Socially-Conditioned Task Reasoning for a Virtual Tutoring Agent

Zian Zhao
Shanghai Jiao Tong University
Shanghai, SH
zi.an.zhao@sjtu.edu.cn

Michael Madaio
Carnegie Mellon University
Pittsburgh, PA
mmadaio@andrew.cmu.edu

Yoichi Matsuyama
Carnegie Mellon University
Pittsburgh, PA
yoichim@andrew.cmu.edu

Florian Pecune
Carnegie Mellon University
Pittsburgh, PA
fpecune@andrew.cmu.edu

Justine Cassell
Carnegie Mellon University
Pittsburgh, PA
justine@cs.cmu.edu

ABSTRACT
Virtual agents have been shown to be more effective when incorporating social factors such as trust into task action selection. However, there has been less work on how virtual tutoring agents can incorporate social factors into pedagogical action selection. We propose and evaluate how a socially-conditioned task reasoner for a virtual pedagogical agent can incorporate both task and social factors into task reasoning. Our work contributes to the autonomous agent community by providing further evidence that incorporating information about dyadic social factors (e.g. rapport) can be beneficial for agents’ task reasoning in the case of a tutoring agent.

KEYWORDS
Socially-aware; pedagogical agent; rapport; RL-LSTM

ACM Reference Format:

1 INTRODUCTION
With the increasing ubiquity of virtual agents in a wide variety of domains, it is becoming ever more critical for those agents to incorporate awareness of social dimensions of interactions [13]. While prior work has investigated the role that social factors like trust play in interactions between humans and agents [3, 7, 14], and modeled the human-agent relationship to inform the agent’s selection of high-level tasks [4], there has been less work on using that social dimension to inform the selection of low-level task-related dialogue moves, particularly for tutoring agents.

Over the last several decades, intelligent tutoring systems have been developed to support student learning by providing adaptive problem-selection, step-level instructions, hints or feedback [16]. More recently, researchers have identified optimal pedagogical strategies through reinforcement learning [10, 11]. Prior research suggests that the best teachers build interpersonal closeness with their students, and change their teaching style accordingly [5]. However, while some work has studied virtual tutors’ action selection based on students’ individual affect [1, 12], those prior work didn’t take into account dyadic social states, such as rapport.

In this paper, we investigate how an awareness of dyadic interpersonal social closeness (or, ‘rapport’) and the social conversational strategies that may build that rapport can be incorporated into and improve an agent’s task reasoning. Specifically, we focus on two research questions: (1) Will an agent’s pedagogical task selection be more effective for learning when incorporating the rapport and social behaviors between tutor and tutee? (2) Will using the task and social interaction history instead of only the current observation improve the performance of a pedagogical reasoner?

2 MODEL CONSTRUCTION
The task of a socially-conditioned pedagogical reasoner is as follows: at each time step, the pedagogical reasoner will choose the optimal tutoring strategy from a set of available strategies, to maximize tutees’ learning gains over time (normalized gain between pre and post-test [11]), given data of (a) tutees’ previous problem-solving performance; (b) the tutees’ and tutoring agent’s prior labeled utterances; and (c) the rapport level between tutee and agent [8].

We use a peer tutoring dialogue corpus (described in [9]) of 22 pairs of 12-15 year old students tutoring each other in algebra, for a total of 30 hours of multimodal data. The transcripts were annotated for tutoring and learning strategies (TS) (e.g. propose step/answer, ask for help, etc) and two social conversational strategies (SS) (e.g. self-disclosure, praise, etc), described in [8, 9]). Each tutoring session was scored for the problems solved during the session and the normalized learning gain from pre-test to post-test (as in [11]). To find the rapport level for each 30-second “thin-slice” of our corpus, we used a crowd-sourcing approach, described in [8, 17].

To combine social factors (rapport and SS) and task factors (learning performance and TS) into pedagogical reasoning, we model the process into a reinforcement learning problem as follows: States are defined as the TS or SS the tutee has just used (e.g. propose step), concatenated with rapport level, if included. Actions are the set of TS available for the tutoring agent to choose from, used to tutor the tutees in algebraic problem-solving (e.g. provide hint, feedback,
prompt for step). The **Reward function** assigns rewards according to learners’ performance (1 for problem solved, 10 for positive, and -10 for negative learning gains at the end of the session). The **Environment** is modeled by a Markov Decision Process (MDP) with **Discount coefficient** set to 0.98 to ensure the tutoring sessions’ duration is realistic. We will refer to this model as an MDP reasoner.

To investigate RQ2, we replace the Environment with a long-short term memory (LSTM) to model possible long-term dependencies in pedagogical reasoning. We refer to this as an RL-LSTM reasoner. In order to find the optimal policy, we train the network to approximate the Q value, following the method in [2]. Our architecture contains a layer of 200 LSTM units, followed by a single hidden layer of 100 hidden units, with the input layer corresponding to the dimensions of the concatenated input vector. The input of the model is a concatenated vector:

\[ < S_{t-1}, A_{t-1}, S_t, R_{pt} > \]

which is the tutee’s action \( S_{t-1} \) and agent’s action \( A_{t-1} \) at the previous time step; the current tutee’s action \( S_t \); and the current rapport level (e.g., low, medium, high) \( R_{pt} \); each one-hot encoded. The output layer is the same length as the set of available task strategy actions for the virtual agent. The output layer is activated by the tanh function and other layers are activated by ReLU function. At each time step, the network selects the action with the largest expected reward to be the agent’s next pedagogical action. To compare performance across multiple models (i.e., MDP and RL-LSTM), we evaluate the models by comparing them to the ground truth expectation in a deployed system. Further, tutors often teach differently according to their individual tutees, while our model presents a general approach to action selection, given the rapport level. A personalized model may thus be more effective. For future studies, given the flexibility of RL-LSTM models, we also intend to incorporate more features, such as nonverbal and acoustic features, and study their effect on pedagogical reasoning or broader applications for socially-conditioned task reasoning in virtual agents.

We intend for this article to contribute to the autonomous agent community for its study on incorporating social factors into task reasoning, and intend for it to contribute to the design of virtual agents that can interact with people on more human terms.

**ACKNOWLEDGMENTS**

The work is supported by NSF Award No.1523162, IES Grant R305B150008 to CMU, and the IT R&D program of MSIP/IITP [2017-0-00255].
REFERENCES