Personalized Robot Tutoring using the Assistive Tutor POMDP (AT-POMDP)

Aditi Ramachandran*, Sarah Strohkorb Sebo*, Brian Scassellati
Computer Science, Yale University
New Haven, Connecticut, USA
{aditi.ramachandran, sarah.sebo, brian.scassellati}@yale.edu
*equal contribution

Abstract

Selecting appropriate tutoring help actions that account for both a student's content mastery and engagement level is essential for effective human tutors, indicating the critical need for these skills in autonomous tutors. In this work, we formulate the robot-student tutoring help action selection problem as the Assistive Tutor partially observable Markov decision process (AT-POMDP). We designed the AT-POMDP and derived its parameters based on data from a prior robot-student tutoring study. The policy that results from solving the AT-POMDP allows a robot tutor to decide upon the optimal tutoring help action to give a student, while maintaining a belief of the student's mastery of the material and engagement with the task. This approach is validated through a between-subjects field study, which involved 4th grade students ($n = 28$) interacting with a social robot solving long division problems over five sessions. Students who received help from a robot using the AT-POMDP policy demonstrated significantly greater learning gains than students who received help from a robot with a fixed help action selection policy. Our results demonstrate that this robust computational framework can be used effectively to deliver diverse and personalized tutoring support over time for students.

Introduction

Robots have shown increasing promise as a technology to emulate many of the benefits of one-on-one human tutoring in a variety of domains (Belpaeme et al. 2018; Hood, Lemaignan, and Dillenbourg 2015; Kanda et al. 2004). Notably, the physical presence of robots has demonstrated enhanced student learning gains, compliance, and enjoyment in learning interactions (Bainbridge et al. 2011; Leyzberg et al. 2012; Pereira et al. 2008). Many aspects of the robot-student tutoring interaction have been investigated such as building models of student knowledge (Spaulding, Gordon, and Breazeal 2016), evaluating different teaching and instruction paradigms (Hood, Lemaignan, and Dillenbourg 2015), and determining how and when a specific type of help would be best administered to students (Ramachandran, Litoiu, and Scassellati 2016).

Although most prior work has focused on modifying and optimizing a single component of the robot-student tutoring interaction (e.g. building a model of student knowledge, determining when hints should be provided), robot tutors could be even more effective by, like human tutors, working from a unified model of the student that encompasses the student's knowledge and engagement and employing a diverse set of assistive behaviors in order to facilitate a more comprehensive, diverse, and personalized robot-student tutoring interaction. Studies of highly successful human tutors describe how human tutors "maintain a 'working model' of each tutee" enabling students to do as much of the work as possible as well as receive sufficient guidance to prevent frustration and confusion (Lepper and Woolverton 2002; Merrill et al. 1992). For example, human tutors are excellent at giving minimal help to students who have proven their competence with the tutoring content, guiding students with low knowledge step-by-step through the process of how to solve a problem, and discerning when to stop and take a break when students get frustrated or bored.

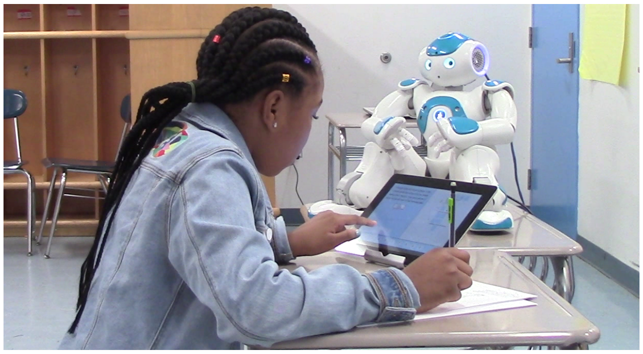


Figure 1: We studied the effects of the AT-POMDP policy on student learning in multiple tutoring sessions with a robot.

In this work, we emulate expert human tutors in the design of a robot tutor that models the affective and knowledge states of students and, using that model, selects appropriate actions to assist student learners. We design the Assistive Tutor partially observable Markov decision process (AT-POMDP), a unified framework in which a robot tutor can maintain a belief estimate of student knowledge and engagement via observations that can be directly sensed through the tutoring system and subsequently plan what supportive help actions to provide to enhance learning. We also validate the model's effectiveness in decision making by conducting a user study in which fourth grade students interact with a robot tutor that employs AT-POMDP to provide support to the students over five tutoring sessions (Figure 1). We found that the AT-POMDP policy significantly improves learning gains for students when compared to a fixed policy for choosing help actions, indicating the importance of providing personalized tutoring support for young children.

Background

In this section, we present work on help action selection for tutoring systems, other computational approaches to tutoring, and the generalized POMDP framework.

Assistive Actions Employed by Tutors

In previous work in intelligent tutoring systems, many forms of help have been developed to assist learners. Below we describe some common types of help used in tutoring systems. Hints, or help messages, are one of the most common forms of help and typically contain information about specific features of the given problem. Tutoring systems often use multiple levels of ordered hints, where subsequent hints contain more specific information to solving the problem at hand. Worked examples refer to problems that show all the necessary steps to solve them successfully, and have been shown to be an effective form of help (McLaren et al. 2014). Self-explanations are defined as the generation of explanations to oneself and have been shown to improve understanding (Chi et al. 1994). This process can also be referred to as thinking aloud which has been investigated as a metacognitive strategy and has led to improved performance (Aleven and Koedinger 2002). Step-based tutoring refers to tracking progress of individual steps in a problem and providing feedback on these intermediary steps rather than all at once after the student provides an answer (Vanlehn et al. 2005). In tutoring systems, interactive tutorials implement this step-based tutoring and leverage this idea that more interactivity and feedback on each step may lead to stronger understanding of where a misconception occurs (VanLehn 2011). Researchers have designed principles on how to provide help in tutoring systems, suggesting that the system should first elicit as much self-explanation (requesting the student to think aloud) as possible from the student and then provide instructional explanations, progressing from minimalistic to extensive forms of help (Renkl 2002).

In addition to help actions that relate to improving mastery of the educational concepts within tutoring, other support mechanisms are employed by intelligent tutoring systems to maintain learner engagement. Prior work in robot tutoring systems suggests that short, non-task breaks can be used throughout a tutoring interaction to sustain and restore engagement when provided based on student progress (Ramachandran, Huang, and Scassellati 2017).

Based on these design recommendations from prior work, we find that thinking aloud, hints, worked examples, tutorials, and breaks are all useful actions that can be used to benefit learners during a tutoring interaction. Our work focuses on planning which supportive tutoring actions to provide to a given student from a bank of actions we identify as potentially helpful to students.

Computational Approaches to Tutoring

Since the field of Intelligent Tutoring Systems (ITS) began, researchers have been investigating ways to computationally model the student in order to inform effective tutoring strategies. The most widely used tactic for modeling student knowledge is a method called Bayesian Knowledge Tracing (BKT) (Corbett and Anderson 1994). Though this is an effective technique for modeling a student's knowledge state, the computational focus is on accurately estimating student knowledge of skills and utilizing this information to inform content selection based on fixed mastery thresholds (Lee and Brunskill 2012; Yudelson, Koedinger, and Gordon 2013).

Recent work has explored employing POMDPs to plan actions in a teaching setting (Rafferty et al. 2016; Folsom-Kovarik, Sukthankar, and Schatz 2013). Rafferty et al. used a POMDP to find optimal teaching actions during a concept learning task, in which they minimize the expected time for a learner to acquire a new concept in a short interaction by balancing three teaching actions: presenting an example, giving a quiz question, and asking a question with feedback. They demonstrate that adults can learn a fabricated, simple concept mapping faster when receiving teaching actions dictated by the POMDP model versus a random baseline (Rafferty et al. 2016). In contrast to this work, our approach uses a POMDP to plan supportive help actions, rather than instructive teaching actions, to students during a real-world robot-child tutoring task in which students practice a difficult mathematical concept they have learned in school over several tutoring sessions. In addition, our approach considers the learner's engagement rather than just the learner's knowledge of a given concept and contains a rich repertoire of supportive help actions.

POMDPs

A partially observable Markov decision process (POMDP) (Kaelbling, Littman, and Cassandra 1998) can be represented as a 7-tuple $(S, A, \Omega, T, R, O, \gamma)$, where:

- S is the set of partially observable states with $s \in S$
- A is the set of possible actions with $a \in A$
- Ω is the set of observations with $o \in \Omega$
- $T : S \times A \times S \rightarrow [0, 1]$ is a probabilistic transition function such that $T(s, a, s') \equiv \Pr(s'|a, s)$
- $R : S \times A \times S \rightarrow \mathbb{R}$ is a reward function mapping state-action-state tuples to rewards
- $O : S \times A \times \Omega \rightarrow [0, 1]$ is a probabilistic observation function such that $O(a, s', o) \equiv \Pr(o|a, s')$
- $\gamma \in [0, 1]$ is the discount factor

After each time step, the agent making decisions updates its belief, b, a probability distribution over S, where $b(s)$ represents the belief relative to state s. The belief can be updated according to the following:

$$b'(s') = \eta O(s', a, o) \sum_{s \in S} T(s, a, s') b(s)$$

where $\eta = 1/\text{Pr}(o|a, b)$, a normalization term to ensure that $\sum_{s \in S} b(s) = 1$.

The solution to a POMDP is a policy that maps beliefs to actions, $\pi : B \rightarrow A$, and selects actions that maximize the value function, the expected discounted reward.

The Assistive Tutor POMDP

This section formulates the robot tutor action selection problem as a POMDP called the Assistive Tutor POMDP (AT-POMDP) and describes the model design and parameters[1]. Many of the model parameters were derived from a previous robot-student tutoring data set (Ramachandran, Litoiu, and Scassellati 2016). From this data set, we retrieved timing and accuracy data of students doing math problems with a robot tutor as well as their engagement during the math tutoring sessions by annotating low (bored, blind guessing) and high engagement (focused) in the video files.

State Space

The state space of AT-POMDP consists of three dimensions: knowledge level, engagement level, and math problem attempt number. There are four domain-independent knowledge levels that roughly equate to: little to no mastery, some mastery, moderate mastery, and near-complete mastery. There are two engagement levels: low and high. High engagement is marked by the students' attention being focused on the math problem at hand, making honest attempts at the problems. Low engagement is marked by either rapid guessing on problems without knowing the correct answer or boredom and off-task behavior. For each problem the students complete, they have three attempts to answer correctly. If all three attempts are answered incorrectly, they will be moved onto the next question. Attempts 1-[after a correct attempt], 1-[after an incorrect attempt], 2, and 3 are encoded into our state space. With four knowledge levels, two engagement levels, and four attempt definitions, the size of the state space can be defined as $|S| = 4 \times 2 \times 4 = 32$.

Action and Observation Spaces

The robot's action space consists of six tutoring actions, which during the tutoring session are administered before an attempt is made by the student on a problem.

For each attempt made by the student on a problem, we observe the accuracy of the attempt (correct or incorrect) and the speed (slow, medium, or fast) at which the student answers the question. Thus, the size of the observation space $|\Omega|$ is $2 \times 3 = 6$. To account for individual differences in attempt speeds, the z-score of attempt timing, which measures how many standard deviations the current data point is from the mean, is used to determine the timing categorization (slow, medium, fast) of a student on a particular attempt.

Reward Model

The reward function formalizes the robot tutor's goal of aiding the student to transition from lower to higher knowledge states and to transition from low to high engagement by rewarding those transitions. Each action is also taken with a cost proportional to the time it takes for the student to complete that action (e.g. interactive-tutorials take many times longer to complete than a hint). We also penalized any action other than no-action heavily on the first attempt so that students would have a chance to answer the question before a help action was chosen and performed by the robot.

Transition and Observation Models

The transition model $T(s, a, s') \equiv \text{Pr}(s'|a, s)$ can be derived by examining the likelihood of change in the attempt s_a, engagement s_e, and knowledge s_k dimensions of a single state s: $\text{Pr}(s'|a, s) \equiv c_{action} \cdot \text{Pr}(s'_a|a, s) \cdot \text{Pr}(s'_e|a, s) \cdot \text{Pr}(s'_k|a, s)$, where $\text{Pr}(s'_a|a, s)$ represents the likelihood that a student transitions to a particular attempt state by answering a question incorrectly/correctly, $\text{Pr}(s'_e|a, s)$ represents the likelihood that a student moves from one engagement level to another, $\text{Pr}(s'_k|a, s)$ represents the likelihood that a student transitions from one knowledge level to another, and c_{action} represents an action-specific constant multiplier where each action will uniquely influence state transitions. In our model, we make the assumption that students can only increase their knowledge level and cannot 'lose' knowledge.

We derived these transition probabilities from the prior robot-child tutoring study data set. We determined $\text{Pr}(s'_a|a, s)$ by examining the attempt accuracy of students in each knowledge level, which depended on whether they were on the first attempt or a subsequent attempt (student accuracy was lower for subsequent attempts) and the students' engagement (student accuracy was lower if they were not engaged). We derived $\text{Pr}(s'_e|a, s)$ by calculating the proportion of attempts where students moved engagement states, which depended on their knowledge state (students in higher knowledge states were less likely to move from high to low engagement). We determined $\text{Pr}(s'_k|a, s)$ by computing the proportion of attempts where a student crosses the threshold from one knowledge level to another, which depended on their engagement state (students in our data set did not increase knowledge state when at a low engagement level). In addition to deriving these variables from our data set, we also tested them in simulation.

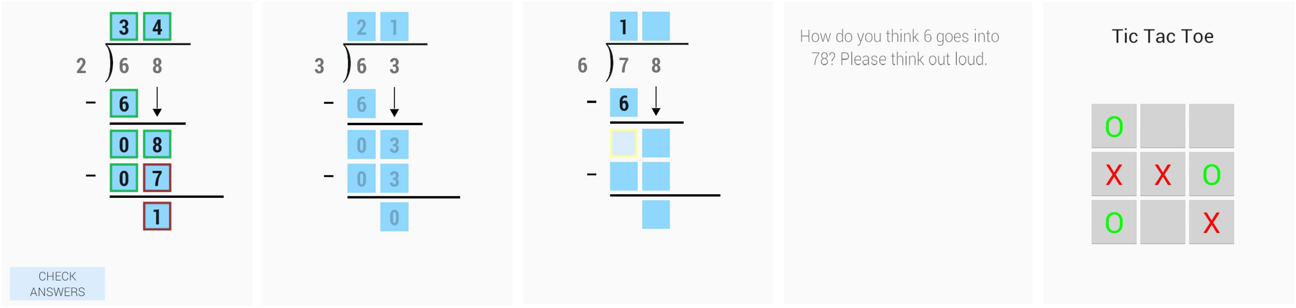
The observation model $O(a, s', o) \equiv \text{Pr}(o|a, s')$, was derived by computing the likelihood of the observed accuracy and speed on the attempt just completed, given the action and end state.

Using the prior robot-child tutoring study data set, we were able to derive $\text{Pr}(o|a, s')$ by computing the proportion of student attempts in the three speed categories (slow, medium, fast), which depended on the engagement of the student, since students with low engagement were more likely to exhibit slow and fast attempt speeds.

AT-POMDP Policy Computation

We used an offline POMDP solver originally implemented by (Kaelbling, Littman, and Cassandra 1998) and modified by (Roncone, Mangin, and Scassellati 2017) to solve for the AT-POMDP policy. We computed the AT-POMDP's belief update online and the robot's action selection based on

[1] Code used to create AT-POMDP is available at: https://github.com/ScazLab/AT-POMDP



How do you think 6 goes into 78? Please think out loud.

(a) (b) (c) (d) (e)

Figure 2: Tablet screenshot examples of each help action used in the tutoring system setup: (a) interactive-tutorial, (b) worked-example, (c) hint, (d) think-aloud, (e) break.

our solved policy was determined in real-time during the repeated tutoring sessions with fourth grade students.

Methodology

In this section, we describe a user study that explores the effects of an autonomous robot tutoring system that employs the AT-POMDP policy on student learning outcomes. We describe the user study's educational context and the design of our integrated robot tutoring system, followed by our experimental conditions, procedure, and participants.

Evaluation Context

We developed long division math content for the tutoring sessions for 4th grade students. Long division is a challenging concept for students and one where they could benefit from extra practice. We designed the content to be in line with common core standards and with the aid of a 4th grade teacher. Although the 4th grade students in this study had been taught long division in the classroom, many students failed to grasp the full process or had forgotten several steps. We focused our curriculum design on fostering improvement with both long division strategy use as well as successful application of division concepts (both obtaining a correct answer and using the correct strategy).

Robot Tutoring System

Our tutoring system consisted of a Nao robot, a tablet device for input, and several software components that enabled the flow of an autonomous tutoring interaction. We used a ROS architecture to coordinate communication between the robot, tablet, and software components of the system that implemented our help action selection method (Quigley et al. 2009). The Nao robot acted as a tutoring agent throughout each interaction and facilitated the interaction by introducing each question, giving feedback on whether an entered answer was correct or incorrect, and proactively providing help according to the participant's experimental condition.

The tablet application was used to display questions, feedback, and help to the students and was used as an input device for entering answers to each question. The tablet had two panels, one in which the current question and input box was displayed, and the other in which interactive help would be displayed (see Figure 2). In the following, we describe each help action in detail:

- *Interactive-tutorial.* The tablet displayed a long division box structure and allowed the user to only interact with the boxes associated with one "step" of a problem at a time. Students were required to correctly enter the numbers associated with one long division step before progressing to the next step. The robot provided feedback and guidance to the student through each step.
- *Worked-example.* The tablet displayed a comparable problem in difficulty and solution process to the problem at hand. The robot verbally described each step of the long division process for that problem, filling in the non-interactive boxes in the displayed structure in time with the verbal explanation.
- *Hint.* The tablet displayed an interactive long division structure on the tablet help panel for the student's current problem. The robot suggested that the student use the structure to help solve the current problem.
- *Think-aloud.* The robot made requests of the student to verbalize their long division solving process. These think-aloud prompts were also displayed in text on the tablet.
- *Break.* The student played one game of tic-tac-toe with the robot as a non-task break.
- *No-Action.* Students received no action if the system determined they did not need help on a given attempt.

Experimental Design

We designed a between-subjects study involving two experimental conditions—the AT-POMDP condition and the fixed condition. We want to understand whether the AT-POMDP policy benefits students when compared to a 'best practice' fixed policy for selecting help actions. All students received the same educational math content and help actions regardless of their experimental group. What differed between the groups was the decision of which help action to provide to the student when help was being given. Below, we describe the action selection policies for each condition.

AT-POMDP Condition - The students in this group received help actions according to the AT-POMDP policy. The AT-POMDP selects the best help action to give to the student based on its belief of the knowledge and engagement of the student. The model's belief is updated after each action-observation pair and is preserved between sessions. The initial state for each student was determined by their pretest score and the assumption that the children would be engaged during the first problem of the first session.

Fixed Condition - The students in the fixed condition received help according to a fixed policy we designed based on current best practice in education and intelligent tutoring systems. Each time a student gets a question incorrect, they receive a help action, in order from the "smallest" to the "largest" help actions, excluding *no-action*. We created a fixed policy to provide progressive help in the following order: *think-aloud*, *hint*, *worked-example*, *interactive-tutorial*. This mimics hint systems commonly used in ITSs where subsequent help given to students becomes more specific and helpful to solving the problem. When a student answers a question correctly, the level of help provided resets to the smallest amount of help (think-aloud). Students in the fixed condition received a break once per session, approximately halfway through the 15-minute session.

Experimental Procedure

Parental and child consent forms were collected for each student prior to participation in this experiment. Before interacting with the robot tutoring system, students completed an 8-question pretest designed to assess incoming knowledge about the division concepts covered during the tutoring sessions. Students were randomly assigned to one of the experimental conditions and interacted with the robot tutoring system for five sessions, each lasting approximately 15 minutes. Each student completed as many problems as they could from a bank of practice problems within the 15-minute session. The AT-POMDP model did not influence problem selection, and each student received the same series of long division problems regardless of experimental condition. During each session, students sat at a table facing the robot and tablet and used scratch paper if needed. All one-on-one interactions were autonomous, requiring no input from the experimenter during tutoring. Each of the five sessions was completed on separate days spaced out over approximately three weeks. On a separate day after the last session, participants completed a posttest. The pretest and posttest were identical, each consisting of the same questions that encompassed relevant long division concepts that were represented during the tutoring interactions.

Participants

We recruited 30 participants from a local elementary school to participate in this study. Two participants were excluded in this data analysis (one due to non-compliance and one due to a perfect pretest score), resulting in a total of 28 participants. Participants were randomly distributed into the two experimental groups, resulting in 14 students per condition. The two groups were gender-balanced, each having exactly 6 males and 8 females. Between groups, there was no significant difference in starting knowledge levels as measured by pretest scores between the fixed ($M = .44$, $SD = .31$) and AT-POMDP ($M = .30$, $SD = .36$) conditions, $t(26) = 1.146$, $p = .262$. All students in the study were in fourth grade, resulting in comparable ages between the AT-POMDP ($M = 9.29, SD = .47$) and fixed ($M = 9.21, SD = .43$) conditions, $t(26) = -.422, p = .676$.

Results

In this section, we present findings characterizing participants' interactions with the system over the five tutoring sessions and results on differences in learning outcomes for students between our two experimental groups. We also show metrics of how the AT-POMDP policy's decisions differed from the fixed policy's decisions and highlight case studies of participants who benefited from the decisions of the AT-POMDP policy. When comparing our two experimental groups directly, we use independent t-tests and when assessing one group's progress by comparing within-subjects measures, we use paired t-tests. For all statistical tests, we used an α level of .05 for significance in our analysis.

Action Selection in Tutoring Sessions

Participants in both groups received a similar number of help actions over all five sessions. Participants in the fixed condition received an average of 19.43 ($SD = 5.00$) help actions and participants in the AT-POMDP condition received an average of 19.57 ($SD = 10.82$) help actions, $t(26) = -.045$, $p = .965$. For each of the five 15-minute sessions, participants received on average 3.90 ($SD = 1.65$) help actions.

Participants in the AT-POMDP and fixed conditions received a significantly different distribution of help actions across all five sessions, $\chi^2(5, N = 28) = 168.78, p < .001$, as shown in Figure 3a. In addition to analyzing the difference in the actions chosen between the fixed and AT-POMDP conditions, we examined the differences in the actions chosen for participants in the AT-POMDP condition and the actions that would have been chosen for those same participants if they had been in the fixed condition. We found a similar result in that the distribution of help actions across all five sessions was significantly different, $\chi^2(5, N = 14) = 98.23, p < .001$. Additionally, 85.4% of the 274 total actions the participants in the AT-POMDP condition received were different than the actions they would have received had they been in the fixed condition.

These results support the conclusion that participants in the AT-POMDP and fixed conditions received significantly different distributions of help actions and, additionally, that the actions chosen by the AT-POMDP and fixed policies were also significantly different.

Learning Gains Results

Students completed a pretest before the first tutoring session and a posttest after the fifth session. Each student received a

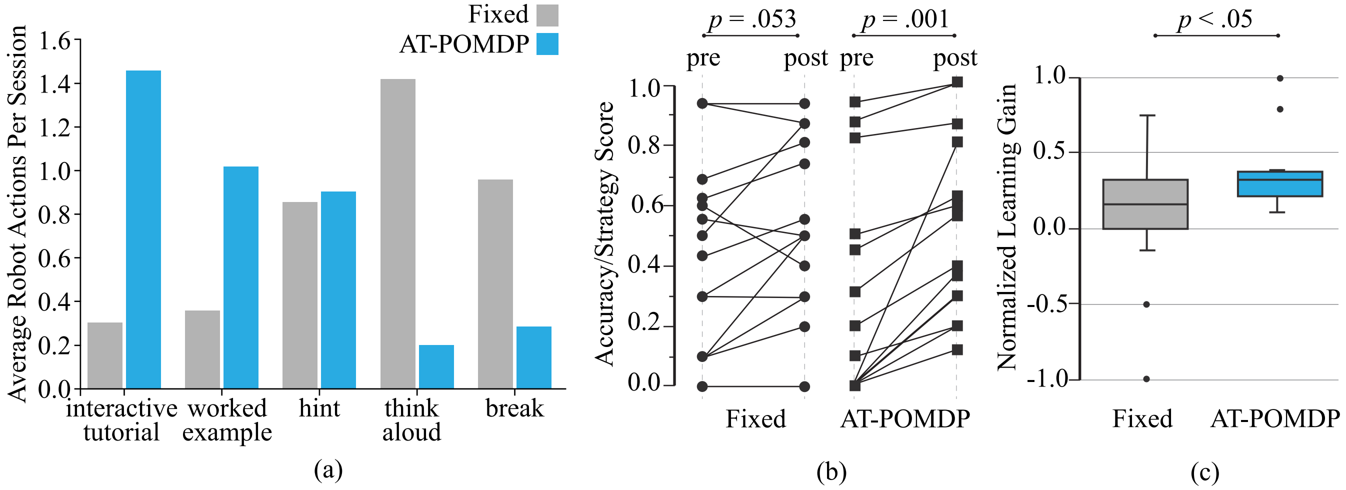


Figure 3: (a) Distribution of actions averaged per session per student. (b) Students in the AT-POMDP condition significantly improved their test scores from pretest to posttest. (c) Students in the AT-POMDP condition improved their scores based on accuracy and strategy use significantly more than those in the fixed condition.

test score for both the pretest and posttest, where each question was scored for both accuracy (0 or 1) and correct long division strategy use (0 or 1). This scoring scheme was in line with the design of our mathematical content, awarding points for both accuracy and correct strategy use. Below we define normalized learning gain to measure improvement from pretest to posttest for each student i:

$$nlg(i) = \frac{score_{\text{post}}(i) - score_{\text{pre}}(i)}{1 - score_{\text{pre}}(i)}$$

These scores were calculated by diving the number of points received by the total number of points it was possible to receive for each test. The metric of nlg provides a measure of learning improvement for each student, accounting for different starting knowledge levels. Similar uses of nlg can be found in the review (Belpaeme et al. 2018).

For students in the fixed condition, posttest ($M = .54, SD = .28$) scores marginally improved from pretest scores ($M = .44, SD = .31$), $t(13) = -2.128$, $p = .053$. Students in the AT-POMDP condition had posttest scores ($M = .53, SD = .30$) that were significantly higher than their pretest scores ($M = .30, SD = .36$), $t(13) = -4.473$, $p = .001$ (Figure 3b). In comparing nlg between the two conditions, we found that average nlg for the AT-POMDP condition ($M = .41, SD = .30$) was significantly higher than for the fixed condition ($M = .08, SD = .43$), $t(26) = -2.326, p = .028$ (Figure 3c). These results indicate that the students who received help actions from the robot according to the AT-POMDP policy improved their strategy use and accuracy on long division concepts significantly more than the students who received help actions according to a best-practice fixed policy. Given that students across groups only received 3.83 help actions per session on average, this difference between groups further highlights the impact of the decisions made by the AT-POMDP policy.

Case Studies

In order to further examine the actions chosen by the AT-POMDP policy, in this section we take a more in-depth look at three individual students in the AT-POMDP condition. For these students, we examine the tutoring action choices made by the AT-POMDP policy and evaluate their effectiveness.

Participant 11 (P11) was one of the highest performing students in our sample with an attempt accuracy of 92.1%, as compared to the 41.2% attempt accuracy of the entire sample. Of the 114 attempts P11 made on problems over the 5 sessions, P11 received 7 tutoring help actions from the robot, selected by the AT-POMDP policy: 1 hint, 1 break, 1 think-aloud, and 4 no-actions. Given that P11 displayed a high mastery of the long-division material, the AT-POMDP estimated that P11 was in a high knowledge state and thus, the cost of selecting help actions like hints, worked-examples, and interactive-tutorials would have too high to be worthwhile, so the model selected a majority of no-action help actions for P11. Despite not receiving help when the model selected no-action, P11 answered the next attempt correctly 3 out of the 4 times this occurred.

Participant 25 (P25) was one of the lower performing students in our sample. P25's accuracy on the attempts (19.1%) was lower than the entire sample's attempt accuracy (41.2%). Additionally, P25 answered merely 1.8 attempts correctly out of 9.4 on average per session. Of the 25 help actions the AT-POMDP policy selected for P25, 12 were interactive-tutorials and 7 were worked-examples, the two most comprehensive and involved tutoring help actions. P25 did not answer any questions correctly on either the pretest or the posttest, however, attempted 2 more long division problems on the posttest than the pretest, showing an increased confidence with attempting long division problems. Had P25 been in the fixed condition, the fixed policy would have selected 9 think-alouds and 8 hints, the two most minimal tutoring help actions, and only 4 worked-examples and

3 interactive-tutorials. It seems unlikely that if P25 had been in the fixed condition, P25 would have grown in confidence and familiarity with long division from the pretest to posttest since P25 would have received considerably less long division assistance as compared with the help P25 received in the AT-POMDP condition.

Participant 12 (P12) was also one of the lower performing students in our sample. P12's accuracy on question attempts (9.8%) was substantially lower than the entire sample's attempt accuracy (41.2%) and P12 answered a meager 0.8 attempts correctly out of 8.2 attempts on average per session. From watching P12's tutoring session videos, P12 tended to be more distracted and disengaged than the average student, likely due to the difficulty of the problems and P12's low attempt accuracy. The AT-POMDP policy selected a total of 5 tic-tac-toe breaks across the 5 sessions: 1 break in sessions 2, 3, and 4, and 2 breaks in session 5. P12 received all of these breaks after an incorrect answer on the previous question with a faster speed ($M = 27.2s, SD = 9.0s$) than P12's average question answering speed ($M = 67.6s, SD = 43.5s$), indicating that P12 was presumably making blind guesses and that a break would likely be useful for reengaging P12. After the tic-tac-toe breaks, P12's accuracy on the next attempt was 40.0%, much higher than P12's overall attempt accuracy during all of the sessions, 9.8%, suggesting that the breaks were well-timed and effective for P12.

Through the examination of these case studies, we encounter three diverse action selection approaches by the AT-POMDP policy: giving limited help to a student who displayed mastery of the material, providing significant help to a student who showed little mastery of long division, and administering appropriately timed breaks to a student who was frequently disengaged. For these three students it is possible that separate individual fixed policies could be developed to support each student's learning behavior, however, it is extremely unlikely that one overall fixed policy could be developed to provide the learning support needed for all three students. In contrast, the AT-POMDP can produce a policy that makes successful and effective action choices, supporting the learning and engagement of a diverse set of students in one unified model.

Discussion

In this work, we implemented the AT-POMDP that enabled robot tutors to autonomously provide an appropriate help action to students based on an estimate of their knowledge and engagement levels. We demonstrated that with a single, unified model, we could provide help actions to individual students according to their needs. By evaluating the effectiveness of the AT-POMDP in a five-session long-term tutoring interaction, we demonstrated that students strengthened their learning on a long division task by exhibiting improved test scores based on accuracy and correct strategy use. Furthermore, these students improved more than students who received help from the robot tutoring system according to a fixed policy. By examining certain participants in closer detail, rather than just looking at the average learning gains across groups, we can see specific instances in which the AT-POMDP policy selected appropriate actions for the individual child. Our results highlight the value of building robust, computational frameworks to deliver personalized tutoring support over time for young students.

Other investigations into probabilistic models for teaching have also demonstrated the benefits of an approach that can plan under uncertainty in finding useful policies for teaching tasks (Rafferty et al. 2016; Murray and VanLehn 2006). Our work is in agreement with this body of work, and we provide further evidence for the usefulness of a POMDP model used to plan under uncertainty in a long-term tutoring setting for children. Rather than focus on the sequencing of teaching content, the AT-POMDP we designed selects supportive help actions the tutor can take to strengthen student learning of a concept that is challenging for them.

Our results indicated a difference in average learning gains between the AT-POMDP condition and the fixed condition. However, the two conditions differed in a number of ways that could have led to the AT-POMDP condition participants showing increased learning gains over the fixed condition participants including differences in the distribution of help actions between conditions and the variation in the ordering of the help actions given. While we demonstrated that the AT-POMDP could be used for personalized help action selection in tutoring, additional research and user studies must be conducted to tease apart exactly which factors led to the difference in learning gains between conditions and to what degree each factor influenced the results.

Though the AT-POMDP policy was effective in strengthening learning outcomes, we found that not all students improved their long division skills. The lowest performing students often received "larger" help actions frequently (e.g. interactive-tutorials, worked examples), and this may have helped them improve their attempt rate as well as their tendencies to employ the correct strategy when solving long division problems. However, we noticed that those who started with extremely low incoming pretest scores, were typically unable to demonstrate strong mastery of complex long division skills even after five sessions. We acknowledge that our model could still benefit from additional personalization, such as adapting the help action choice according to individual preferences.

Conclusion

In this paper, we designed the Assistive Tutor POMDP (AT-POMDP) to provide personalized support to students practicing a difficult math concept over several tutoring sessions. The AT-POMDP estimated a student's individual knowledge level and engagement level and computed a policy to make decisions on the appropriate help action to take to increase the likelihood of the student reaching higher knowledge and engagement levels. The AT-POMDP was effective in providing a personalized approach to planning and balancing several different help actions in a tutoring setting. Our evaluation demonstrated the effectiveness of using the AT-POMDP to help students with a long division math task as students receiving help from the AT-POMDP policy showed improved learning gains when compared to students receiving help from a fixed policy to govern help action selection.

References

Aleven, V., and Koedinger, K. R. 2002. An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive science* 26(2):147–179.

Bainbridge, W. A.; Hart, J. W.; Kim, E. S.; and Scassellati, B. 2011. The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics* 3(1):41–52.

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

Chi, M. T.; De Leeuw, N.; Chiu, M.-H.; and LaVancher, C. 1994. Eliciting self-explanations improves understanding. *Cognitive science* 18(3):439–477.

Corbett, A. T., and Anderson, J. R. 1994. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction* 4(4):253–278.

Folsom-Kovarik, J. T.; Sukthankar, G.; and Schatz, S. 2013. Tractable pomdp representations for intelligent tutoring systems. *ACM Transactions on Intelligent Systems and Technology (TIST)* 4(2):29.

Hood, D.; Lemaignan, S.; and Dillenbourg, P. 2015. When children teach a robot to write: An autonomous teachable humanoid which uses simulated handwriting. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, 83–90. ACM.

Kaelbling, L. P.; Littman, M. L.; and Cassandra, A. R. 1998. Planning and acting in partially observable stochastic domains. *Artificial intelligence* 101(1-2):99–134.

Kanda, T.; Hirano, T.; Eaton, D.; and Ishiguro, H. 2004. Interactive robots as social partners and peer tutors for children: A field trial. *Human–Computer Interaction* 19(1-2):61–84.

Lee, J. I., and Brunskill, E. 2012. The impact on individualizing student models on necessary practice opportunities. *International Educational Data Mining Society*.

Lepper, M. R., and Woolverton, M. 2002. The wisdom of practice: Lessons learned from the study of highly effective tutors. *Improving Academic Achievement: Impact of Psychological Factors on Education* 135–158.

Leyzberg, D.; Spaulding, S.; Toneva, M.; and Scassellati, B. 2012. The physical presence of a robot tutor increases cognitive learning gains. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 34.

McLaren, B. M.; van Gog, T.; Ganoe, C.; Yaron, D.; and Karabinos, M. 2014. Exploring the assistance dilemma: Comparing instructional support in examples and problems. In *International Conference on Intelligent Tutoring Systems*, 354–361. Springer.

Merrill, D. C.; Reiser, B. J.; Ranney, M.; and Trafton, J. G. 1992. Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *The Journal of the Learning Sciences* 2(3):277–305.

Murray, R. C., and VanLehn, K. 2006. A comparison of decision-theoretic, fixed-policy and random tutorial action selection. In *International Conference on Intelligent Tutoring Systems*, 114–123. Springer.

Pereira, A.; Martinho, C.; Leite, I.; and Paiva, A. 2008. icat, the chess player: the influence of embodiment in the enjoyment of a game. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 3*, 1253–1256. International Foundation for Autonomous Agents and Multiagent Systems.

Quigley, M.; Conley, K.; Gerkey, B.; Faust, J.; Foote, T.; Leibs, J.; Wheeler, R.; and Ng, A. Y. 2009. Ros: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3, 5. Kobe.

Rafferty, A. N.; Brunskill, E.; Griffiths, T. L.; and Shafto, P. 2016. Faster teaching via pomdp planning. *Cognitive science* 40(6):1290–1332.

Ramachandran, A.; Huang, C.-M.; and Scassellati, B. 2017. Give me a break!: Personalized timing strategies to promote learning in robot-child tutoring. In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 146–155. ACM.

Ramachandran, A.; Litoiu, A.; and Scassellati, B. 2016. Shaping productive help-seeking behavior during robot-child tutoring interactions. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, 247–254. IEEE Press.

Renkl, A. 2002. Worked-out examples: Instructional explanations support learning by self-explanations. *Learning and instruction* 12(5):529–556.

Roncone, A.; Mangin, O.; and Scassellati, B. 2017. Transparent Role Assignment and Task Allocation in Human Robot Collaboration. *IEEE International Conference on Robotics and Automation (ICRA)*.

Spaulding, S.; Gordon, G.; and Breazeal, C. 2016. Affect-aware student models for robot tutors. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, 864–872. International Foundation for Autonomous Agents and Multiagent Systems.

Vanlehn, K.; Lynch, C.; Schulze, K.; Shapiro, J. A.; Shelby, R.; Taylor, L.; Treacy, D.; Weinstein, A.; and Wintersgill, M. 2005. The andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence in Education* 15(3):147–204.

VanLehn, K. 2011. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist* 46(4):197–221.

Yudelson, M. V.; Koedinger, K. R.; and Gordon, G. J. 2013. Individualized bayesian knowledge tracing models. In *International Conference on Artificial Intelligence in Education*, 171–180. Springer.