

# Inferring learners’ knowledge from observed actions

Anna N. Rafferty  
Computer Science Division  
University of California  
Berkeley, CA 94720

rafferty@cs.berkeley.edu

Michelle M. LaMar  
Graduate School of Education  
University of California  
Berkeley, CA 94720

mlamar@berkeley.edu

Thomas L. Griffiths  
Department of Psychology  
University of California  
Berkeley, CA 94720

tom\_griffiths@berkeley.edu

## ABSTRACT

Teachers gain significant information about their students through close observation of classroom activities. By noting which actions a student takes to achieve particular goals, a teacher can often infer the knowledge possessed by the student and diagnose misconceptions. In this work, we develop a framework for automatically inferring a student’s underlying beliefs from a set of observed actions. This framework relies on modeling how student actions follow from beliefs about the effects of those actions. We demonstrate the practicality of this approach by modeling empirical student data from an educational game and validate its performance via a controlled lab experiment. In the educational game, inferences were consistent with conventional assessment measures; in the lab experiment, the model’s inferences reflect participants’ stated beliefs.

## 1. INTRODUCTION

By observing a student work towards a goal, a teacher can infer what actions the student believes are necessary to achieve the goal and how the student believes those actions affect her progress towards the goal. Critically, a teacher might observe a student’s actions and realize that the student has misconceptions about the effects of the actions. This allows for intervention and correction of those specific misconceptions – something that is vital in an educational setting [5]. In this work, we formalize a student model in which the student’s knowledge is characterized by her beliefs about how her actions affect the state of the world and what states are most beneficial for achieving her goals. We propose a framework for automatically inferring these beliefs based on observed actions, drawing on ideas from inverse reinforcement learning. In order to make inferences about students’ beliefs, we use Markov decision processes to model how those beliefs combine with their goals to determine their actions. Previous work that has focused on understanding the actions of others has typically assumed that the person taking the actions has full knowledge of how those actions affect the world. For example, work in plan recognition has focused on identifying what the intended plan of action is from some set of already observed actions (e.g., [4]), as well as on categorizing sets of individual actions into strategies and larger semantic parts (e.g., [1]). This work has shown plan recognition to be useful both in human-computer interaction and in understanding data from educational programs. Our

approach extends the idea of automated inference about a student’s actions to a context in which there may be many actions performed and the person may have misconceptions about how actions affect progress toward the goal.

## 2. INFERRING STUDENT BELIEFS

We consider tasks in which students are trying to achieve some known goal (e.g., win a level in a game) but may have misconceptions about how to achieve that goal. We model these misconceptions using Markov decision processes (MDPs). MDPs provide a natural framework for sequential planning problems in which people must reason about the immediate gains or costs of an action and how that action affects the ease with which the goal can be achieved in the future (see [7] for an overview). An MDP is a tuple  $\langle S, A, T, R, \gamma \rangle$ , where  $S$  is the set of possible states of the world and  $A$  is the set of actions that one can take.  $T$  represents the transition model  $p(s'|s, a)$  specifying the probability of transitioning to a state  $s'$  given that the action  $a$  was taken in state  $s$ .  $R$  corresponds to the reward model  $r(s, a, s')$  that specifies the reward for taking action  $a$  in state  $s$  and entering state  $s'$ , while  $\gamma$  represents the relative value of immediate versus future rewards. From this specification, one can calculate the expected sum of discounted rewards obtained from each state  $s$  and action  $a$ :

$$Q(s, a) = \sum_{s' \in S} p(s'|a, s) \left( r(s, a, s') + \gamma \sum_{a' \in A} p(a'|s') Q(s', a') \right),$$

which is known as the  $Q$ -function and can be calculated via a dynamic program [3]. The distribution  $p(a|s)$ , known as the policy, gives the probability an agent will choose action  $a$  given state  $s$ . As in [2], we model people using a noisily optimal policy in which  $p(a|s, T, R, \gamma) \propto \exp(\beta Q(s, a|T, R, \gamma))$ , where  $\beta$  is a noise parameter.

We assume that students may have misunderstandings about the effects of their actions, and thus their beliefs may not reflect the true transition model. Our goal is to infer what transition model the student believes is correct. Formally, we consider a hypothesis space  $\mathcal{T}$  of possible transition models and infer a probability distribution over this space based on the observed student actions. Using Bayes’ rule, we can compute the posterior distribution  $p(T|\mathbf{a}, \mathbf{s}) \propto p(\mathbf{a}|T, \mathbf{s})p(T)$ , where  $\mathbf{a} = (a_1, \dots, a_n)$  is the series of observed actions and  $\mathbf{s} = (s_1, \dots, s_n)$  is the corresponding series of states. This posterior distribution represents how likely it is that the student’s beliefs correspond to a given hypothesis  $T$  by combining the prior  $p(T)$  and the likelihood  $p(\mathbf{a}|\mathbf{s}, T, R, \gamma)$ .

The prior can encode knowledge about which misconceptions are common. The likelihood  $p(\mathbf{a}|\mathbf{s}, T, R, \gamma)$  represents how well the data fit hypothesis  $T$  and can be computed via the Markov property. We can use the posterior distribution over transition models to determine how probable it is that the student has an incorrect understanding of her actions and to calculate what misconceptions are most likely.

### 3. APPLICATIONS

We used the MDP framework to infer learner beliefs in both an educational game and in a learning task in the lab. The first task allowed us to compare the model to traditional assessment measures, while the latter provided an opportunity to more carefully validate the predictions of the model.

#### 3.1 Microbe Game

We first applied the MDP framework to data from a publicly available educational game in which students learned about cell biology by playing the part of a microbe navigating through increasingly challenging environments [6]. The student’s goal on each of the ten levels is to maximize their chances of surviving the level by purchasing appropriate amounts of mitochondria and chloroplasts. Students may play a level multiple times if they are initially unsuccessful. We modeled data from Level 6 of the game, which introduces sunlight into the environment for the first time. In applying the MDP framework, we consider transition models defined by the number of mitochondria and chloroplasts the student believes are optimal for success. We then infer these beliefs from the series of buying actions and play attempts.

Data came from a pilot study of the educational game conducted in seven schools. A total of 127 students played the game in class or at home and then participated in a post-test to measure content understanding. Post-test scores were analyzed using a standard Rasch Item Response Theory model [8], which yields an ability estimate for each student. The MDP model was used to calculate *maximum a posteriori* (MAP) estimates of students’ beliefs about the ideal number of mitochondria and chloroplasts. An analysis of variance on these MAP estimates shows that inferred beliefs were highly significant predictors for estimated ability scores on the post test (mitochondria:  $F = 4.9$ ,  $p < 0.001$ ; chloroplasts:  $F = 2.9$ ,  $p < 0.01$ ). The relationship between average ability estimates of students and their MAP estimates for ideal mitochondria and chloroplasts shows ability peaks at moderate levels of both features. This tracks well with the assumption that the game requires both mitochondria and chloroplasts but excessive amounts waste resources.

#### 3.2 Flight Planning Experiment

To validate the MDP framework, we also applied it to modeling learners’ actions in a lab experiment where we could collect explicit reports about their beliefs. In this experiment, 25 participants learned to control a spaceship by pressing different buttons. The experiment alternated between phases in which learners could choose what button to press and observe the effect of that action, and flight planning phases in which learners were asked to enter a series of button presses that would move the ship from its current location to another specified location; all participants completed six flight planning phases. Each button press moved the ship by a fixed amount, and learners were told that each

button either usually moved the ship in one particular direction or in a direction at random. They could indicate their beliefs about how a button worked using drop-down menus below each button.

We evaluated the model’s performance based on how well it matched each participant’s stated beliefs in the flight planning phases. Overall, the model achieved relatively high accuracy at inferring learner’s beliefs about the buttons that they used: The MAP estimate of the model matched the stated beliefs of the learner in 73% of flight plans. Additionally, in cases where the data were inherently ambiguous such that a human observer would also have difficulty inferring the learner’s beliefs, the model tended to place similar posterior mass on all supported hypotheses. This feature suggests the importance of modeling a full posterior distribution rather than only considering the MAP estimate.

### 4. CONCLUSION

We have developed a framework using Markov decision processes for inferring learners’ beliefs about the effects of their actions. Such a model has the potential to provide useful feedback to students about their misunderstandings and to provide information to teachers about their students’ knowledge. Designing the computational framework for this model is a first step towards applying it in more complex education settings such as virtual labs or games in which more information about the students precise behavior is known.

**Acknowledgements.** We thank WestEd for the use of the microbe game data (supported by NSF grant number DRL-0816359). We also thank Benj Shapiro and HyeYoung Shin for help with data collection. This work was supported by a DoD NDSEG Fellowship to ANR and NSF grant number IIS-0845410 to TLG.

### 5. REFERENCES

- [1] O. Amir and Y. Gal. Plan recognition in virtual laboratories. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI)*, pages 2392–2397, 2011.
- [2] C. L. Baker, R. R. Saxe, and J. B. Tenenbaum. Action understanding as inverse planning. *Cognition*, 113(3):329–349, 2009.
- [3] R. E. Bellman. *Dynamic Programming*. Princeton University Press, Princeton, NJ, USA, 1957.
- [4] H. Kautz and J. F. Allen. Generalized plan recognition. In *Proceedings of the Fifth National Conference on Artificial Intelligence*, pages 32–37. Philadelphia, PA, 1986.
- [5] T.-C. Liu, Y.-C. Lin, and Kinshuk. The application of Simulation-Assisted Learning Statistics (SALS) for correcting misconceptions and improving understanding of correlation. *Journal of Computer Assisted Learning*, 26(2):143–158, 2010.
- [6] Red Hill Studios. Lifeboat to Mars. <http://www.pbskids.org/lifeboat>, 2009.
- [7] R. S. Sutton and A. G. Barto. *Reinforcement learning*. MIT Press, 1998.
- [8] M. Wilson. *Constructing measures: An item response modeling approach*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, 2005.