

The More the Merrier? Examining Three Interaction Hypotheses

Min Chi (minchi@cs.cmu.edu)

Machine Learning Department, Carnegie Mellon University, 5000 Forbes Avenue
Pittsburgh, PA 15213 USA

Kurt VanLehn (Kurt.Vanlehn@asu.edu)

School of Computing, Informatics and Decision Systems Engineering, Arizona State University
Tempe, AZ 85287 USA

Diane Litman (litman@cs.pitt.edu)

Department of Computer Science, University of Pittsburgh, 210 South Bouquet Street
Pittsburgh, PA 15260 USA

Abstract

While high interactivity is one of the key characteristics of one-on-one human tutoring, a great deal of controversy surrounds the issue of whether interactivity is indeed the key feature of tutorial dialogue that impacts students' learning. In this paper we investigate three interaction hypotheses: a widely-believed monotonic interactivity hypothesis, a better supported interaction plateau hypothesis, and our tactical interaction hypothesis. The monotonic interaction hypothesis predicts that increasing interactivity causes an increase in learning; the plateau hypothesis states that increasing interactivity yields increasing learning until it hits a plateau, and further increases in interactivity do not cause noticeable increases in learning. Finally, the tactical interaction hypothesis predicts that interactivity only increases learning when interactions are guided by effective tutorial tactics. In this paper, we examine each hypothesis in the context of an empirical study, the results of which support the tactical interaction hypothesis.

Keywords: machine learning; reinforcement learning; pedagogical strategy; Intelligent Tutoring Systems.

Introduction

One-on-one tutoring is a highly effective educational intervention. Tutored students often perform significantly better than students in classroom settings (Bloom, 1984). Computer learning environments that mimic aspects of human tutors have also been highly successful. Intelligent Tutoring Systems (ITSs) have been shown to be highly effective at improving students' learning in real classroom settings (Koedinger et al., 1997; VanLehn, 2006). A key characteristic of one-on-one tutoring, with both human and computer tutors, is high interactivity.

A common assumption, often referred to as the *monotonic interaction hypothesis* (VanLehn, Graesser, et al., 2007) is that greater interactivity leads to greater learning.

However, several studies have failed to confirm this hypothesis. Experiments with human tutors found no significant differences in learning gains when content was carefully controlled and interactivity was directly manipulated (M. T. H. Chi et al., 2001, 2008; Rose et al., 2001). Experiments that compared human tutors and several Natural Language dialogue-based computer tutors also found no significant differences in learning as interactivity varied across students (Evens & Michael, 2006; VanLehn, Graesser, et al., 2007; Reif & Scott, 1999; Katz et al., 2003; Fossati et al.,

2008). In a meta-analysis of the tutoring literature, VanLehn found little support for the monotonic interactivity hypothesis and instead proposed the interaction plateau: the hypothesis that increased interactivity increases learning up to a point (roughly, the level of interactivity afforded by conventional step-based ITSs); beyond that threshold, however it does not yield any noticeable increases in learning (VanLehn, submitted).

On the other hand, for any form of tutoring the tutor's behaviors can be viewed as a sequential decision processes wherein, at each discrete step, the tutor is responsible for selecting the next action to take. Each of these tutorial decisions affects successive actions. Some existing theories of learning suggest that when making tutorial decisions, a tutor should adapt its actions to the students' needs based upon their current knowledge level, affective state, and other salient features (Vygotsky, 1971; Collins et al., 1989; Koedinger & Aleven, 2007). Most studies cited above made use of human tutors for their highly-interactive condition, simply assuming that expert tutors will take optimal actions. However, Chi et al. and others have argued that human tutors may not always make optimal tutorial decisions (M. T. H. Chi et al., 2001, 2008). Given that tutoring is a rather complex procedure and tutors have to make many decisions fairly rapidly, even expert human tutors may not take the full advantage of the tutorial alternatives.

Therefore, in this paper we propose a third hypothesis: the *tactical interaction hypothesis*. It states that interactivity only increases learning when interactions are guided by effective tutorial tactics. By "tutorial tactics" we refer to the policies used for selecting the tutorial action taken at each step when there are multiple actions available. In other words, we hypothesize that the tutors' success will not be governed by *how often* they give interactive prompts or ask the students questions but *how well*.

To investigate the three hypotheses, we focused on two tutorial actions: *elicit* and *tell*. During the course of one-on-one tutoring, tutors often face a simple question, should they *elicit* the next step information from the student, or should they *tell* the student the next step directly? There are many theories, but no widespread consensus on how or when an *elicit* or a *tell*

(a) Elicit Version	(b) Tell Version
1. T: So let's start with determining the value of v_1 . 2. T: Which principle will help you calculate the rock's instantaneous magnitude of velocity at T1? {ELICIT} 3. S: definition of kinetic energy 4. T: Please write the equation for how the definition of kinetic energy applies to this problem at T1 {ELICIT} 5. S: $ke_1 = 0.5 * m * v_1^2$ 6. T: From $KE_1 = 0.5 * m * v_1^2$, ...	1. T: So let's start with determining the value of v_1 . 2. T: To calculate the rock's instantaneous magnitude of velocity at T1, we will apply the definition of kinetic energy again. {TELL} 3. T: Let me just write the equation for you: $KE_1 = 0.5 * m * v_1^2$. {TELL} 4. T: From $KE_1 = 0.5 * m * v_1^2$, ...

Figure 1: Elicit vs. Tell

should be taken (Vygotsky, 1971; Aleven et al., 2004; Collins et al., 1989). Generally speaking, eliciting more information from the student will result in a more interactive tutorial dialogue. Figure 1 compared a pair of dialogues extracted from logs in this study. Both dialogues begin and end with the same tutor turn (lines 1 and 6 in (a) and 1 and 4 in (b)). In dialogue (a) the tutor chooses to elicit twice (lines 2-3 and 4-5 respectively). Dialogue (b), by contrast, *covers the same domain content* with two tell actions (lines 2 and 3). As a consequence, dialogue (a) is more interactive than (b).

In this paper, we quantify the *interactivity* of a dialogue via the Interactivity ratio (I-ratio) which we define as the number of elicitation decisions divided by the total number of elicit or tell decisions in a given dialogue. The higher this value, the more interactive the tutorial dialogue.

$$I - \text{ratio} = \frac{N_{Elicit}}{N_{Elicit} + N_{Tell}} \quad (1)$$

Unlike the monotonic and plateau hypotheses, validation of the tactical interaction hypothesis requires effective tutorial tactics. In most computer learning environments the pedagogical tutorial tactics are hard-coded rules designed to implement preexisting cognitive and/or pedagogical theories. Typically, these theories are considerably more general than the specific interaction decisions that designers must make. This makes it difficult to tell if a specific policy is consistent with the theory. Moreover, it is often difficult to empirically evaluate these tactics because the tutor's overall effectiveness depends upon many factors, such as the usability of the system, how easily the dialogues are understood, and so on. Ideally, several versions of a system are created, each employing different tutorial tactics. Data is then collected with human subjects interacting with these different versions of the system and the results are compared. Due to the high cost of experiments, however, only a handful of policies are typically explored. Yet, many other reasonable policies are possible.

In recent years, work on the design of dialogue systems has involved several data-driven methodologies. Among these, Reinforcement Learning (RL) has been widely applied (Singh

et al., 2002). In this work, rather than implementing pedagogical policies drawn from human experts or theories, we applied and evaluated RL to derive pedagogical tutorial tactics using pre-existing interactivity data.

General Approach

For this study, we induced two sets of tutorial tactics: the Normalized Gain (NormGain) tactics, derived with the goal of making tutorial decisions that contribute to students' learning, and the Inverse Normalized Gain (InvNormGain) tactics, induced with the goal of making less beneficial, or possibly useless, decisions. The two sets were then compared with human students on Cordillera (VanLehn, Jordan, & Litman, 2007), a Natural Language Tutoring System teaching students introductory college physics. Using Cordillera in lieu of human tutors allowed us to rigorously control the content and vary only the interactivity. In order to avoid artifacts due to imperfect natural language understanding, Cordillera incorporated a human wizard whose sole task was to rapidly match students' actual utterance to one of the expected student utterances displayed in a menu. The wizard made no tutorial decisions.

In the learning literature, it is commonly assumed that relevant knowledge in domains such as math and science is structured as a set of independent but co-occurring Knowledge Components (KCs) and that KC's are learned independently. A KC is "a generalization of everyday terms like concept, principle, fact, or skill, and cognitive science terms like schema, production rule, misconception, or facet" (VanLehn, Jordan, & Litman, 2007). For the purposes of tutoring, these are the atomic units of knowledge. It is assumed that a tutorial dialogue focusing on a single KC will not affect the student's understanding of any other KC. This is an idealization, but it has served developers well for many decades, and is a fundamental assumption of many cognitive models (Anderson, 1983; Newell, 1994). When dealing with a specific KC, the expectation is that the tutor's best policy for teaching that KC (e.g., when to Elicit vs. when to Tell) would be based upon the student's mastery of the KC in question, its intrinsic

difficulty, and other relevant, but not necessarily known, factors specific to that KC. In other words, an optimal policy for one KC might not be optimal for another. In this study, we focused on eight KCs. We induced eight policies and conducted eight tests of the three hypotheses, one per KC.

Later results indicated that on average the percentage of elicit prompts students received during the tutoring is more than 70% for both groups in this study, thus based on the standard set in (VanLehn, submitted) the tutorial dialogues reported here are well beyond the threshold of the level of interactivity afforded by conventional step-based ITs. Therefore, we expect that on each KC:

1. If the monotonic hypothesis is correct, the group that learned more would have a higher I-ratio.
2. If the interaction plateau hypothesis is correct, both NormGain and InvNormGain students would learn equally well regardless of interactivity difference.
3. If the tactical interaction hypothesis is correct and our RL-based tutorial tactics are indeed effective, NormGain students would learn more than InvNormGain peers regardless of interactivity difference.

First we will briefly describe how we apply machine learning to induce tutorial dialogue tactics. Then we will describe our study and its results.

Applying RL to Induce Tutorial Tactics

Much of the previous research on the use of RL to improve dialogue systems has typically used Markov Decision Problems (MDPs) (Sutton & Barto, 1998) to model dialogue data (Singh et al., 1999). An MDP formally corresponds to a 4-tuple (S, A, T, R) , in which: $S = \{S_1, \dots, S_n\}$ is a state space; $A = \{A_1, \dots, A_m\}$ is an action space represented by a set of action variables; $T : S \times A \times S \rightarrow [0, 1]$ is a set of transition probabilities $P(S_j|S_i, A_k)$, which is the probability that the model would transition from state S_i to state S_j after the agent takes action A_k ; $R : S \times A \times S \rightarrow R$ assigns rewards to state transitions. Finally, $\pi : S \rightarrow A$ is defined as a policy, which determines which action the agent should take in each state in order to maximize the expected reward.

The central idea behind our approach is to transform the problem of inducing effective pedagogical tactics into computing an optimal policy for choosing actions in an MDP. Inducing pedagogical tactics can be represented using an MDP: the states S are vector representations composed of relevant student-tutor interaction characteristics; $A = \{Elicit, Tell\}$ in this study, and the reward function R is calculated from the system's success measures and we used learning gains. Once the (S, A, R) has been defined, the transition probabilities T are estimated from the training corpus, which is the collection of dialogues, as: $T = \{p(S_j|S_i, A_k)\}_{i,j=1,\dots,n}^{k=1,\dots,m}$. More specifically, $p(S_j|S_i, A_k)$ is calculated by taking the number of times that the dialogue is in state S_i , the tutor took action A_k , and the dialogue was next in state S_j divided by the number of times

the dialogue was in S_i and the tutor took A_k . Once a complete MDP is constructed, a dynamic programming approach can be used to learn the optimal control policy π^* and here we used the toolkit developed by Tetreault and Litman (Tetreault & Litman, 2008).

In this study, the reward functions for inducing both the NormGain and the InvNormGain sets were based on Normalized Learning Gain (NLG) defined as: $NLG = \frac{posttest - pretest}{1 - pretest}$ because it measures a student's gain *irrespective of his/her incoming competence*. Here *posttest* and *pretest* refer to the students' test scores before and after the training respectively; and 1 is the maximum score. More specifically, the NormGain tutorial tactics induced by using the student's $NLG \times 100$ as the final reward while the InvNormGain ones was induced by using the student's $(1 - NLG) \times 100$ as the final reward. Apart from the reward functions, the two sets were induced using the same general procedure.

In order to learn a policy for each KC, we annotated our tutoring dialogues and action decisions based on which KCs a tutor action or tutor-student pair of turns covered ($\kappa \geq 0.77$ for each of the eight KCs). Additionally, we have mapped students' pre- and post-test scores to the relevant KCs for each test item. The rest of this section presents a few critical details of the process, but many others must be omitted to save space. Overall, the RL approach in this study differed from that of the previous study (M. Chi et al., 2009) in many aspects. First, we have three training corpora in this study: the Exploratory corpus collected in 2007, the Dich-Gain corpus collected in 2008, and a Combined training corpus. Second, in order to examine a range of possible tactics we included 50 features based upon six categories of features considered by previous research (Moore et al., 2004; Forbes-Riley et al., 2007) to be relevant. Additionally, we also used a different method of searching the power set of the 50 features. Finally we directly used the $NLG \times 100$ for inducing NormGain policies and $(1 - NLG) \times 100$ for inducing InvNormGain ones instead of dichotomizing the NLGs when inducing policies previously.

Figure 2 shows an example of a learned NormGain policy on one KC, "Definition of Kinetic Energy". The policy involves three features:

[StepDifficulty:] encodes a step's difficulty level. Its value is estimated from the students' log files based on the percentage of correct answers given on the step.

[TutorConceptsToWords:] which represents the ratio of the physics concepts to words in the tutor's dialogue. This feature also reflects how often the tutor has mentioned physics concepts overall.

[TutorAvgWordsSession:] The average number of words in the tutor's turn in this session. This feature reflects how verbose the tutor is in the current session.

MDP generally requires discrete features and thus all the continuous features need to be discretized. The top half of Figure 2 lists how each of the three features was discretized. For example, For StepDifficulty, if its value is above 0.38, it

[Feature:]			
StepDifficulty:	$[0, 0.38) \rightarrow 0;$	$[0.38, 1] \rightarrow 1$	
TutorConceptsToWords:	$[0, 0.074) \rightarrow 0;$	$[0.074, 1] \rightarrow 1$	
TutorAvgWordsSession:	$[0, 22.58) \rightarrow 0;$	$[22.58, \infty) \rightarrow 1$	
[Policy:]			
Elicit:	0:0:0	0:0:1	1:0:1 1:1:0 1:1:1
Tell:	0:1:0		
Else:	0:1:1	1:0:0	

Figure 2: A NormGain Policy on KC_{20} For ET Decisions

is 1 (difficult) otherwise, it is 0 (easy). The lower half of Figure 2 shows there are 8 rules learned: in 5 situations the tutor should elicit, in one situation it should tell; in the remaining 2 cases either will do. For example, when all three features are zero (which means when the step is easy, the tutor ratio of physics concepts to words so far is low, and the tutor is not very wordy in the current session), then the tutor should elicit as 0:0:0 is listed next to the [elicit]. As you can see, three features already provide relatively complex tutorial tactics and the induced policies were not like most of the tutorial tactics derived from analyzing human tutorial dialogues.

The resulting NormGain and InvNormGain policies were then implemented back into Cordillera yielding two new versions of the system, named NormGain-Cordillera and InvNormGain-Cordillera respectively. The induced tutorial tactics were evaluated on real human subjects to see whether the NormGain students would out-perform the InvNormGain peers.

Methods

Participants

Data were collected over a period of two months during the summer of 2009. Participants were 64 college students who received payment for their participation. They were required to have a basic understanding of high-school algebra. However, they could not have taken any college-level physics courses. Students were randomly assigned to the two conditions. Each took from one to two weeks to complete the study over multiple sessions. In total, 57 students completed the study (29 in the NormGain condition and 28 in the InvNormGain condition).

Domain & Procedure

The tutoring addressed work-energy problem solving from a first-year college physics course. The eight primary KCs were: the weight law (KC_1), definition of work (KC_{14}), Definition of Kinetic Energy (KC_{20}), Gravitational Potential Energy (KC_{21}), Spring Potential Energy (KC_{22}), Total Mechanical Energy (KC_{24}), Conservation of Total Mechanical Energy (KC_{27}), and Change of Total Mechanical Energy (KC_{28}).

All participants in the study followed the same procedures and used the same training and testing materials as were used when collecting the training corpora. More specifically, the participants all: completed 1) a background survey; 2) read a text covering the target domain knowledge; 3) took a pretest;

4) solved the same seven training problems in the same order on Cordillera; and 5) finally took a posttest. The pretest and posttest were identical. Except for following the policies (NormGain vs. InvNormGain), the remaining components of Cordillera, including the GUI interface, the same training problems, and the tutorial scripts, were identical for all students.

Grading

The tests contained 33 test items covering 168 KC occurrences. Each occurrence was graded by a single experienced grader who was not aware of the study condition from which it arose. These were then summed and normalized to the range of [0,1]. Other grading rubrics were also tried. They presented the same pattern of results as the ones presented next.

Results

No significant difference was found between the two conditions in terms of the total training time spent on Cordillera: $t(55) = 0.27, p = .79$. The NormGain group spent ($M = 259.98$ mins, $SD = 59.22$) and the InvNormGain group spent ($M = 264.57$ mins, $SD = 67.60$). For each student, Cordillera had made on average 260 decisions on whether to Elicit or to tell during the training and on a KC by KC basis, the number of such decisions varies from 4 on KC_1 to 72 on KC_{20} .

Learning Performance

First, we investigated whether students learned by training on Cordillera. A one-way ANOVA was used to test for learning performance differences between the pre- and posttests. Both groups made reliable learning gains from pre-test to post-test: $F(1,56) = 31.34, p = .000$ for the NormGain condition and $F(1,54) = 6.62, p = .013$ for the InvNormGain condition respectively. On a KC by KC basis, the NormGain conditions learned reliably on all the eight primary KCs while the InvNormGain learned reliably on five primary KCs save for KC_{14} , KC_{22} , and KC_{28} .

Next, we compared the learning performance between the two conditions. Random assignment appears to have balanced the incoming student competence across conditions. There were no statistically significant differences between the two conditions on the mathSAT scores nor in the pre-test scores: $t(55) = 0.71, p = .48$. On a KC by KC basis, no significant difference was found between the two conditions across all eight primary KCs except that on KC_{27} , the NormGain group score marginally higher than the InvNormGain group: $t(55) = 1.74, p = 0.088$ (see Table 1). In order to account for varying pretest scores, the adjusted Post-test scores were compared between the two conditions by running an ANCOVA using the corresponding pre-test score as the covariate.

The NormGain condition out-performed the InvNormGain on the overall adjusted posttest scores: $F(1,54) =$

Table 1: Between-Group Comparison on Pre-Test and Adjusted Post-Test Scores Across Primary KCs

KC	TestScore	NormGain	InvNormGain	Stat	d
KC_1	Pretest	0.42 (0.15)	0.39 (0.22)	$t(55) = 0.66, p = 0.51$	0.16
	Adjusted Posttest	0.64 (0.12)	0.54 (0.12)	$F(1, 54) = \mathbf{9.80}, p = \mathbf{0.0028}$	0.85
KC_{14}	Pretest	0.43 (0.23)	0.44 (0.25)	$t(55) = -0.17, p = 0.86$	-0.04
	Adjusted Posttest	0.65 (0.17)	0.53 (0.17)	$F(1, 54) = \mathbf{6.47}, p = \mathbf{0.014}$	0.72
KC_{20}	Pretest	0.38 (0.17)	0.37 (0.22)	$t(55) = 0.31, p = 0.76$	0.05
	Adjusted Posttest	0.67 (0.11)	0.58 (0.11)	$F(1, 54) = \mathbf{10.30}, p = \mathbf{0.002}$	0.83
KC_{21}	Pretest	0.45 (0.20)	0.43 (0.24)	$t(55) = 0.35, p = 0.72$	0.09
	Adjusted Posttest	0.75 (0.13)	0.65 (0.13)	$F(1, 54) = \mathbf{7.62}, p = \mathbf{0.008}$	0.78
KC_{22}	Pretest	0.42 (0.25)	0.39 (0.26)	$t(55) = 0.41, p = 0.68$	0.12
	Adjusted Posttest	0.63 (0.17)	0.51 (0.17)	$F(1, 54) = \mathbf{7.77}, p = \mathbf{0.007}$	0.72
KC_{24}	Pretest	0.46 (0.15)	0.41 (0.23)	$t(55) = 0.89, p = 0.38$	0.26
	Adjusted Posttest	0.64 (0.11)	0.58 (0.11)	$F(1, 54) = \mathbf{4.22}, p = \mathbf{0.045}$	0.56
KC_{27}	Pretest	0.53 (0.21)	0.42 (0.24)	$t(55) = 1.74, p = 0.088$	0.5
	Adjusted Posttest	0.74 (0.18)	0.63 (0.18)	$F(1, 54) = \mathbf{5.88}, p = \mathbf{0.019}$	0.62
KC_{28}	Pretest	0.37 (0.20)	0.36 (0.26)	$t(55) = 0.13, p = 0.90$	0.04
	Adjusted Posttest	0.53 (0.17)	0.47 (0.17)	$F(1, 54) = 1.61, p = 0.21$	0.36

10.689, $p = .002$, $d^1 = 0.86$. On a KC by KC basis, Table 1 summarize the comparisons on the pre-test and adjusted posttest scores between the two conditions. The third and fourth columns in Table 1 list the means and SDs σ of the NormGain and InvNormGain groups' pretest or adjusted posttest scores on the corresponding KC. The fifth column lists the corresponding statistical comparison and the sixth column lists the Cohen's d of the comparison. Table 1 shows that the NormGain condition out-performed the InvNormGain across all primary KCs (in bold) except for KC_{28} , on which no significant difference was found between the two groups.

I-ratios

We next investigated the interactive characteristics of the derived tutorial tactics by comparing the tutorial dialogues' I-ratios between the two groups. Surprisingly, there were no significant differences between the two groups on the overall I-ratio: $t(55) = -0.395$, $p = 0.694$. More specifically, we have $M = 0.758$, $SD = 0.073$ (maximum is 1) for the NormGain group and $M = 0.763$, $SD = 0.018$ for the InvNormGain group respectively.

However, once the results were examined on a KC by KC basis there were significant differences between the two groups on each of the eight primary KCs. Figure 3 shows that the NormGain condition was more likely to get elicits than the InvNormGain condition on KC_{14} , KC_{20} , KC_{21} , and KC_{22} ; and the InvNormGain condition was more likely to get elicits than the NormGain condition on KC_1 , KC_{24} , KC_{27} , and KC_{28} .

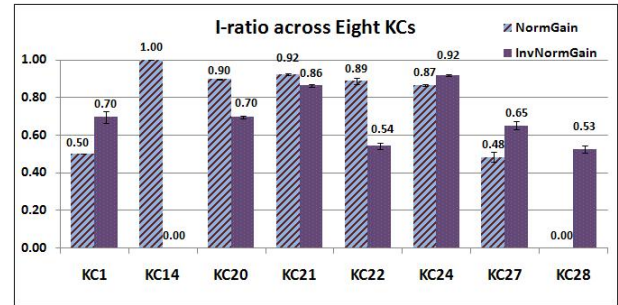


Figure 3: Compare I-ratio Across KCs

Examining The Three Interaction Hypothesis

The monotonic interactivity hypothesis states that more interactivity should lead to increased learning. Because the NormGain group learned more than the InvNormGroup across all eight KCs except KC_{28} , which was a null result, the NormGain group should also have a larger I-ratio on all seven KCs. From Figure 3, it was shown that this was *not* the case for KC_1 , KC_{24} and KC_{27} . Thus, our data are not consistent with the monotonic interactivity hypothesis.

The interaction plateau hypothesis states that increasing interactivity yields increasing learning until it hits a plateau, and further increases in interactivity do not cause noticeable increases in learning. The main difference between this hypothesis and monotonic interactivity hypothesis is once beyond a certain level of interactivity whether increasing interaction would impact students' learning gain or not. In order to test this hypothesis, we mainly focused on the six KCs (all but KC_{14} and KC_{28}). This is because on these six KCs both NormGain and InvNormGain groups' I-ratios were more than 48% (see Figure 3) which is well beyond the threshold of the level of interactivity afforded by conventional step-based ITSs based on the definition set in (VanLehn, submitted). If

¹Cohen's d , which is defined as the mean learning gain of the experimental group minus the mean learning gain of the control group, divided by the groups' pooled standard deviation.

the interaction plateau hypothesis is true, then the NormGain group should learn just as much as the InvNormGain group on each of the six KCs. Table 1 however shows that the NormGain group learned more than the InvNormGain group across all six KCs. Thus, the interaction plateau hypothesis is not consistent with our data.

Finally, the tactical interaction hypothesis states that interactivity does not increase learning unless they are governed by effective tutorial tactics. If this is true and all our derived RL-based policies were indeed effective, the NormGain group would learn more than the InvNormGain group across all KCs. This hypothesis was supported by seven of the KCs, and on KC₂₈ there was only an unreliable trend in the expected direction. Thus, of all three hypotheses, the tactical interaction hypothesis receives the most support from our data.

Discussion

Overall, our results inform the ongoing discussion of Socratic vs. didactic tutoring by suggesting that a tutor's success is not governed by *how often* they prompt or ask the students questions but *how well*. In particular, the reason human tutors so often failed to be more effective than simple, unoptimized dialogue-based tutors in those previous studies may be that effective policies for tutorial interaction are complex and not easily derived from the tutors' experience. This in turn suggests that an optimized dialogue-based tutoring system, such as NormGain-Cordillera, would be potentially even more effective than expert human tutors. Although controlling for content is difficult when human tutors are involved, testing this speculative hypothesis would certainly be interesting.

Finally, this study suggests that instead of using an overall tutorial tactics for all KCs, inducing KC-based tutorial tactics seems is necessary in that the induced tutorial tactics seems generated different tutorial decisions for different KCs in this study. Additionally, our results demonstrate that RL may be fruitfully applied to derive adaptive pedagogical tutorial tactics from student-computer interactivity data. However, this technique is not yet well understood. It is not completely clear to us, for instance, why our first attempt at inducing policies was suboptimal. In future work, we plan to explore the use of richer POMDP models, and do additional empirical evaluation of the RL approach.

Acknowledgments NSF (#0325054) supported this work and NSF (#SBE-0836012) supported its publication. We thank LRDC for providing all the facilities for this work.

References

- Aleven, V., Ogan, A., Popescu, O., Torrey, C., & Koedinger, K. R. (2004). Evaluating the effectiveness of a tutorial dialogue system for self-explanation. In *Intelligent tutoring systems* (Vol. 3220, p. 443-454). Springer.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, Mass.: Harvard University Press.
- Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13, 4-16.
- Chi, M., Jordan, P. W., VanLehn, K., & Litman, D. J. (2009). To elicit or to tell: Does it matter? In V. Dimitrova, R. Mizoguchi, B. du Boulay, & A. C. Graesser (Eds.), *Aied* (p. 197-204). IOS Press.
- Chi, M. T. H., Roy, M., & Hausmann, R. G. M. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science*, 32(2), 301-342.
- Chi, M. T. H., Siler, S., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, 25, 471-533.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. In L. B. Resnick (Ed.), *Knowing, learning and instruction* (p. 453-494).
- Evens, M., & Michael, J. (2006). *One-on-one tutoring by humans and machines*. Mahwah, NJ: Erlbaum.
- Forbes-Riley, K., Litman, D. J., Purandare, A., Rotaru, M., & Tetreault, J. R. (2007). Comparing linguistic features for modeling learning in computer tutoring. In R. Luckin, K. R. Koedinger, & J. E. Greer (Eds.), *Aied* (Vol. 158, p. 270-277). IOS Press.
- Fossati, D., Eugenio, B. D., Brown, C., & Ohlsson, S. (2008). Learning linked lists: Experiments with the ilist system. In *Intelligent tutoring systems* (Vol. 5091, p. 80-89). Springer.
- Katz, S., Allbritton, D., & Connelly, J. (2003). Going beyond the problem given: How human tutors use post-solution discussions to support transfer. *International Journal of AI and Education*, 13, 79-116.
- Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3), 239-264.
- Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of AI in Education*, 8(1), 30-43.
- Moore, J. D., Porayska-Pomsta, K., Varges, S., & Zinn, C. (2004). Generating tutorial feedback with affect. In V. Barr & Z. Markov (Eds.), *Flairs conference*. AAAI Press.
- Newell, A. (1994). *Unified theories of cognition*. Harvard University Press; Reprint edition.
- Reif, F., & Scott, L. A. (1999). Teaching scientific thinking skills: Students and computers coaching each other. *American Journal of Physics*, 67(9), 819-831.
- Rose, C. P., Moore, J. D., VanLehn, K., & Allbritton, D. (2001). A comparative evaluation of socratic versus didactic tutoring. In *Proc. of cognitive sciences society* (p. 869-874).
- Singh, S. P., Kearns, M. J., Litman, D. J., & Walker, M. A. (1999). Reinforcement learning for spoken dialogue systems. In S. A. Solla, T. K. Leen, & K.-R. Müller (Eds.), *Nips* (p. 956-962). The MIT Press.
- Singh, S. P., Litman, D. J., Kearns, M. J., & Walker, M. A. (2002). Optimizing dialogue management with reinforcement learning: Experiments with the njfun system. *J. Artif. Intell. Res. (JAIR)*, 16, 105-133.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning*. MIT Press Bradford Books.
- Tetreault, J. R., & Litman, D. J. (2008). A reinforcement learning approach to evaluating state representations in spoken dialogue systems. *Speech Communication*, 50(8-9), 683-696.
- VanLehn, K. (2006). The behavior of tutoring systems. *International Journal AI in Education*, 16(3), 227-265.
- VanLehn, K. (submitted). The two-sigma effect revisited: A meta-analysis of human tutoring and several types of computer tutoring.
- VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science*, 31(1), 3-62.
- VanLehn, K., Jordan, P., & Litman, D. (2007). Developing pedagogically effective tutorial dialogue tactics: Experiments and a testbed. In *Proc. of slate workshop on speech and language technology in education isca tutorial and research workshop*.
- Vygotsky, L. (1971). Interaction between learning and development. In T. M. Cole (Ed.), *In mind in society*. (p. 79-91). Harvard University Press: Cambridge Massachusetts.