Towards an Ontology for Generating Behaviors for Socially Assistive Robots Helping Young Children

Yuqi Yang[1], **Allison Langer**[2], **Lauren Howard**[1], **Peter J. Marshall**[2], **Jason R. Wilson**[1]
[1] Franklin & Marshall College
[2] Temple University
yyang5@fandm.edu, allison.langer@temple.edu, lauren.howard@fandm.edu, peter.marshall@temple.edu, jrw@fandm.edu

Abstract

Socially assistive robots (SARs) have the potential to revolutionize educational experiences by providing safe, non-judgmental, and emotionally supportive environments for children's social development. The success of SARs relies on the synergy of different modalities, such as speech, gestures, and gaze, to maximize interactive experiences. This paper presents an approach for generating SAR behaviors that extend an upper ontology. The ontology may enable flexibility and scalability for adaptive behavior generation by defining key assistive intents, turn-taking, and input properties. We compare the generated behaviors with hand-coded behaviors that are validated through an experiment with young children. The results demonstrate that the automated approach covers the majority of manually developed behaviors while allowing for significant adaptations to specific circumstances. The technical framework holds the potential for broader interoperability in other assistive domains and facilitates the generation of context-dependent and socially appropriate robot behaviors.

Introduction

The future of socially assistive robots (SARs) has great potential to supplement limited resources of service, where SARs could systematically provide a safe, non-judgmental, emotionally supportive environment for the social development of children (Belpaeme et al. 2018; Langer, Marshall, and Levy-Tzedek 2023). There are potential benefits of SAR in child-robot-interaction, such as tutoring, assisting language learning (van den Berghe et al. 2019), and providing structured environments with less face-to-face social pressure (Kim et al. 2013; Alcorn et al. 2019).

For the best educational experience, robot modalities must coordinate to maximize effective and natural interaction. To generate autonomous robot behaviors robustly and succinctly, we propose an upper ontology of SAR behaviors. An ontology describes entities, classes, and relations between them. An upper ontology defines general abstract components that are relevant to a broader range of domains. In the case of SAR, the ontology could produce action sequences from simple intents while adapting to task conditions and user-specific needs. Our hypothesis is that this ontology of social assistive intents can describe the behaviors necessary for assisting young children on puzzle tasks. We validate the approach by comparing ontology-generated behaviors to hand-coded robot behaviors which were demonstrated to be effective in an experiment with 5-8-year-old children. The comparison supports that the SAR behavior ontology generates effective assistance in child-robot interaction while providing greater flexibility and scalability.

Background

Creating an appropriate conversational agent is essential to increasing rapport in human-robot interaction (Nomura and Kanda 2016; Seo et al. 2018). In reference to socio-linguistic theories, we examined turn-taking behavior, which follows an implicit structure of claiming the floor, holding the floor, and releasing the floor in multiparty conversations (Sacks, Schegloff, and Jefferson 1978). Research has shown that both verbal and nonverbal signals, such as gaze and gestures, play a significant role in presenting and recognizing turn-taking (Holler and Kendrick 2015; Duncan 1972). These social cues serve as floor management skills that regulate conversational dynamics. The cognitive paradigm is also applied to virtual agents (Bohus and Horvitz 2010), incorporating gesture and gaze to influence multiparty conversation flow.

To facilitate turn-taking, provide clear instruction, and develop rapport, the SAR may use gestures and other nonverbal cues. Effective robot gaze could involve preceding linguistic references with a referential gaze, looking towards the listener at the end of a turn, and directing greetings towards the person (Huang and Mutlu 2012). Gestures may also play a critical role, as deictic gestures improve information recall (Huang and Mutlu 2013) and iconic gestures enhance rapport (Wilson et al. 2017). The robot's expression, encompassing emotions, mood, and attitudes, influences how people perceive the interaction, and simulating human-like emotions during conversations could foster robot-human social relationships (Kirby, Forlizzi, and Simmons 2010; Chuah and Yu 2021).

For robots to be effective assistants, they need to be able to adapt supportive behavior according to how and when the user needs help (Wilson, Aung, and Boucher 2022). However, there is a trade-off between directness and politeness (Goldsmith 2007). To pertain to the autonomy and individuality of the user without over-helping or under-helping, a SAR needs to adopt recognition for levels of need (Begum et al. 2013; Greczek et al. 2014; Wilson et al. 2018; Wilson, Tickle-Degnen, and Scheutz 2020). Accordingly, it begins with minimal assistance and escalates when necessary to increase directness and specification.

Moreover, behavior representation is essential for robust and efficient execution of complex tasks in unstructured environments (Nakawala et al. 2018). State machines may be used to represent sequences of actions a robot may perform (Bohren and Cousins 2010), but they are statically defined and lack flexibility. Behavior trees, which provide a modular and adaptable approach consisting of hierarchical descriptions of actions (Colledanchise and Ögren 2018), allow for the reuse of particular behaviors but do not provide a structure defining the social intent of the behaviors. We propose that extending our upper ontology affords a general approach across domains while providing novel integration of social and task-related behaviors.

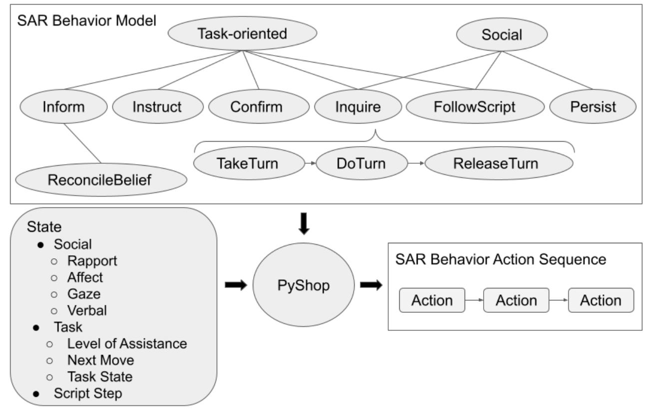

Figure 1: An upper ontology that defines SAR behaviors and specifications of input state are being processed through a Hierarchical Task Network (HTN) planner to output a sequence of actions.

Technical Approach

To enable autonomous behavior generation, we created an upper ontology of robot behaviors for guiding children with assembly tasks. The ontology defines the primary intents of a SAR, the structure of an assistive turn, and some of the key properties needed in constructing and selecting the appropriate behaviors.

Assistive Intents

We draw from the speech act theory (Searle 1975) to describe in this ontology the 8 primary intentions a SAR utilizes when assisting a user, categorized into task-oriented ones and socially-focused ones.

Task-oriented:

- **Confirm** - Confirmation on correctness of progress.
- **Instruct** - Indirect and direct support to correct mistakes.
- **Inform** - General directions on the way the task should be done.
- **Reconcile beliefs** - Help the user with a misunderstanding, incorrect assumption, or other false belief.
- **Inquire** - Inquiries that prompt verbal communication and increase mutual attentiveness. This may be used both to advance the task and to socially connect with the user.

Socially-oriented:

- **Social** - Common phrases that follow social courtesy.
- **Follow Script** - Following a scripted outline of a speech, such as in the introduction during the first encounter.
- **Persist** - Encouragement to show emotional support.

Assistive Turn

Structurally, the execution of an intent is composed of three parts: claiming the floor of conversation, holding the turn, and releasing the turn back to the floor. While claiming and releasing the floor are turn-taking elements that facilitate the flow of the interaction, actively holding the turn is the body of communication intent. Breaking down the three parts as sub-tasks of an objective, the ontology enables reusable assistive turns to be flexibly molded into assistive behaviors.

Furthermore, each body of intent has a variety of strategies the robot may exercise to express the desired intent. For example, direct support instruction can be accompanied by varying degrees of emotions and complexity of gestures.

Key Properties

Correspondingly, the ontology necessitates 8 input facts concerning the state of the task and of the participant:

- **Level of assistance** - How much assistance the user currently needs – this allows the selection of the assistive behavior to offer the proper amount of help.
- **Affect** - The emotional positivity of the subject.
- **Rapport** - When rapport is high, there is a high level of positivity, mutual attentiveness, and coordination (Tickle-Degnen and Rosenthal 1990).
- **Gaze** - Where the subject is directing their gaze: looking at the robot, the task, or other places in the room.
- **Verbal** - If the subject initiated verbal communication. For example, if the subject says “thank you” the robot may respond “you are welcome”.
- **Next move** - An inference regarding the next appropriate move. Such as “Spin the small blue triangle left”.
- **Task state** - The current state of progress: unstarted, in progress, has no error, just made a mistake, mistake escalated, nearly fixed, almost done (Görür et al. 2017).
- **Step** - The step strictly concerns the Follow Script intent, where this property specifies the current script being followed and the step in that script.

The initial description of key input properties acts as contingencies that navigate extensive variance of combinations for the same tasks. Such recognition is required by the ontology to generate context-dependent behaviors that are personalized, socially appropriate, and hopefully well-received.

Implementation

For the implementation, the upper ontology is manifested as a Hierarchical Task Network (HTN). Using the HTN and a planner, we generate behaviors for the robot to perform. An HTN characterizes domain knowledge as a network of compound and primitive tasks (Ilghami et al. 2005). Each compound task can be accomplished through a variety of methods. A method specifies premises of conditions for its activation; it also includes a list of sub-tasks that the planner recursive breaks down into a sequence of primitive tasks. A primitive task cannot be decomposed; it represents a particular executable action of the robot.

The major intents in the upper ontology are characterized as the highest compound tasks in the HTN. When behaviors are generated, the planner takes a top-level intent as the objective and generates a plan which is a sequence of actions. Because there are many ways a task can be done, the planner requires knowledge of certain facts in the input state to perform bindings with preconditions of different methods. The preconditions allow the SAR to act on recognition of the immediate situation.

The HTN affords scalability and re-usability in turn-taking implementation. The definition of a top-level compound task would begin with a ”take turn” sub-task and end with a ”release turn” sub-task. Turn-taking behaviors are lower compound tasks that also could be realized by adopting various methods. Because the execution of each intent is compartmentalized, this structure allows significant flexibility by enabling conversational turn-taking to adapt in a way that is independent of the content of communication.

Evaluation

As an initial validation of our approach, we compare a set of hand-coded behaviors to ones automatically generated from our HTN. We demonstrate the effectiveness of the hand-coded behaviors using a child-robot interaction experiment, then we show that the generated behaviors are able to cover the majority of those capacities while providing additional variation under various circumstances.

Hand-Coded Behaviors

We designed 70 hand-coded behaviors for a robot to perform while assisting young children in a puzzle task. They are based on the combination of six basic actions: SayText, SetEyes, LookInDirection, TiltHead, PointAt, and Pause. This set of operative modalities enables complex behaviors for the robot to introduce the task to the child, encourage the child, give suggestions and hints on the position and movement of the pieces, help the child find a missing piece, and congratulate the child upon completion. In the iterative development process of these capacities, we integrated feedback from developmental psychologists and results from testing on adults and children.

To demonstrate the effectiveness of the behavioral design, we conducted an experiment where a Misty Robot helped 5-8 year old children (N=32) to assemble tangram pieces into the shape of either a fox or rabbit. One piece of the puzzle was hidden, thus requiring each participant to seek assistance from the robot instructor.

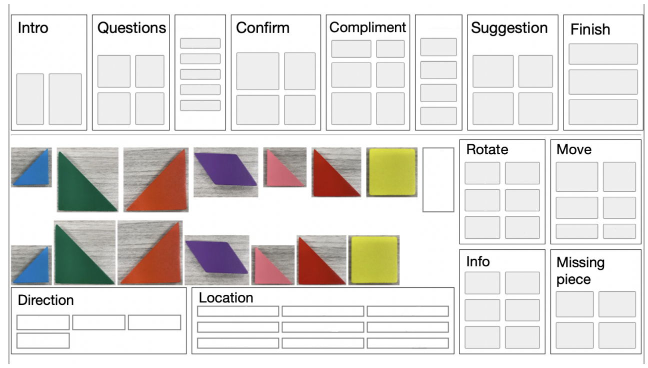

Figure 2: Interface to select the hand-coded behaviors.

The robot was controlled in a Wizard-of-Oz setup by a human observing from behind a one-way mirror. The robot controller used a web interface (see Figure 2) to select appropriate behaviors that the robot performs. After completing the task, the children answered a short survey about the helpfulness, likability, and trustworthiness of the robot.

32 participants sufficiently completed the study, the average time of completion was 3 minutes and 28 seconds. Average ratings for the robot on a 5-point Likert scale were 4.86 for helpfulness, 4.78 for smartness, 4.61 for trustworthiness, and 4.53 for interest to do the task again with the robot. The results indicate successful interactions with the majority of participants, therefore validating the hand-coded behaviors' capacity to provide effective and friendly assistance.

Generated Behaviors

The generated behaviors are encoded in an HTN with eight assistive intents as top-level tasks. Using a state that encapsulates the key properties, we then use PyShop (a variant of Pyhop (Nau 2013) that supports HDDL) to find a sequence of output actions (see Figure 1). They consist of the same types of basic actions as the hand-coded behaviors.

Behavior Comparison

We used the series of hand-coded behaviors as a blueprint of capacities to incorporate into the ontology. The comparison aims to measure how much of the hand-coded capacities the ontology can reproduce and validate. It assesses the extent of coverage by checking if the ontology can generate behaviors with the same action sequence or audio-visual similarity as the hand-coded model. Overall, the automated behaviors show similar abilities. The comparison is divided into the following categories:

- Exact match: The sequence of actions is identical to the original hand-coded behavior.
- Deletion: Behaviors we did not incorporate into the HTN because of their limited utility in the experiment.
- Split-string match: In the scope of this study, the HTN on occasion splits a SayText input string into multiple subtasks, leading to a short pause between phrases.

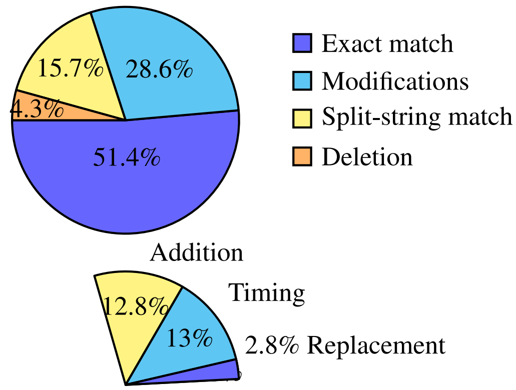

Figure 3: The ontology supports most of the hand-coded behaviors, with some minor modifications.

- Modification: The sequence of actions undergoes gestural adjustments for hypothetical improvements.
  - Timing: The content of actions is the same as the original, but timing adjustments are made with the intention of better delivery. Timing adjustments are adding, removing pauses, or re-ordering the sequence of actions.
  - Addition: An additional gestural movement added to the sequence of action, e.g. adding an extra arm movement to do a wave in greetings.
  - Replacement: Some gesture is removed and replaced with another.

As shown in Figure 3, the ontology covers 51.4% of capacities in the hand-coded behaviors by matching exact actions; 15.7% of the capacities are retained with the sole difference of a split in the string input for SayText action; 28.6% of capacities are slightly modified due to a gesture addition, removal, or different timing, which altered the delivery but retained the core semantics of the behavior. On top of that, 4.3% of the original behaviors were not retained due to their limited utility in the experiment.

Moreover, we added different methods and classified the sequence of the turn-taking actions, consequently, the ontology could generate 1500 variances on top of current capacities This allows the opportunity for adaptations in complex situations. Examples of typical differences in each of the 5 categories as well as in additional variations can be seen in the demonstration video[1].

Discussion

Through the comparison, we confirmed our hypothesis that the ontology is able to generate the necessary behaviors while also adding more variability to adapt to subtle differences in the situation.

Advantages

The goal of defining a hierarchical ontology is to support broad interoperability (Olivares-Alarcos et al. 2019). The ontology builds a framework that specifies the major categories of assistive intent and defines properties that the SAR's perceptive and cognitive components must provide to allow context-adaptive behavior generation.

Compartmentalizing turn-taking increases the adaptability of robot behavior, enabling robust turn-initiation and turn-release. The variance of methods during turn-holding, furthermore, increases the flexibility of the assistive behavior and appeals to children with diverse needs. The HTN processes contextual inputs and search for task bindings, and ultimately generates action sequences that align with social norms, facilitate interaction, and increase rapport.

Limitations

We recognize that our technical approach is based on a limited validation of behavior design. The hand-coded behaviors were shown through an empirical study to be successful, which provides a single measure of validity, yet lacks the ability to inform future developments. In the scope of this paper, we illustrate the ability to generate necessary behaviors through the demonstration of holistic coverage of the manual design. However, we have not determined the effectiveness of each individual behavior and how well the behavior has the intended effects on a child's affect, engagement, and rapport.

To further validate our approach we need to examine how the generated behaviors affect young children. The modifications, while generally minor, may have unanticipated effects on children, and thus an evaluation with young children will be necessary. Furthermore, we want to test the interoperability of the ontology by applying it to other domains, and by assisting children on different tasks.

Nevertheless, the theoretical basis of social and linguistic norms is widely applicable in social interactions independent of the type of agent and the task. The ontology-defined intents appeal to categories of existing speech act (Searle 1975), enabling know-what for SAR. Moreover, the ways in which intents are executed are based on contextualized variables of politeness, which enables know-how for the robot to mitigate face-threatening factors in its behaviors.

Conclusion

This paper presents an upper ontology for generating socially assistive robot (SAR) behaviors. The ontology's primary assistive intents enable flexible and scalable behavior generation, maximizing adaptability while using a combination of speech, gestures, and gaze. However, there is a lack of extensive research on the association between robot behaviors and impacts on child development (Langer, Marshall, and Levy-Tzedek 2023). Therefore, it is important that we proceed with caution in autonomous behavior generation to make sure they follow social norms and facilitate the learning experience. On the complementary, the ontology has flexibility to allow for easy integration of new findings in SAR behavior. Thus, it provides an avenue for SARs that could potentially enhance the educational experience of young children.

[1] https://youtu.be/sEvmIufoHH0

References

Alcorn, A. M.; Ainger, E.; Charisi, V.; Mantinioti, S.; Petrović, S.; Schadenberg, B. R.; Tavassoli, T.; and Pellicano, E. 2019. Educators' views on using humanoid robots with autistic learners in special education settings in England. *Frontiers in Robotics and AI*, 6: 107.

Begum, M.; Wang, R.; Huq, R.; and Mihailidis, A. 2013. Performance of daily activities by older adults with dementia: The role of an assistive robot. In *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*, 1–8. IEEE.

Belpaeme, T.; Kennedy, J.; Ramachandran, A.; Scassellati, B.; and Tanaka, F. 2018. Social robots for education: A review. *Science Robotics*, 3(21): eaat5954.

Bohren, J.; and Cousins, S. 2010. The smach high-level executive [ros news]. *IEEE Robotics & Automation Magazine*, 17(4): 18–20.

Bohus, D.; and Horvitz, E. 2010. Facilitating multiparty dialog with gaze, gesture, and speech. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*, 1–8.

Chuah, S. H.-W.; and Yu, J. 2021. The future of service: The power of emotion in human-robot interaction. *Journal of Retailing and Consumer Services*, 61: 102551.

Colledanchise, M.; and Ögren, P. 2018. *Behavior trees in robotics and AI: An introduction*. CRC Press.

Duncan, S. 1972. Some signals and rules for taking speaking turns in conversations. *Journal of personality and social psychology*, 23(2): 283.

Goldsmith, D. J. 2007. Brown and Levinson's politeness theory. *Explaining communication: Contemporary theories and exemplars*, 219–236.

Görür, O. C.; Rosman, B. S.; Hoffman, G.; and Albayrak, S. 2017. Toward integrating theory of mind into adaptive decision-making of social robots to understand human intention. In *ACM/IEEE International Conference on Human-Robot Interaction (HRI) Workshop on the Role of Intentions in HRI*. IEEE.

Greczek, J.; Kaszubski, E.; Atrash, A.; and Matarić, M. 2014. Graded cueing feedback in robot-mediated imitation practice for children with autism spectrum disorders. In *The 23rd IEEE international symposium on robot and human interactive communication*, 561–566. IEEE.

Holler, J.; and Kendrick, K. H. 2015. Unaddressed participants' gaze in multi-person interaction: optimizing recipiency. *Frontiers in psychology*, 6(98): 1–14.

Huang, C.-M.; and Mutlu, B. 2012. Robot behavior toolkit: generating effective social behaviors for robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 25–32.

Huang, C.-M.; and Mutlu, B. 2013. Modeling and Evaluating Narrative Gestures for Humanlike Robots. In *Proceedings of Robotics: Science and Systems*. Berlin, Germany.

Ilghami, O.; Nau, D.; Muñoz-Avila, H.; and Aha, W. 2005. Learning Preconditions for Planning from Plan Traces and HTN Structure. *Computational Intelligence*, 21: 388–413.

Kim, E. S.; Berkovits, L. D.; Bernier, E. P.; Leyzberg, D.; Shic, F.; Paul, R.; and Scassellati, B. 2013. Social robots as embedded reinforcers of social behavior in children with autism. *Journal of autism and developmental disorders*, 43: 1038–1049.

Kirby, R.; Forlizzi, J.; and Simmons, R. 2010. Affective social robots. *Robotics and Autonomous Systems*, 58(3): 322–332.

Langer, A.; Marshall, P. J.; and Levy-Tzedek, S. 2023. Ethical considerations in child-robot interactions. *Neuroscience & Biobehavioral Reviews*, 151: 105230.

Nakawala, H.; Goncalves, P. J.; Fiorini, P.; Ferringo, G.; and De Momi, E. 2018. Approaches for action sequence representation in robotics: a review. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 5666–5671. IEEE.

Nau, D. 2013. Game applications of HTN planning with state variables. In *Planning in Games: Papers from the ICAPS Workshop*.

Nomura, T.; and Kanda, T. 2016. Rapport–expectation with a robot scale. *International Journal of Social Robotics*, 8: 21–30.

Olivares-Alarcos, A.; Beßler, D.; Khamis, A.; Goncalves, P.; Habib, M. K.; Bermejo-Alonso, J.; Barreto, M.; Diab, M.; Rosell, J.; Quintas, J.; and et al. 2019. A review and comparison of ontology-based approaches to robot autonomy. *The Knowledge Engineering Review*, 34: e29.

Sacks, H.; Schegloff, E. A.; and Jefferson, G. 1978. A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction*, 7–55. Elsevier.

Searle, J. R. 1975. Indirect speech acts. In *Speech acts*, 59–82. Brill.

Seo, S. H.; Griffin, K.; Young, J. E.; Bunt, A.; Prentice, S.; and Loureiro-Rodríguez, V. 2018. Investigating people's rapport building and hindering behaviors when working with a collaborative robot. *International Journal of Social Robotics*, 10: 147–161.

Tickle-Degnen, L.; and Rosenthal, R. 1990. The nature of rapport and its nonverbal correlates. *Psychological inquiry*, 1(4): 285–293.

van den Berghe, R.; Verhagen, J.; Oudgenoeg-Paz, O.; van der Ven, S.; and Leseman, P. 2019. Social Robots for Language Learning: A Review. *Review of Educational Research*, 89(2): 259–295.

Wilson, J.; Lee, N.; Saechao, A.; Hershenson, S.; Scheutz, M.; and Tickle-Degnen, L. 2017. Hand Gestures and Verbal Acknowledgments Improve Human-Robot Rapport. 334–344. ISBN 978-3-319-70021-2.

Wilson, J. R.; Aung, P. T.; and Boucher, I. 2022. When to Help? A Multimodal Architecture for Recognizing When a User Needs Help from a Social Robot. In *International Conference on Social Robotics*, 253–266. Springer.

Wilson, J. R.; Lee, N. Y.; Saechao, A.; Tickle-Degnen, L.; and Scheutz, M. 2018. Supporting human autonomy in a robot-assisted medication sorting task. *International Journal of Social Robotics*, 10: 621–641.

Wilson, J. R.; Tickle-Degnen, L.; and Scheutz, M. 2020. Challenges in designing a fully autonomous socially assistive robot for people with parkinson's disease. *ACM Transactions on Human-Robot Interaction (THRI)*, 9(3): 1–31.