Interactive Robot Training for Temporal Tasks

Ankit Shah
ajshah@mit.edu
MIT
Cambridge, MA, USA

Julie Shah
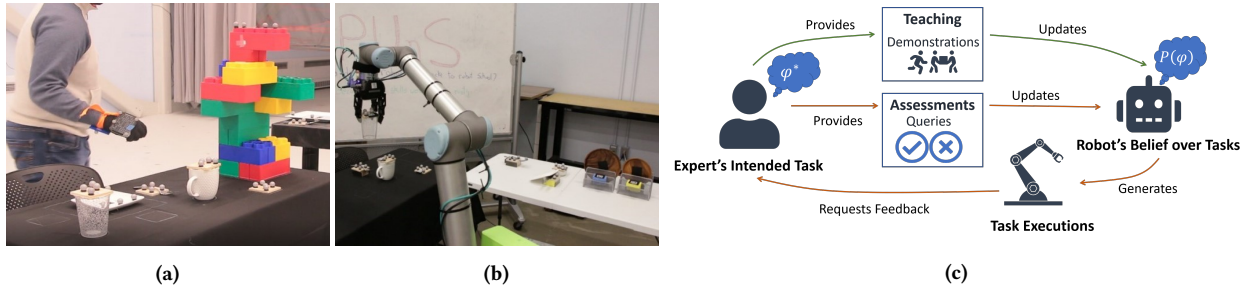julie_a_shah@csail.mit.edu
MIT
Cambridge, MA, USA

Figure 1: Figures 1a and 1b depicts the table-setting task as demonstrated by the human and performed by the robot respectively. Figure 1c depicts our active learning framework for unifying demonstrations and query assessments.

KEYWORDS

Learning from Demonstrations, Active learning

1 INTRODUCTION AND RELATED WORK

Imagine a future where a domestic robot ships with a state-of-the-art learning from demonstrations (LfD) system to learn household tasks. You would like the robot to set the dinner-table for you when you get home at dinner time. After you demonstrate how to set the dinner table a couple of times. Would you be confident that robot will not try to place the saucer on top of the cup, or finish as much of the task as possible if an object was missing?

Formal languages like linear temporal logic (LTL) are ideal for specifying temporal properties pertaining to the dinner setting task due to their unambiguous semantics and expressivity. However, formal languages are an unwieldy tool for an untrained user. This has lead to an interest in developing models to infer formal specifications from data provided through intuitive modalities. For example, there exist models to infer LTL formulas from demonstrations [16],[11], natural language instructions [15],[8], and preferences elicited from multiple experts [10]. However these modalities are inherently ambiguous and potentially mutually contradictory. Often, there is no 'correct' answer, and the task specification is best

described as a belief over multiple LTL formulas $P(\varphi)$. Our proposed approach is to decompose the problem of training a robot into inferring the specification [16] as a belief over candidate LTL formulas, followed by computing policies that best satisfy this belief [17]. Further, I propose an interactive training framework (Figure 1c) that allows the user to begin the robot training by providing demonstrations, and later provide feedback on the acceptability of task executions performed by the robot.

Argall et al. [1] and Chernova et al. [6] provide a comprehensive survey of methods for robot learning from demonstrations (LfD). Prior research into specification inference from demonstrations [9, 12, 20, 21] generate a point estimate of the specifications; this makes them unsuitable in scenario when multiple hypotheses are equally valid. Our prior work on Bayesian specification inference [16] computes the belief over candidate LTL formulas thus providing an estimate of the model's (un)certainty.

Specification inference from a batch of demonstrations yields an initial belief over specifications, however active learning can guide further exploration to reduce the uncertainty of the distribution. Cakmak et al. [3, 4] developed a taxonomy of types of queries that can be asked to obtain a better understanding of task specifications. Biyik et al. [2, 2] have designed inference models that learn the reward function for a task based on queries that elicit the user's preference over multiple trajectories. However, the prior work has largely focused on tasks with a Markovian reward function. Temporal tasks that we consider do not satisfy the Markov assumption as they require the entire state history to evaluate acceptability. Our proposed research extends active learning to tasks that can be described by formulas belonging to a well-defined segment of LTL.

Adopting LTL specifications as goal specifications would require abandoning the Markov assumption. Algorithms for planning with goals defined through LTL formulas have been proposed by symbolic planning [5], hybrid controller synthesis [13] and reinforcement learning [14, 19]. However these algorithms are not applicable to tasks where the specifications are expressed as a beliefs over
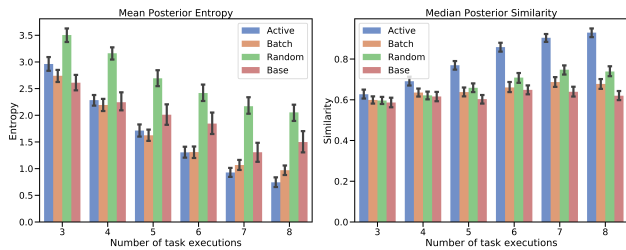
Figure 2: The average entropy (left; lower is better) of the final posterior and the similarity of the posterior with the ground truth formula (right; higher is better) for the four training conditions. All error bars indicate 95% confidence interval.

LTL formulas. Our prior work [17] addresses this by developing algorithm to compile planning with uncertain LTL specifications into an equivalent MDP formulation.

2 BAYESIAN SPECIFICATION INFERENCE

As a part of our prior research [16], we developed a probabilistic model, inspired by Bayesian concept learning [18], to infer specifications given task demonstrations. In context of the table-setting task, given a batch of demonstrations for setting a table $\{D\}$ (Figure 1a), which objects need to be placed, and what are the allowable placement orders? A single hypothesis for such specification can be represented as an LTL formula φ and the robot's belief over the specification can be expressed as a conditional posterior $P(\varphi \mid \{D\})$. We defined priors over a concept class of candidate LTL formulas selected from a library of temporal properties defined by Dwyer et al. [7]. Further, we proposed a domain independent approximation to the likelihood function that satisfies the size principle [18] (the likelihood should favor the most restrictive hypothesis that explains the observed data). We used a sampling-based algorithm to compute an approximation to $P(\varphi \mid \{D\})$. We demonstrated the efficacy of our model by inferring specifications for setting a dinner table using demonstrations of people performing the task. The true specification was one of the top-5 most likely formulas with as few as 30 demonstrations, while with all 71 demonstrations, it was the most likely formula.

3 PLANNING WITH UNCERTAIN SPECIFICATIONS (PUNS)

Next, we considered the problem of computing the robot policy that best adheres to a belief over the true specification $P(\varphi \mid \{D\})$ instead of a single formula φ^*. We first proposed four evaluation criteria that captured the semantics of satisfying a distribution over LTL formulas. We demonstrated that each of these evaluation criteria is equivalent to optimizing a non-Markov reward function. Next we developed algorithms to construct a minimal finite state machine (FSM) that simultaneously tracks the (dis)satisfaction of all the formulas in $P(\varphi \mid \{D\})$. The reformulation as an MDP allows the agent to use a suite of popular reinforcement learning algorithms to compute an optimal policy for the learner. As depicted in Figure 1b, a robot trained using the PUnS formulation for the table setting task was able to reliably set the dinner table with a very small error rate ($\sim 10^{-4}$ as estimated through simulations).

4 PROPOSED RESEARCH

Bayesian specification inference and PUnS can be used in conjunction with each other to allow a robot to learn a task given a batch of demonstration data. However, the initial belief $P(\varphi)$ and the minimal FSM constructed for PUnS allow the learner to identify a task execution whose acceptability is the most uncertain. This execution can then be demonstrated back to the user to ascertain whether it acceptable or unacceptable (Q represents the query and the user's assessment). Given Q, the learner can once again apply Bayesian specification inference; this time using $P(\varphi)$ as the prior to compute a posterior $P(\varphi \mid Q)$ in an active learning setting.

In preliminary experiments, we compared the performance of this approach (termed 'Active'), interactions with random query generation (termed 'Random') and a heuristic query generation baseline (termed 'Base') based on modifying only the most likely formula in $P(\varphi)$. We also compared it against belief computed by directly providing the robot with just an equal number of demonstrations (termed 'Batch'). We conducted 200 simulated trials. For each trial, we sampled a ground truth formula and automatically generated demonstrations that satisfied it. The three query based models were initialized with $P(\varphi)$ generated using specification inference conditioned on two demonstrations. This was then followed by sequentially updating the belief as per the responses provided by the ground truth based oracle to the generated queries to yield $P_{final}(\varphi)$. In the 'Batch' case, the final posterior was generated conditioned on the same number of demonstrations as the total number data points in the query based cases. We compared the Entropy of final distribution indicating the confidence of the model, and the expected value of the similarity metric measured as an intersection-over-union of the clauses included in the candidate formula compared to the ground truth.

The results of the simulation experiment (Figure 2) indicate that the 'Batch' and the 'Active' approach had the lowest entropy values compared to 'Base' and 'Random' indicating high confidence distributions. However the higher similarity score for the 'Active' approach indicates that it demonstrates better convergence to the true specification compared to 'Batch'. While the preliminary results are promising, we would like to repeat this experiment as a human-participant study where the participant provides both demonstrations and assessments. Besides the objective metrics evaluated here, we would also like to assess subjective measures for how well the participant's perception of the robot's competence aligns with the objective metrics.

5 CONCLUSION

In conclusion we propose an active learning framework that accepts user's input through both demonstrations and query assessments for temporal tasks. We hypothesize that initializing the robot's belief through demonstrations followed by refining it through demonstrating the most informative query would result in lower uncertainty in the resulting belief and better alignment with the ground truth specification. Our preliminary simulation experiments support his hypothesis, and we propose to conduct human-participant studies to evaluate the alignment of participants' subjective perceptions with objective measures.

REFERENCES

[1] Brenna D Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. 2009. A survey of robot learning from demonstration. *Robotics and autonomous systems* 57, 5 (2009), 469–483.

[2] Erdem Bıyık and Dorsa Sadigh. 2018. Batch active preference-based learning of reward functions. *arXiv preprint arXiv:1810.04303* (2018).

[3] Maya Cakmak, Crystal Chao, and Andrea L Thomaz. 2010. Designing interactions for robot active learners. *IEEE Transactions on Autonomous Mental Development* 2, 2 (2010), 108–118.

[4] Maya Cakmak and Andrea L Thomaz. 2012. Designing robot learners that ask good questions. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction.* ACM, 17–24.

[5] Alberto Camacho, Jorge A Baier, Christian Muise, and Sheila A McIlraith. 2018. Finite LTL synthesis as planning. In *Twenty-Eighth International Conference on Automated Planning and Scheduling.*

[6] Sonia Chernova and Andrea L Thomaz. 2014. Robot learning from human teachers. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 8, 3 (2014), 1–121.

[7] Matthew B Dwyer, George S Avrunin, and James C Corbett. 1999. Patterns in property specifications for finite-state verification. In *Proceedings of the 21st international conference on Software engineering.* ACM, 411–420.

[8] Nakul Gopalan, Dilip Arumugam, Lawson LS Wong, and Stefanie Tellex. 2018. Sequence-to-Sequence Language Grounding of Non-Markovian Task Specifications.. In *Robotics: Science and Systems.*

[9] Bradley Hayes and Brian Scassellati. 2016. Autonomously Constructing Hierarchical Task Networks for Planning and Human-Robot Collaboration. In *International Conference on Robotics and Automation (ICRA).* IEEE.

[10] Joseph Kim, Christopher J Banks, and Julie A Shah. 2017. Collaborative Planning with Encoding of Users' High-Level Strategies.. In *AAAI.* 955–962.

[11] Joseph Kim, Christian Muise, Ankit Shah, Shubham Agarwal, and Julie Shah. 2019. Bayesian Inference of Linear Temporal Logic Specifications for Contrastive

[12] Explanations. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19.* https://doi.org/10.24963/ijcai.2019/776

[12] Zhaodan Kong, Austin Jones, Ana Medina Ayala, Ebru Aydin Gol, and Calin Belta. 2014. Temporal logic inference for classification and prediction from data. In *Proceedings of the 17th international conference on Hybrid systems: computation and control.* ACM, 273–282.

[13] Hadas Kress-Gazit, Georgios E Fainekos, and George J Pappas. 2009. Temporal-logic-based reactive mission and motion planning. *IEEE transactions on robotics* 25, 6 (2009), 1370–1381.

[14] Michael L Littman, Ufuk Topcu, Jie Fu, Charles Isbell, Min Wen, and James MacGlashan. 2017. Environment-independent task specifications via GLTL. *arXiv preprint arXiv:1704.04341* (2017).

[15] Yoonseon Oh, Roma Patel, Thao Nguyen, Baichuan Huang, Ellie Pavlick, and Stefanie Tellex. 2019. Planning with State Abstractions for Non-Markovian Task Specifications. In *RSS.* https://doi.org/10.15607/RSS.2019.XV.059

[16] Ankit Shah, Pritish Kamath, Julie A Shah, and Shen Li. 2018. Bayesian Inference of Temporal Task Specifications from Demonstrations. In *Advances in Neural Information Processing Systems 31.*

[17] Ankit Shah, Shen Li, and Julie Shah. 2019. Planning With Uncertain Specifications (PUnS). *arXiv preprint arXiv:1906.03218* (2019).

[18] Joshua B Tenenbaum. 2000. Rules and similarity in concept learning. In *Advances in neural information processing systems.* 59–65.

[19] Rodrigo Toro Icarte, Toryn Q Klassen, Richard Valenzano, and Sheila A McIlraith. 2018. Teaching multiple tasks to an RL agent using LTL. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems.* International Foundation for Autonomous Agents and Multiagent Systems, 452–461.

[20] Marcell Vazquez-Chanlatte, Susmit Jha, Ashish Tiwari, Mark K Ho, and Sanjit Seshia. 2018. Learning Task Specifications from Demonstrations. In *Advances in Neural Information Processing Systems.* 5372–5382.

[21] Chanyeol Yoo and Calin Belta. 2017. Rich Time Series Classification Using Temporal Logic. In *Robotics: Science and Systems.*