Procedure

We evaluate our learning system during a live DTT interaction. The robot was controlled semi-autonomously by having the learning system utilize the learned policy to output a behavior during each state of the interaction to the wizard and allowing the wizard to confirm whether the proposed behavior is correct. When proposing a behavior, the interface notes whether the policy is certain of its decision based on a strongly similar example or uncertain of its decision. The wizard is also presented with alternative learned discrete behaviors from within that skill so that if the proposed behavior is incorrect than the wizard can select the appropriate behavior in the given interaction state. This semi-autonomous approach was used to evaluate our learning system as it provides a ground-truth behavior for each interaction state of the live DTT interaction while simultaneously ensuring that the interaction with the participant progresses correctly.

Overall, we evaluated our system with one adult participant interacting with the robot over 144 trials (16 sessions). The participant was instructed to deliver a mix of compliant and noncompliant responses at their discretion. Each trial consisted of teaching a target skill randomly selected from the set of learned skills (i.e., S^D's found in section III-A). The 16 sessions with the participant were divided evenly so that unrepaired policy was used for half the sessions (72 trials) and the repaired policy was used for the other half (72 trials). Due to data corruption 9 trials were lost from one of the policy tests. Consequently, both policy conditions were trimmed to 63 trials.

B. Measures and Metrics

Cluster accuracy was measured by comparing the original behavior and target skill clusters against the repaired clusters. We compute skill clustering accuracy as the percentage of actions whose repaired skill matched their original assigned skill. Similarly, we compute behavior clustering accuracy as the percentage of actions whose repaired behavior matched their original behavior. As described in section IV-D, we additionally include the number of hidden and deleted actions.

Repaired cluster policy accuracy was measured by training the policy on repaired clusters and comparing the suggested behaviors output from the policy against behaviors selected by the wizard. The use of repaired clusters isolates the policy from any errors in clustering and allows us to measure the performance of policy learning directly. We define policy accuracy as the percentage of accepted policy suggestions over all actions performed in the live experiment. We separately score accuracy for certain and uncertain policy decisions in addition to an overall accuracy. We also separately score the subset of skills targeting a gestural response from the participant and those skills targeting a verbal response.

Unrepaired cluster policy accuracy was evaluated following the same procedure, but using raw clusters instead of repaired clusters to learn the policy. This metric effectively evaluates the end-to-end performance of segmentation, clustering, and the policy under unsupervised conditions. Similarly to the repaired condition, we score accuracy for all decisions, for certain decisions, and for uncertain decisions, as well as separating the target skill types.

C. Results

For the 421 actions segmented from our dataset, the clustering method achieves near-perfect accuracy on skill clustering (99.5%). Upon further inspection, the 0.5% of incorrectly-assigned actions were due to a failure of the human to signal the start of a new trial. Behavior clustering also achieves a high accuracy of 75.8%. The 24% that were clustered incorrectly were mostly due to false separations of verbal praise into separate behaviors. DTT can have high variance in the phrasing of verbal praise to avoid monotony, which led to high distances from the sentence embeddings. Additionally, ~10% of the actions were selected to be hidden, almost entirely due to speech-to-text errors. With these actions hidden the repaired-cluster policy was able to deliver better responses free of grammatical mistakes, an important factor for deployment in clinical settings. Sixteen actions were also selected for deletion; these corresponded to failures in the action-segmentation system caused by false-positive gestures or divisions of one action into fragments.

The performance of both policies are presented in Table I. The repaired cluster policy achieves a high overall accuracy of 90.3% in selecting the correct behavior in a given interaction state. However, there is a significant disparity in accuracies for the policy between verbal and gestural skills. The policy achieves perfect accuracy (100%) on verbal responses but moderate accuracy (66.1%) in the gestural skills categories. We observed a strong bias against delivering verbal praise on the gestural skills. The policy also seems strongly biased to mark its decisions as certain for gestural target skills (56 out of 59). However, the perfect verbal scores indicate that the transformer-based verbal features provide excellent performance when the clustering errors are repaired.

The policy trained on unrepaired clusters performs similarly overall to the policy trained on the repaired clusters but with several key differences. As expected, the cluster noise reduced the total accuracy by 9.4%. Verbal skill accuracy was decreased by 13% but it is important to note that the certain verbal accuracy decreased by only 3%. This indicates the policy certainty correlates strongly with actual accuracy. Overall the unrepaired policy's ability to accurately select behaviors during gestural skill teaching trials was similar to the repaired policy condition. However, one key difference is the higher accuracy of the unrepaired policy on uncertain behavior predictions during gestural trials. While it is difficult to draw conclusion about this accuracy given

the limited number of instances, the disproportionate ratio of certain to uncertain gestural instances suggests the policy was overconfident in gestural target skills.

TABLE I: POLICY PERFORMANCE

		Repaired		Unrepaired	
Target Skills	**Prediction Type**	**Policy Accuracy**	**# of Instances**	**Policy Accuracy**	**# of Instances**
All	All	90.3%	207	80.7%	212
	Certain	87.1%	147	81.3%	139
	Uncertain	98.3%	60	79.5%	73
Gestural	All	66.1%	59	70.4%	81
	Certain	66.1%	56	66.7%	72
	Uncertain	66.7%	3	100.0%	9
Verbal	All	100.0%	148	87.0%	131
	Certain	100.0%	91	97.0%	67
	Uncertain	100.0%	57	76.6%	64

VI. DISCUSSION

In this work, we present a novel approach to robot-mediated therapy that can learn multiple tasks from limited human demonstrations and make interpretable decisions using its learned model. The system achieves high accuracy in end-to-end unsupervised learning and even higher accuracy with cluster repair as a form of semi-supervision that is enabled by the transparency of the models.

In future work, we hope to further improve system accuracy to the point where clinical evaluations with real children can be conducted. While the verbal analysis capabilities of our system perform extremely well, gestural analysis has room for improvement. The interpretable KNN policy was useful for identifying instances where the skeletal tracking of the participant failed due to occluded limbs or insufficient contrast. Improving the input capture of joint positions might significantly improve performance on gestural skills. Replacing the Dynamic Time Warping with other approaches could also improve gestural comparison. A learning-based gestural comparison system such as DNNs could provide better participant invariance, positional invariance, and noise tolerance which simple time-alignment techniques cannot provide. Another important piece necessary for clinical deployment is the addition of turn-taking behavior. Similar to previous work we utilized fixed-time thresholds for responses [7], [16], [17]. However, children should be praised or corrected immediately after they respond because additional delay could confuse the child or reduce engagement [10].

Finally, we note that further HRI studies would be beneficial to evaluate interpretability and explainability from therapists' perspectives. While the models used in this paper are easily examined by roboticists, it is a further challenge to intuitively convey the meaning of this data to ABA therapists not well-versed in robotics and machine learning.

REFERENCES

[1] M. J. Maenner, *et al.*, "Prevalence and characteristics of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2018," *MMWR Surveillance Summaries*, vol. 70, no. 11, p. 1, 2021.

[2] A. American Psychiatric Association *et al.*, *Diagnostic and statistical manual of mental disorders*. American Psychiatric Association Washington, DC, 1980, vol. 3.

[3] R. M. Foxx, "Applied behavior analysis treatment of autism: The state of the art," *Child and Adolescent Psychiatric Clinics of North America*, vol. 17, no. 4, pp. 821–834, 2008.

[4] L. K. Koegel, *et al.*, "The importance of early identification and intervention for children with or at risk for autism spectrum disorders," *International Journal of Speech-Language Pathology*, vol. 16, no. 1, pp. 50–56, 2014.

[5] N. Peters-Scheffer, *et al.*, "A meta-analytic study on the effectiveness of comprehensive aba-based early intervention programs for children with autism spectrum disorders," *Research in Autism Spectrum Disorders*, vol. 5, no. 1, pp. 60–69, 2011.

[6] Y. X. Zhang and J. R. Cummings, "Supply of certified applied behavior analysts in the united states: Implications for service delivery for children with autism," *Psychiatric Services*, vol. 71, no. 4, pp. 385–388, 2020.

[7] M. Clark-Turner and M. Begum, "Deep reinforcement learning of abstract reasoning from demonstrations," *ACM/IEEE International Conference on Human-Robot Interaction*, p. 372, 2018.

[8] A. Hijaz, *et al.*, "In-the-wild learning from demonstration for therapies for autism spectrum disorder," *IEEE International Conference on Robot and Human Interactive Communication*, pp. 1224–1229, 2021.

[9] J. Korneder, *et al.*, "Robot-mediated interventions for teaching children with ASD a new intraverbal skill," *Assistive Technology*, 2021.

[10] M. Sochanski, *et al.*, "Therapists' perspectives after implementing a robot into autism therapy," *IEEE International Conference on Robot and Human Interactive Communication*, pp. 1216–1223, 2021.

[11] A. M. Alcorn, *et al.*, "Educators' views on using humanoid robots with autistic learners in special education settings in England," *Frontiers in Robotics and AI*, vol. 6, p. 107, 2019.

[12] P. G. Esteban, *et al.*, "How to build a supervised autonomous system for robot-enhanced therapy for children with autism spectrum disorder," *Paladyn, Journal of Behavioral Robotics*, vol. 8, no. 1, pp. 18–38, 2017.

[13] A. F. Markus, *et al.*, "The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies," *Journal of Biomedical Informatics*, vol. 113, pp. 1–11, 2021.

[14] Centers for Medicare & Medicaid Services, "The Health Insurance Portability and Accountability Act of 1996 (HIPAA)," Online at http://www.cms.hhs.gov/hipaa/, 1996.

[15] P. Liu, *et al.*, "Data-driven hri: Learning social behaviors by example from human–human interaction," *IEEE Transactions on Robotics*, vol. 32, no. 4, pp. 988–1008, 2016.

[16] E. Senft, *et al.*, "SPARC: Supervised progressively autonomous robot competencies," *International Conference on Social Robotics*, vol. 9388 LNCS, pp. 603–612, 2015.

[17] E. Senft, *et al.*, "Teaching robots social autonomy from in situ human guidance," *Science Robotics*, vol. 4, no. 35, 2019.

[18] K. Winkle, *et al.*, "In-situ learning from a domain expert for real world socially assistive robot deployment," *Proceedings of Robotics: Science and Systems*, 2020.

[19] Y. Lou, *et al.*, "Intelligible models for classification and regression," in *ACM SIGKDD international conference on knowledge discovery and data mining*, 2012, pp. 150–158.

[20] Z. C. Lipton, "The mythos of model interpretability," *Communications of the ACM*, vol. 61, no. 10, pp. 35–43, 2018.

[21] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, 2019.

[22] D. Watson, *et al.*, "Clinical applications of machine learning algorithms: Beyond the black box," *BMJ Clinical Research*, vol. 364, p. l886, 03 2019.

[23] T. Smith, "Discrete trial training in the treatment of autism," *Focus on Autism and Other Developmental Disabilities*, vol. 16, no. 2, pp. 86–92, 2001.

[24] S. Yilmazyildiz, *et al.*, "Voice modification for wizard-of-oz experiments in robot-child interaction." Workshop on Affective Social Speech Signals, 08 2013.

[25] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," in *Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 11 2019.