Modeling Student Performance to Enhance the Pedagogy of AutoTutor

Tanner Jackson [1], Eric Mathews[1], King-Ip Lin[2], Andrew Olney [1], and Art Graesser[1]

[1]Department of Psychology
[2]Department of Computer Science
University of Memphis
Memphis, TN 38152, USA
{gtjacksn, emathews, davidlin, aolney, a-graesser}@memphis.edu

Abstract. The Tutoring Research Group from the University of Memphis has developed a pedagogically effective Intelligent Tutoring System (ITS), called AutoTutor, that implements conversational dialog as a tutoring strategy for conceptual physics. Latent Semantic Analysis (LSA) is used to evaluate the quality of student contributions and determine what dialog moves AutoTutor gives. By modeling the students' knowledge in this fashion, AutoTutor successfully adapted its pedagogy to match the ideal strategy for students' ability.

1 Introduction

Recent educational technologies attempt to engage users in the learning process in an active manner and to simultaneously address the needs of the user through interactive displays and discourse. Tutorial dialog systems are interactive, conversing directly with the user by providing hints, corrections, and support throughout the entire learning process. The advantages of such an educational technology tool are numerous, but one of the foremost advantages lies in the fact that the educational tool can be personalized and tailored to the needs of the individuals, who vary in ability, background, and learning styles.

Research continually points to the fact that "Designers of collaborative HCI face the formidable task of writing software for millions of users, at design time, while making it work as if it were designed for each individual, at use time" [2]. Once emerging educational technologies can intelligently adapt to the need of the user, the learning process presumably can be optimized to achieve the greatest learning gains in the shortest amounts of time.

This paper outlines our efforts at the University of Memphis to build an adaptive educational technology that optimizes the learning process. We provide a brief overview of our system, AutoTutor, its user modeling, and evaluation data.

2 Overview of AutoTutor

From a large amount of research on human-human tutoring conducted by our group and others, we have determined the underlying principles of tutoring that are found in our intelligent tutoring system called AutoTutor [4], [9], [11], [12]. AutoTutor engages the learner in a conversation while simulating the dialog moves of human tutors. The user and AutoTutor collaboratively improve the quality of the student's contributions to problems and questions while participating in a mixed-initiative dialog, distinguishing it from mere information delivery systems. During the conversation, AutoTutor implements a constructivist-based tutoring strategy. That is, in order for AutoTutor to believe that a student knows something, the student must actually state it during the conversation.

AutoTutor has recently been developed to help college students learn Newtonian conceptual physics (previous versions handled computer literacy). AutoTutor has been described in several previous publications, so it will not be detailed here (See [3], [5], [8], [10], [13]).

AutoTutor's dialog moves lead the student towards the correct ideas in answering the question, all the while trying to let the student provide as much of the information as possible. The types of dialog moves used by AutoTutor vary on a continuum of specificity from the student supplying information (pumps and hints), to information delivery on the part of AutoTutor (prompts and assertions). Within AutoTutor, pumps are the least specific form of dialog elicitation and they provide the least amount of information delivery (e.g., "Tell me more", "What else?"). In contrast, assertions have the highest amount of information delivery because the correct piece of the answer is directly expressed.

2.2 Tailoring the Tutoring within AutoTutor

Developers of intelligent tutoring systems continually make use of discourse planning models that adapt to the user's ability [1], [2], [6], [7], [14]. AutoTutor is no different in this respect, for it tailors its tutoring to the student in two specific ways: pedagogical feedback and dialog move selection. Without properly modeling the ability of the user within AutoTutor, the task of tailoring tutoring would be impossible. If AutoTutor operated on the assumption that all users were the same, we would find one of two things happening: (1) High ability students being frustrated and bored by AutoTutor being pedantic or (2) Low ability students being overly challenged and frustrated by getting nothing correct.

Additionally, feedback is a necessary component of tutoring; without it the user cannot engage in needed metacognitive analyses to determine whether they are providing good or bad material. In AutoTutor, we should find that higher ability students receive more positive feedback (positive correlation with ability) and less negative feedback (negative correlation with ability), in contrast to lower ability students where we should find the opposite to be true.

AutoTutor needs to be able to adapt the conversation at various levels of specificity, accommodating student ability. For instance, high ability users should receive less content delivery and more general requests for information, which require active knowledge construction. Hence, in AutoTutor we should find that high ability users receive more pumps and hints, but fewer assertions.

3 The Experiment

The learners were 24 undergraduate students recruited from the University of Memphis, University of Pittsburgh, Christian Brothers University, and Rhodes College. Each participant was given a pre- and a post-test consisting of 4 different essay questions (8 total) and 40 different selected items from the Force Concept Inventory (80 total from FCI). The Force Concept Inventory is an established multiple-choice test in physics, from which we selected questions addressing Newtonian physics. Though the pre- and post-tests were administered for each participant, they are not factored into the user modeling in AutoTutor.

The focus of this study is on the actual tutoring sessions of AutoTutor, during which students completed the 10 conceptual physics dialog. The 10 problems were split into two sessions, one week apart, for approximately 3 hours total. During the first session, each participant took a pre-test and then spent the rest of their time (approximately 70 min.) going through the first five conceptual physics problems with AutoTutor. At the second session, each participant spent about the same amount of time working with AutoTutor through the second five conceptual physics problems, and then took a post-test.

4 Determining Student Ability

To determine student ability from this experiment we examined the participants' pretest scores. Specifically we looked at their FCI pretest ability. We used a median split to distinguish between high and low ability. For all further analyses, high ability students are those who had a proportion of correct answers higher than .63, while low ability students are those who had a proportion of correct answers lower than .63.

5 Experimental results and discussion

We examined the application of AutoTutor's user modeling capabilities by analyzing correlations between student ability and the proportions of various dialog moves. Specifically, we examined the distribution of dialog moves (pumps, hints, prompts, and assertions) and the proportions of positive and negative feedback.

An analysis of the dialog moves was performed that correlated student ability with each dialog move proportion. As briefly mentioned earlier, we expected the following ordering of proportions for high ability students, to the extent that the student is supplying information, as opposed to the tutor: pumps > hints > prompts > assertions. The analysis resulted in the following correlations: pumps (r = .49), hints

(r = .24), prompts (r = -.19), and assertions (r = -.40). As you can see, AutoTutor tended to deliver pumps and hints to high ability students, but was forced to deliver prompts and assertions to low ability students. In this sense, AutoTutor's user modeling components are valid. There was also a significant positive correlation (r = .38) between positive feedback and objective physics knowledge, FCI score, and a negative correlation in the case of negative feedback (r = -.37).

The correlation of r = .49, p= .016, between student ability and the proportion of pumps demonstrates that the high ability students receive more pumps than anything else. The fact that pumps are correlated with ability demonstrates that AutoTutor's user modeling properly tracks a good student's behavior and ability and then properly adapts by 1) using less specific dialog moves and 2) moving faster through the material instead of exhausting all possible dialog moves for a particular piece of information.

Similarly, the marginally significant negative correlation between student ability and the proportion of assertions, r= -0.40, p= .055, demonstrates that low ability students require a higher proportion of assertions. This means that AutoTutor was able to correctly identify those students who could not actively construct the knowledge on their own, and who needed more of the information to be provided for them. These correlations between the pretest ability and the proportion of dialog moves show us that AutoTutor is doing an appropriate job of modeling user ability.

The results from the short feedback analysis follow the same trend as the dialog moves. We found that the high ability students received a higher proportion of positive feedback from AutoTutor, while the low ability students received more negative feedback. This means that AutoTutor did an adequate job of discriminating the contributions from low versus high ability students, and was able to respond with appropriate levels of pedagogical feedback.

6 Concluding Remarks

These analyses of student ability correlated with dialog move and short feedback proportions provide evidence of effective user modeling within AutoTutor. This recent analysis of AutoTutor supports the claim that it does an effective job of modeling user ability, and adapting accordingly with appropriate pedagogical strategies. Although some of these phenomena are often easily implemented within computer systems, it is not so easily implemented within natural language dialog. The unique contribution of this work is the fact that we have a natural language system that extracts semantic intent and properly models the user.

7 Acknowledgements

The Tutoring Research Group (TRG) is an interdisciplinary research team comprised of approximately 35 researchers from psychology, computer science, physics, and education (Visit http://www.autotutor.org). This research conducted by the authors and the TRG was done in collaboration with members of the CIRCLE Group at the

University of Pittsburgh, and was supported by grants from the National Science Foundation (SBR 9720314 and REC 0106965) and the Department of Defense Multidisciplinary University Research Initiative (MURI) administered by the Office of Naval Research under grant N00014-00-1-0600. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of ONR or NSF.

References

1. Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4*, 167-207.
2. Fischer, G. (2001). User modeling in human-computer interaction. *User-Modeling-and-User-Adapted-Interaction, 11(1-2)*: pp 65-86.
3. Foltz, P.W. (1996). Latent semantic analysis for text-based research. *Behavior Research Methods, Instruments, and Computers, 28,* 197-202.
4. Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialog patterns in naturalistic one-on-one tutoring. *Applied Cognitive Psychology, 9,* 359-387.
5. Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25,* 259-284.
6. Lepper, M. R., Woolverton, M., Mumme, D.L., & Gurtner, J.L. (1991). Motivational techniques of expert human tutors: Lessons for the design of computer-based tutors. In S.P. Lajoie & S.J. Derry (Eds.), *Computers as cognitive tools* (pp. 75-105). Hillsdale, NJ: Erbaum.
7. Merrill, D. C., Reiser, B. J., Ranney, M., & Trafton, J. G. (1992). Effective tutoring techniques: A comparison of human tutors and intelligent tutoring systems. *The Journal of the Learning Sciences, 2,* 277-305.
8. Person, N. K., Graesser, A. C., Kreuz, R. J., Pomeroy, V. & the Tutoring Research Group. (2000). *Simulating human tutor dialog moves in AutoTutor*. Submitted to *International Journal of Artificial Intelligence in Education*.
9. Person, N. K., Graesser, A. C., Magliano, J. P., & Kreuz, R. J. (1994). Inferring what the student knows in one-to-one tutoring: The role of student questions and answers. *Learning and Individual Differences, 6*, 205-29.
10. Person, N. K., Klettke, B., Link, K., Kreuz, R. J., & the Tutoring Research Group (1999). The integration of affective responses into AutoTutor. *Proceeding of the International Workshop on Affect in Interactions* (pp. 167-178). Siena, Italy.
11. Person, N. K., Kreuz, R. J., Zwaan, R., & Graesser, A. C. (1995). Pragmatics and pedagogy: Conversational rules and politeness strategies may inhibit effective tutoring. *Cognition and Instruction, 13,* 161-188.
12. Putnam, R. T. (1987). Structuring and adjusting content for students: A study of live and simulated tutoring of addition. *American Educational Research Journal, 24,* 13-48.
13. Wiemer-Hastings, P., Graesser, A. C., Harter, D., & the Tutoring Research Group (1998). The foundations and architecture of AutoTutor. *Proceedings of the 4th International Conference on Intelligent Tutoring Systems* (pp. 334-343). Berlin, Germany: Springer-Verlag.
14. Zukerman, I. & McConachy, R. (2001) WISHFUL: A discourse planning system that considers a user's inferences. Computational Intelligence. 17(1): 1-61