Learning Manner of Execution from Partial Corrections

Extended Abstract

Mattias Appelgren
University of Edinburgh
United Kingdom
mattias.appegren@ed.ac.uk

Alex Lascarides
Unviersity of Edinburgh
United Kingdom
alex@inf.ed.ac.uk

ABSTRACT

Some actions must be executed in different ways depending on the context. Wiping away marker requires vigorous force while almonds require gentle force. We provide a model where an agent learns which manner to execute in which context, drawing on evidence from trial and error and verbal corrections when it makes a mistake (e.g., "no, do it gently"). The learner's initial domain model lacks the concepts denoted by the words in the teacher's feedback: both those describing the context (e.g., almonds) and those describing manner (e.g., gently). We show that discourse coherence helps the agent refine its domain model and perform the symbol grounding that's necessary for using the guidance to solve its planning problem: to perform its actions in the current context in the correct way.

KEYWORDS

Interactive task learning, Symbol grounding, Knowledge representation and reasoning

ACM Reference Format:
Mattias Appelgren and Alex Lascarides. 2023. Learning Manner of Execution from Partial Corrections: Extended Abstract. In *Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), London, United Kingdom, May 29 – June 2, 2023*, IFAAMAS, 3 pages.

1 INTRODUCTION

Some actions are similar in many ways but also call for some differences in execution, dependent on context. For instance, the goal of wiping and the bulk of its movement are the same across different substances, but wiping away marker requires pressing hard and moving quickly, while wiping almonds requires moving slowly and gently. We address the task of learning in which contexts to perform which manner of action using the Interactive Task Learning paradigm (ITL, [3, 6]): i.e., the agent learns incrementally from trial and error and teacher feedback on its actions.

As this extended embodied interaction proceeds, our model supports online incremental learning of two things. The first is symbol grounding: i.e., learning to map the teacher's words, including adverbs like "gently", to their referents in the embodied environment. The second is a domain-level policy for choosing which manner to perform in the current state. Following other works in ITL, we are interested in developing learners that can cope with unforeseen changes to their environment after deployment. Accordingly, the learner's initial domain model lacks concepts that are critical to success—in the wiping example, that's analogous to the learner's domain model lacking almonds, let alone the links from almonds to their desired wiping manner. This makes symbol grounding more challenging. The learner must not only use the latest observations to update its probabilistic mappings from symbols to referents in the visual scene, but also estimate whether it needs to expand the set of possible domain states by adding a new category to it.

The task the agent faces in our experiments is an abstract version of the wiping example. The results show that exploiting coherence constraints on interpretation makes learning more data efficient.

Proc. of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2023), A. Ricci, W. Yeoh, N. Agmon, B. An (eds.), May 29 – June 2, 2023, London, United Kingdom.

2 EXPLOITING DISCOURSE COHERENCE

The agent must learn a set of rules of the form *concept* → *behaviour* (e.g., *concept* is almonds, and *behaviour* is gently), learn to recognise the *concept* from visusal features, and learn to generate the *behaviour*. Our basic claim is that discourse coherence aids this learning. Discourse coherence models constrains the meaning of an utterance via its semantic connection to its context [2, 4], including its embodied context [5]. We will assume that the teacher's coherent feedback is semantically connected to the learner's latest action. With this in mind, consider what the agent can learn from the following three types of teacher feedback (allowing terms to be replaced as appropriate):

(1) yes
(2) no, wipe almonds gently and slowly
(3) no, do it gently and slowly.

The teacher saying "yes" can strengthen the agent's belief that its action was correct. Following Lascarides and Stone [7], when uttering a correction (ie., (2) or (3)), the teacher must be denying that something the agent did was correct. For utterance (2), there are several (not mutually exclusive) sources of error:

(i) the substance was almonds, but the agent didn't believe this.
(ii) Rules *almonds* → *gently* and *almonds* → *slowly* are both true, but the agent didn't believe at least one of these.
(iii) the wiping wasn't gentle and slow, even if the agent believed it was.

For utterance (3), the potential sources of error (i) and (ii) are underspecified with respect to the substance. Either way, on observing a correction, the agent must make an inference to estimate which of these types of errors is the case.

3 EXPERIMENTS

Our experiments incorporate an abstract task that is analogous to the wiping example: the concepts on which desired manners depend are shape (square, ellipse, heart) and/or colour categories (of many hues for each category), and the manner is a point on a bezier curve. Shape is observed, but the agent must learn colour categories, and starts out unaware of which colour categories exist (ie, it does not

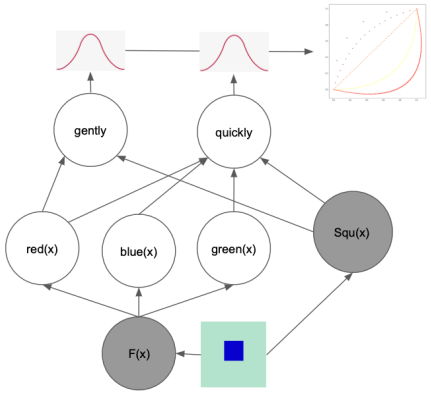

Figure 1: The graphical component of the agent's Bayes Net when in prior moves the teacher said "when you see red squares do it gently", has also mentioned "green" and "blue", and the teacher now says "no, do it quickly". Grey nodes are observed; white nodes are latent.

Dataset	Random	Just No	Full
Partial Corrections	31 ± 18	16 ± 8	11 ± 4
Fully Expressed	16 ± 11	10 ± 5	5 ± 2

Table 1: Mean and standard deviation of terminal regret.

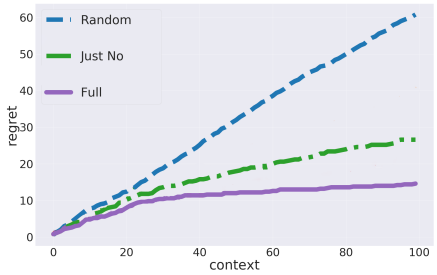

Figure 2: Average cumulative regret for Partial Corrections.

know how to partition the RGB values, let alone the labels for the partitions). It also doesn't know which points on the bezier curve correspond to which manner.

Figure 1 illustrates the agent's probabilistic model. It supports decisions based on the three considerations mentioned earlier. I.e. it incorporates estimates of what objects are in the visual scene; it has beliefs about rules that link concepts to desired behaviour; and it generates behaviour given these prior two beliefs. The model is updated incrementally on any of these three counts. It treats a rule of the form $concept \rightarrow behaviour$ as a conditional probability, $P(behaviour|concept)$, that captures the strength of belief in the rule. It treats identifying concepts in a similar way, via $P(concept|F(s))$, where $F(s)$ are the observable visual features of situation s. Whenever the teacher utters a neologism (e.g., utters "red" for the first time), a new Boolean chance node is added to the Bayes Net, with dependencies dependent on the constraint the teacher expressed. A full description of all the components of this model appears in [1].

Each experiment is made up of five trials. Each trial corresponds to a different set of mutually consistent ground-truth rules: e.g., we forbid $red \rightarrow slowly$ and $red \wedge square \rightarrow quickly$. Each trial consists of 100 different situations, which are chosen so that the ground-truth rules don't conflict: e.g., if $red \rightarrow gently$ and $square \rightarrow firmly$ then there are no red squares. Generating a situation begins by selecting one of eight major colour categories, then a random hue of that category, then a random shape. We make 90% of the 100 situations feature a shape and/or colour that features in a ground truth rule.

The agent observes each situation in turn and observes what shape is depicted and its RGB value. Given these observations, it selects a point in behaviour space which is used to generate a behaviour curve. The teacher observes the generated behaviour and gives feedback as described earlier. The agent then updates its models of symbol grounding, action selection and beliefs about the ground-truth rules. We run two experiments: in Fully Expressed the teacher makes all corrections like utterance (2), and in Partial Corrections the teacher chooses uniformly at random whether to make the correction partial (i.e., like utterance (3)) or fully expressed (like utterance (2)).

We compare our Full Agent against two baselines: the *Random Baseline* simply selects a point in behaviour space uniformly at random and doesn't attempt to learn; and the *Just No* agent learns only from the teacher saying "yes" or "no". To evaluate an agent we measure *terminal regret* (Table 1): the number of mistakes it makes over the 100 situations. We also look at the curvature of *cumulative regret* as it experiences the situations (Figure 2 for the Partial Corrections dataset; the Fully Expressed dataset has similar curves). The t-test on terminal regrets shows that the Full agent outperforms the "Just No" agent ($t = 7.91$ $p = 2.43e^{-5}$), who outperforms the Random agent ($t = 9.25$ $p = 6.82e^{-6}$). In [1] we also provide results for two ablation studies, both of which are also outperformed by our Full agent: one in which the agent does not update when the teacher says "yes"; and one in which the agent does not use negative exemplars to update its models for generating behaviours.

4 CONCLUSION

We tackled the task of an agent learning the way it should perform an action depending on its context, with evidence coming from an extended embodied interaction with a teacher. The agent starts ignorant of how words denoting manner map to particular features of movement, unaware of domain-level concepts in which the constraints on manner are expressed, and ignorant of those constraints. Our experiments support our hypothesis that learning from contentful corrections is more data efficient than learning form just yes/no feedback, even though the additional content requires learning refinements to the domain model and symbol grounding. Our experiments, while analogous to examples like wiping, addressed a highly abstract task. In future work, we plan to test our models on more realistic situations.

Acknowledgements: This work was supported by the UKRI-funded TAS Governance Node (grant number EP/V026607/1).

REFERENCES

[1] Mattias Appelgren and Alex Lascarides. 2023. Learning Manner of Execution from Partial Corrections. https://doi.org/10.48550/ARXIV.2302.03338

[2] Nicholas Asher and Alex Lascarides. 2003. *Logics of conversation*. Cambridge University Press.

[3] Kevin A Gluck and John E Laird. 2019. *Interactive Task Learning: Humans, Robots, and Agents Acquiring New Tasks through Natural Interactions*. MIT press, London.

[4] Jerry R. Hobbs. 1979. Coherence and Coreference. *Cognitive Science* 3, 1 (1979), 67–90.

[5] Julia Hunter, Nicholas Asher, and Alex Lascarides. 2018. A Formal Semantics for Situated Conversation. *Semantics and Pragmatics* 11 (2018). https://doi.org/10.3765/sp.11.10

[6] John E. Laird, Kevin A. Gluck, John R. Anderson, Kenneth D. Forbus, Odest Chadwicke Jenkins, Christian Lebiere, Dario D. Salvucci, Matthias Scheutz, Andrea Lockerd Thomaz, J. Gregory Trafton, Robert E. Wray, Shiwali Mohan, and James R. Kirk. 2017. Interactive Task Learning. *IEEE Intelligent Systems* 32 (2017), 6–21.

[7] Alex Lascarides and Matthew Stone. 2009. A Formal Semantic Analysis of Gesture. *Journal of Semantics* 26, 4 (2009), 393–449.