What Should the Tutor Do When the Student Cannot Answer a Question?

Yujian Zhou[1], Reva Freedman[2*], Michael Glass[1],
Joel A. Michael[3], Allen A. Rovick[3], Martha W. Evens[1]

[1]Department of CSAM
Illinois Inst. of Technology
10 W. 31st Street 236–SB
Chicago, IL 60616

[2]LRDC #819
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

[3]Department of Physiology
Rush Medical College
1750 W. Harrison Street
Chicago, IL 60612

zhouyuj@charlie.iit.edu, freedrk+@pitt.edu, glass@steve.iit.edu,
{jmichael, arovick}@rush.edu, csevens@minna.cns.iit.edu
http://www.csam.iit.edu/~circsim

Abstract

In this paper we describe how to simulate the behavior of a human tutor when the student cannot answer a question in a dialogue-based tutoring system. This paper describes an implementation of our analysis of human tutoring behavior using the simple plan-based framework of our tutoring system CIRCSIM-Tutor v. 2. By carefully choosing the retry strategies according to our categorization of the student's answer, our machine tutor can simulate some sophisticated human tutoring behaviors, such as hinting and "grain of truth" responses.

Introduction

We are building an intelligent tutoring system called CIRCSIM-Tutor (CST) designed to help medical students understand the negative feedback system that controls blood pressure and learn to solve problems in physiology. The system presents the student with a description of a physiological change and asks for predictions about the effect of that change on seven important physiological parameters. Then it conducts a dialogue with the student to correct the errors in the predictions. The interface of the dialogue is free-text input and output.

In order to help the student find the desired answer without being told, it is important for an intelligent tutoring system to have sophisticated retry strategies available when the student cannot answer a question. Among these strategies hinting is particularly interesting because it is frequently used by human tutors in one-on-one dialogue.

This work was supported by the Cognitive Science Program, Office of Naval Research under Grant No. N00014–94–1–0338, to Illinois Institute of Technology. The content does not reflect the position or policy of the government and no official endorsement should be inferred.

*This work was begun while Reva Freedman was at the Illinois Institute of Technology.

It is especially important for CIRCSIM-Tutor to be able to choose appropriate retry strategies, provide meaningful hints, and help students discover the desired answer because it is designed to simulate the behavior of human tutors. In April 1998, 22 first-year medical students from Rush Medical College used our latest upgrade to CST v. 2. We observed that students gave several different types of unexpected answers. Some of these answers showed that they may have some understanding of the concept being taught. As a result it may be inappropriate just to tell them the correct answer. Observation of expert tutors tells us that they usually try to pick up any useful information that shows the student's understanding of the concept and use this information to help the student find the desired answer.

To enrich the retry capability in CST v. 2, we simplified earlier analyses of human tutoring transcripts and implemented them in the current planning framework. This paper shows that by carefully choosing the retry strategy according to our categorization of the student's answer, the machine tutor can simulate some sophisticated human tutoring behavior.

Hints and Other Retry Strategies in Machine and Human Tutoring

Hume et al. (1996) studied the use of hints by experienced tutors in the hope of formulating a strategy for using hints in an ITS. They observed that human tutors frequently use hints as a pedagogical tactic. However, the theory of hints under their framework is too broad and too hard to simulate in CST v. 2 for the following reasons:

1. Hinting is a very subtle tactic and we do not yet fully understand how it is done.
2. The form of hints is very flexible.
3. The content of hints is context sensitive.

It has been observed that most hints are used to help students find the desired answer when they fail to answer a question. For this reason we decided to to implement hints as one of the retry strategies in CST v. 2. In this section we will describe the planner in CST v. 2, then discuss how we can implement hinting as one of the retry strategies under the current planning model. We will also describe the analysis of human tutoring transcripts which we used to identify specific hints and methods of hinting.

The original instructional planner in CST v. 2 was developed by Woo (1991). It includes two levels of planners, a lesson planner and a discourse planner. The lesson planner generates lesson goals and decomposes them into discourse tasks. The discourse planner is responsible for controlling interactions between the tutor and the student. Most of the discourse tasks are executed as questions. When the student cannot answer one of these questions, the tutor needs to find a way to complete the task and carry out the next step. One way to complete a task is simply to give the correct answer. Another way, the alternative our human tutors usually take, is to give the student another chance to find the desired answer.

The ability to retry a task in CST v. 2 is strictly limited. It can construct and ask a question rather than giving the correct answer, and it can optionally produce some declarative material before the question. The original release did not use these capabilities to the fullest; it only generated one type of hint and a few types of questions.

Freedman (1996) modeled the behavior of our expert human tutors as a hierarchical structure of tutoring goals represented as schemata. Modeling this structure requires a more sophisticated planner. In addition to the capabilities in CST v. 2, her planner can back up to a higher goal and retry from any level, not just at the goal where the student erred. Furthermore, it can add interactive subplans to an existing plan, not just declarative material.

Kim et al. (1998) marked up transcripts of human tutors according to Freedman's schemata and discovered several interesting new tutorial patterns. Freedman et al. (1998) applied machine learning techniques to the annotated transcripts as a first step toward discovering rules used by human tutors to choose a response in a given situation.

CIRCSIM-Tutor v. 3, which is currently under development, is based on Freedman's planner. Although the v. 3 planner is required to implement all of the patterns described in these papers, many of them can be used to improve CST v. 2. In some cases we have been able to implement a simplified version in the CST v. 2 framework. Below we will discuss in detail how to adapt these theories for CST v. 2 and how to implement them. Then we will show some sample outputs from the improved version.

Most of the text that CST v. 2 produces, including all the examples shown below, is taken unchanged or with only slight modification from language used by human tutors.

Categorizing Student Answers

Human tutors often use different response strategies for different categories of student answers. If the student answer is correct, the tutor will give a positive acknowledgment and move to the next task. If the student answer is not correct, then depending on the category to which the answer belongs (and other factors), the tutor may just give the correct answer, or retry the task.

The original CST v. 2 differentiated several categories of student answers; Kim and others identified additional categories in transcripts (Kim et al., 1998). But to implement our retry strategies we needed to extend this list. By adding misconceptions, "grain of truth" information, and information for deriving near misses to the knowledge base, we enabled CIRCSIM-Tutor to recognize additional categories of student answers. Below are the categories used by the updated CST v. 2:

1. Correct
2. Partial answer, i.e., the answer is part of the correct answer
3. Near miss answer, which is pedagogically useful but not the desired answer (Glass, 1997)
4. "I don't know" answer
5. "Grain of truth" answer, where the student gives an incorrect answer, but also indicates a partially correct understanding of the problem (Woolf, 1984)
6. Misconception, a common confusion or piece of false knowledge about the concept being tutored
7. Other incorrect answers
8. Mixed answers, i.e., a combination of answers from the other categories

Responses after the Student Cannot Answer a Question

After categorizing possible student answers, we devised algorithms for each category to decide what to say and what to do next. There are two decisions the planner needs to make when the student cannot answer a question:

1. Decide how to satisfy the goal—whether to give the correct answer or retry the task.
2. If retry, then decide what kind of retry and what content to choose.

To make the first decision there are several types of information we could use: the student's performance, the category of the student's answer, the available tutoring options in the knowledge base, and the tutoring history (e.g., which options have already been tried). Currently the most important factor is the category of the student's answer. We almost always prefer to give the student a second chance if that option is available in the knowledge base.

The second decision depends on the category of the student's answer. Below we will describe the retry algorithms for some categories of student answers. They were developed with the goal of simulating our human tutors' behavior as closely as possible.

Retry after a Partial Answer

Since the student has already found part of the correct answer, it is important that the tutor acknowledge this part first. This acknowledgment is usually short. Then the tutor helps the student focus on the missing part. To do so, the tutor will give a hint about the missing part, then ask the student about it. Our machine tutor simulates this behavior. For example:

Tutor> What are the determinants of Stroke Volume?
Student> Inotropic State.
Tutor> Right, Inotropic State is one determinant of Stroke Volume. The other determinant represents preload. What is the other determinant of Stroke Volume?

Retry after a Near Miss Answer

As discussed by Kim et al. (1998), our human tutors' response strategy is to lead the student from a near miss answer, which is a step on the right path, to the desired answer. We adapt this observation to CST v. 2. So our machine tutor will first acknowledge that the near miss answer is correct and then ask a question to lead the student toward the desired answer from the near miss answer. For example:

Tutor> What are the determinants of Stroke Volume?
Student> preload *(This is true, but preload is not one of the variables on the student's worksheet.)*
Tutor> Right, preload determines Stroke Volume. What parameter in the prediction table represents preload?

Retry after a "Grain of Truth" Answer

Woolf (1984) discussed how to pick up and reinforce "grains of truth" in her MENO-Tutor. Her tutor's response was to pick out the "grain of truth" in the student's answer, acknowledge it, then ask a question relating the "grain of truth" to the correct answer. Similar behaviors appear in our tutoring transcripts. We isolated several "grain of truth" answers from human tutoring transcripts and the corresponding response patterns.

Our current implementation recognizes a fixed set of "grain of truth" answers. We use a simpler version of Woolf's schema, limiting the follow-up question to retrying the original question in different words. So, after a "grain of truth" answer, we pick up the grain of truth, acknowledge it briefly, then rephrase the earlier question. For example:

Tutor> What are the determinants of Stroke Volume?
Student> Heart Rate and Cardiac Output. *(i.e., SV = HR * CO.)*
Tutor> Well, it is true that CO = SV * HR. But what I was asking is what determines how much blood is ejected from the heart each time it beats (the SV)?

Retry after an "I Don't Know" Answer

Kim et al. (1998) model our human tutors' response after "I don't know" as a nested schema incorporating a directed line of reasoning, while Hume marked the same text as hints (Hume et al., 1996). To simulate this tutoring behavior in CST v. 2 without the use of nested plans, which our current planner does not support, we try to give hints that offer information relating the variable in question to the desired answer. So the machine tutor's response will be a hint and a follow-up question to retry the task. For example:

Tutor> What determines Central Venous Pressure?
Student> I don't know.
Tutor> Remember, the central venous compartment is very compliant. Again, what determines Central Venous Pressure?

Retry after Other Incorrect Answers

The response strategy to other incorrect answers is similar to the strategy after an "I don't know" answer. But in this case the machine tutor gives a negative acknowledgment first. For example:

Tutor> What determines Central Venous Pressure?
Student> TPR.
Tutor> No. Remember, the central venous compartment is very compliant. Again, what determines Central Venous Pressure?

Response after a Misconception

It is important for CIRCSIM-Tutor to recognize common student misconceptions and provide appropriate remedial responses. We isolated several common student misconceptions from human tutoring transcripts and from interviews with human tutors. For example, the student often confuses the mechanism that controls Inotropic State with the Frank-Starling effect (see the following dialogue example). In our transcripts, the human tutors typically give lengthy explanations about the misconceptions, so the machine tutor will give a similar explanation first and then give the correct answer. For example:

Tutor> By what mechanism is Inotropic State controlled?
Student> It is filling.
Tutor> Well, you are confusing the Frank-Starling effect (increased filling gives greater output) and contractility. Inotropic State (contrac-

tility) is the force of contraction at any given filling or at any end diastolic volume. An increase in sympathetic stimulation of the myocardium increases the Inotropic State of the muscle. So, it is neurally controlled.

Retry after Mixed Answers

The student answer can be a combination of different categories, for example, partial answer plus a near miss, or correct answer plus a misconception. To handle these answers, the machine tutor generally acknowledges any correct part, then handles errors and gives correct answers. Unless there is a near miss answer, our system does not usually ask any follow-up questions. Here is a sample output from the machine tutor showing the near miss case:

Tutor> What are the determinants of Mean Arterial Pressure?
Student> CO and Ra. *(one correct, one near miss)*
Tutor> Right, CO is one determinant of MAP. Think about what parameter in the prediction table relates to Arterial Resistance. What is the other determinant of Mean Arterial Pressure?

Dynamically Generating Hints in Different Situations

After deciding to give a hint, it is necessary to determine how to make the machine tutor generate hints that are meaningful and sensitive to different situations. To be pedagogically helpful, hints must reveal some information, must not conflict with the correct part of the student's answer or any domain knowledge, and at the same time, must be sensitive to the information that is included in the student answer. So our hints cannot be entirely preformulated and stored in a hint list. Compared to hints generated in other systems, such as Andes (Gertner et al., 1998) and Sherlock II (Lesgold et al., 1992), the hints in CIRCSIM-Tutor need to be based more on discourse context since our machine tutor is conducting a dialogue. Andes generates individual hints, while Sherlock II generates a paragraph after the conclusion of the tutoring session.

After carefully categorizing student answers and the follow-up retry strategies, it is easier to decide what kind of information is needed for the hints. For example, if the answer is a partial answer, hints should focus on the information related to the missing part, whereas if the answer is "I don't know", then the machine tutor may just offer a piece of information about the variable in question. It is difficult to find a general form for all hints (although the original CST v. 2 always offers the hint "consider the value of <the desired answer>"). But for each category of retry strategy, it is possible to generate hints from some general templates. So for each category we built a small hint algorithm that is executed when a hint is needed in that case.

Here is an example that shows how we generate a hint in the case of a partial answer. First we search the domain knowledge base to find possible information to give in the hint and then we choose an appropriate template according to the information in the hint.

Algorithm to search for possible hint information:

First find if there is a parameter in the answer that is related to the missing parameter. If so, present that information.
Else find if there is a parameter in the answer that has some of the same characteristics as the missing parameter. If so, present the similarity. For example: "Like HR, the other determinant is neurally controlled."
Else search the domain knowledge base and find a feature which distinguishes the missing answer, for example, whether a parameter is controlled by the nervous system, or whether it is used in one of the fundamental equations.

We have several hint templates related to different types of information. For example:

"Like <the related variable in the student's answer>, the other determinant is also related to <the anatomy object>."

Here are some hints generated by our machine tutor. In the first two examples, the student gave the answer SV.

"The other determinant is affected by SV." *(pointing to the relationship of SV and the desired answer)*

"Like SV, the other determinant is also related to the heart." *(pointing to the anatomy object related to the desired answer, while also remembering that SV is related to the same anatomical object)*

"Tell me, MAP = TPR * what?" *(pointing to a related equation)*

Sample Outputs

Although the scheme described above is a straightforward way to recover from a task failure, it allows us to tailor the system's response to the student's needs and generate a variety of complex dialogues containing question-and-answer series. Some of these have been categorized as "long hint strings" by Hume (Hume et al., 1996). Below are some output samples produced by CST v. 2 using this method.

Example 1: Tutor responds to near miss and partial answer.

Tutor> What are the determinants of Stroke Volume?
Student> edv *(End diastolic volume, near miss to one*

of the two required parts of the answer.)

Tutor> And what determines EDV? *(Ask question based on the near miss.)*

Student> rap *(Right atrial pressure, correct answer to the immediate question but still not an answer to the original question.)*

Tutor> And what determines RAP? *(Tutor recognizes RAP as another near miss, so asks a question based on it.)*

Student> cvp *(Correct answer to the immediate question and one of the two answers required to the original question.)*

Tutor> Right, Central Venous Pressure determines EDV, therefore determines Stroke Volume. The other determinant is neurally controlled. What is the other determinant of Stroke Volume? *(Acknowledge the partial answer and give a hint for the missing part.)*

Student> Inotropic State. *(Correct answer.)*

Tutor> Right, Inotropic State is the other determinant. Stroke Volume is determined by Central Venous Pressure and Inotropic State. *(Acknowledge the second part of the answer and restate the whole answer.)*

Example 2: Tutor responds to incorrect answer.

Tutor> If afterload, i.e. MAP, decreased, which is what happened in this case, what should SV be? What is the correct value of Stroke Volume?

Student> It goes down. *(Incorrect answer.)*

Tutor> No, the value of Stroke Volume is not decreased. Remember if afterload decreased, the heart can pump blood out more easily. What is the correct value of Stroke Volume? *(Give negative acknowledgment and a hint.)*

Student> Up.

Tutor> Correct, the value of Stroke Volume is increased.

Conclusion and Further Studies

This paper describes how our machine tutor tailors its responses to a student when he or she cannot answer a question. By carefully categorizing student answers and adapting earlier analyses of human tutoring transcripts to the current planning framework, we have implemented a variety of retry strategies in CST v. 2. Using this scheme we have been able to simulate some sophisticated human tutoring behaviors, such as hinting.

We are currently developing a new version of CIRCSIM-Tutor, CST v. 3, which will have a more sophisticated tutorial planner than CST v. 2 (Freedman and Evens, 1996). We intend to adapt some of the useful retry strategies to the new version.

Acknowledgments

We wish to thank Jung Hee Kim, Stefan Brandle, and other members of the CIRCSIM-Tutor project.

References

Freedman, R. (1996). Interaction of Discourse Planning, Instructional Planning and Dialogue Management in an Interactive Tutoring System. Ph.D. diss., Dept. of EECS, Northwestern Univ.

Freedman, R. and Evens, M. (1996). Generating and Revising Hierarchical Multi-turn Text Plans in an ITS. *Intelligent Tutoring Systems: Third International Conference (ITS '96)*, Montreal, 1996. (Springer-Verlag Lecture Notes in Computer Science, 1086.) Berlin: Springer-Verlag, pp. 632–640.

Freedman, R., Zhou, Y., Kim, J., Glass, M., and Evens, M. (1998). Using Rule Induction to Assist in Rule Construction for a Natural-Language Based Intelligent Tutoring System. *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, Madison, WI, 1998, pp. 362–367.

Gertner, A., Conati, C., and VanLehn, K. (1998). Procedural Help in Andes: Generating Hints using a Bayesian Network Student Model. *Proceedings of the 15th National Conference on Artificial Intelligence*, Madison, WI, 1998, pp. 106–111.

Glass, M. (1997). Some Phenomena Handled by the CIRCSIM-Tutor Version 3 Input Understander. *Proceedings of the Tenth Florida Artificial Intelligence Research Symposium,* Daytona Beach, FL, 1997, pp. 21–25.

Hume, G., Michael, J., Rovick, A., and Evens, M. (1996). Hinting as a Tactic in One-on-One Tutoring. *Journal of Learning Sciences* 5(1): 32–47.

Kim, J., Freedman, R., and Evens, M. (1998). Responding to Unexpected Student Utterances in CIRCSIM-Tutor v. 3: Analysis of Transcripts. *Proceedings of the Eleventh Florida Artificial Intelligence Research Symposium*, Sanibel Island, FL, 1998, pp. 153–157.

Lesgold, A., Katz, S., Greenberg, L., Hughes, E., and Eggan, G. (1992). Extensions of Intelligent Tutoring Paradigms to Support Collaborative Learning. In Dijkstra, S., Krammer, H., and van Merrienboer, J., eds., *Instructional Models in Computer-based Learning Environments*. (NATO ASI Series, series F: Computer and System Sciences, 104) Berlin: Springer-Verlag, pp. 291–311.

Woo, C. (1991). Instructional Planning in an Intelligent Tutoring System: Combining Global Lesson Plans with Local Discourse Control. Ph.D. diss., Dept. of CSAM, Illinois Institute of Technology.

Woolf, B. (1984). Context-Dependent Planning in a Machine Tutor. Ph.D. diss., Dept. of Computer and Information Science, University of Massachusetts at Amherst. COINS Technical Report 84–21.