The Use of Hints by Human and Computer Tutors: The Consequences of the Tutoring Protocol

Gregory Hume
Mathematics and Computer Science Department
Valparaiso University, Valparaiso, IN 46383
(219)-464-5156
ghume@exodus.valpo.edu

Joel Michael
Department of Physiology
Rush Medical College, Chicago, IL 60612
(312)-942-6426
jmichael@steve.iit.edu

Allen Rovick
Department of Physiology
Rush Medical College, Chicago, IL 60612
(312)-942-6567
aar@steve.iit.edu

Martha Evens
Computer Science Department
Illinois Institute of Technology, Chicago, IL 60616
(312)-567-5150
mwe@math.nwu.edu

Abstract: Our study of two expert human tutors reveals that, while they employ many tactics, they (virtually) always try hinting to remedy problems. Hints are occasionally mentioned in the Intelligent Tutoring System (ITS) literature but there has been no systematic study of this phenomenon. Our ITS, CIRCSIM-Tutor (CST), is designed to assist first year medical students to learn to reason about blood pressure regulation. Our study of human tutoring sessions is the basis for the design of CST. Our protocol (both human and computer tutoring sessions) begins with an explanation of a problem (a disturbance to the circulatory system). The student then makes predictions about the ensuing qualitative causal effects. This is followed by an interactive dialogue. This protocol allows for the student to ask questions, although this rarely happens. Other ITSs and automated training systems require that the student work on a problem and explicitly ask for help. The ITS literature has examples of how other systems differ in tutoring protocol, knowledge domain and student population. How do these differences affect the use of hints?

Introduction

The original motivation of our study of human tutoring sessions was our design of CIRCSIM-Tutor (CST) [Kim et al., 1989], an Intelligent Tutoring System (ITS). We have observed that human tutors use an identifiable algorithm to select tactics [Hume et al., 1995]. What was most interesting was our observation that our tutors constantly provide hints [Hume et al., 1993], [Hume et al., 1996a]. Our results suggest there is a definite relationship between a tutor's evaluation of the student (in an ITS, this is referred to as the student model) and the selection of a tactic [Hume et al., 1995], [Hume et al.,1996b]. In addition, we believe that there are other factors influence the selection of tactics. Specifically, we seek a better understanding of the relationship between the tutoring protocol and the use of hints. This paper will (1) describe our experiments, (2) pose general questions about student modelling, the use of hints and the tutoring protocol, and (3) make an argument for a coordinate effort by the ITS community to address these issues.

Methodology

Fifty-eight human tutoring sessions, conducted with tutor and student using PCs in different rooms, have been recorded using a computer program called CDS [Li et al., 1992]. The transcripts of these dialogues have been analyzed in our attempt to understand:

1. how human tutors plan the tutorial interaction,
2. how human tutors respond to student initiatives,
3. what knowledge needs to be represented in the knowledge base,
4. what sublanguage is used by tutors and students,
5. how tutors generate acknowledgments, and
6. how and when tutors generate hints.

The subjects for the tutoring experiments were first year medical students at Rush Medical College. They ranged in age from 21 to 37 years with a mean age of 25 years; 32 were female and 26 were male. We believe that all subjects, while paid volunteers, had the expectation that they would learn something from their participation in the experiment. Examples of dialogue (human and computer) in this paper have been edited slightly to improve readability

Tutoring Domain and Protocol

The domain of CST, and hence our human tutoring experiments, is cardiovascular (CV) physiology. Specifically, both CST and out tutors assist students to reason about the qualitative causal effects to the human circulatory system when normal blood pressure is perturbed. CST's knowledge base [Khuwaja et al., 1992] contains three layers of physiological parameters and their corresponding relationships. The tutoring protocol (CST and human) follows, in large measure, from the problem solving protocol that the students are required to follow. Students are asked to make predictions about the immediate response (the Direct Response phase, or DR) of seven CV parameters represented in the top layer. For example, if the perturbation is a hemorrhage (a sudden loss of blood) then the first correct prediction is that the central venous pressure (CVP) decreases. This causes the stroke volume (SV) to decrease, which is the second correct prediction. After the predictions are made, the tutor (human or CST) analyzes the patterns of errors and commences an interactive dialogue. The protocol of prediction followed by dialogue is repeated for two more phases of the human body's response to the perturbation: the Reflex Response (RR) and the new Steady State (SS).

The goal of CST is to enable the student to reason about the parameters represented in the top layer of the knowledge base. The parameters represented in the intermediate and deep layers of the knowledge base are utilized in the interactive dialogue to explain and reinforce the relationships represented in the top layer. Parameters represented in the top layer are also represented in deeper layers. For example, an increase in Cardiac Output (CO) causes an increase in Mean Arterial Pressure (MAP). This can be explained by two relationships represented in the intermediate layer:

1. An increase in CO causes an increase in Arterial Blood Volume (ABV; this parameter is not represented in the top layer) and,
2. An increase in ABV causes an increase in MAP

Although students are free to ask questions, or ask for help after predictions are made, they rarely do so in our human tutoring experiments. We primarily attribute this to our protocol which is very tutor-directed. Also, students often do not know when they have a difficulty or a misconception.

Hints and Other Tutorial Tactics

Version 2 of CST runs gracefully. However, the quality of hints is very poor. The hints are often inappropriate and always monotonous. In fact, the poor quality of the dialogue is typical of ITSs that employ natural language. The following is an example of a CST, version 2 hint:

T1: What are the determinants of SV?
S1: RAP.
T2: RAP is the correct answer. Consider the value of RAP. What is the other determinant of SV?

In this example, the student answered with a partially correct answer (the meanings of the acronyms are not important for this discussion). The question in T1 clearly indicates that there is more than one answer. However, the reply in T2 that "RAP is the correct answer" can easily confuse the student. This comment is followed by the directive "Consider the value of RAP." While recalling the value of a parameter is often useful in solving CST problems, it is of no help in this situation where a second determinant of SV needs to be recalled.

The most striking observation we have made from our tutoring transcripts is the prevalence and variety of hints provided by the tutor. We searched the ITS, educational and psychological literature for a study of hints and found little. It appears that ITS researchers are either not fully aware of the importance of hinting or they assume that hinting is a well understood tutorial tactic. Therefore, we initiated a more thorough study to understand how experienced human tutors provide hints. We found that some the best hints are very simple. The following is how one of our experienced tutors handled the identical situation as in the above CST, version 2 example:

T1: Now lets talk about all of the determinants of SV. You have one, RAP. Are there any others?

The subtle reference to "all of the determinants" alludes to that fact that there is more than one answer. Some hints we found were literally one word long ("and...," "so..."). There is implied and contextual meaning conveyed in such hints.

Our tutors have excellent communication skills and hinting appears to be a natural part of conversation. Therefore, we assume the effectiveness of our tutor's hints can be partly attributed to their conversational skills. However, many of hints we identified contained detailed domain knowledge. This is the knowledge included in our knowledge base's intermediate and deep layers. The following, from our transcripts, is an example of the type of hint we would like CST, version 3 to construct:

T1: If the alpha receptors are found ONLY on the blood vessels, then when the drug is administered, what variable will be changed first?

This hint contains knowledge from deeper layers of our knowledge base (about receptors and blood vessels). This knowledge is clearly presented to the student with the expectation that the additional information will enable the student to successfully answer a question (or make a correct prediction). Notice the intended emphasis in the capitalization of "ONLY." Proper consideration of the highlighted fact is required to arrive at a correct answer to the question.

After an analysis of our transcripts, we defined a hint as [Hume et al., 1993] (p. 564):

> A rhetorical device that is intended to either: (1) provide the student with a piece of information that the tutor hopes will stimulate the student's recall of the facts needed to answer a question, or (2) provide a piece of information that can facilitate the student's making an inference that is needed to arrive at an answer to a question or the prediction of system behavior.

Subsequently, we have identified two main categories of hints: convey information hints (CI-Hints) and point to information hints (PT-Hints). In a CI-Hint, the tutor provides one or more pieces of information and prompts the student (implicitly or explicitly) to answer the current question. PT-Hints suggest the availability of pertinent information but do not actually provide that information to the student. They require the student to engage in a sequence of mental activity. The student must:

1. not forget the original question (which may be implied and not explicitly stated),
2. understand the tutor's immediate question (again, it to may be implied),
3. recognize the connection between the tutor's immediate question and the original question, and finally
4. infer or recall the answer to the original question.

We have identified three other tactics used by our tutors: directed line of reasoning (DLR), explanation and summary. A DLR is a sequence of very focused questions by the tutor designed to illustrate a concept or a way of step-wise reasoning about a phenomenon. Explanations and summaries are didactic tactics. Such tactics do not actively involve the student in the dialogue.

We believe that these five tactics can be positioned along a continuum spanning passive to active learning. Explanations and summaries are tutorial tactics that only encourage that the student store information; they are on the passive end of this continuum. DLRs lie near the middle of the continuum; the student is being encouraged to actively consider the significance of the tutor's sequence of questions. Each of the steps places differing cognitive demands on the student. While a DLR often concludes with a summary, we presume that the DLR requires the student to engage in an active cognitive process that aids learning and retention.

CI-Hints require a greater degree of cognitive activity than DLRs, and they are thus positioned on the active end of the continuum. Each sequential question in a DLR is generally very direct and prompts the student for information presumed to be available from previous tutorial interaction. CI-Hints, on the other hand, provide partial information presumed not to be available, but then require the student to use this new information to move towards a correct answer.

PT-Hints provide the student with the opportunity to engage in the most active learning. They allude to information presumed to be available to the student. The student must first retrieve this information, often by answering an internally generated question. They then must understand why the information retrieved is relevant to the issue at hand. It is arriving at an understanding of the relevance that stimulates successful problem solving.

The Student Model

Interviews with our tutors reveal that they provide hints only when they believe that the student can make use of them. To implement such behavior in CST we have designed a student modeller that simulates the process our tutors use to evaluate the state of the student. One use of the student model is to determine *what* topic is to be tutored, but this will not be discussed in detail here. We believe the student modeller should also be used to select instructional strategies and tactics. In other words, the student modeller should be used to help determine *how* to tutor.

Our tutors, and we believe all human tutors, use very coarse grained schemes to evaluate students. Our tutors maintain an overall impression of the competence of the student in this problem solving environment; we call this the global assessment. While working on a particular topic or phase of the response, our tutors assess the student's performance on that part of the domain; we call this the local assessment.

In our tutoring setting, the tutor starts each session with a default global assessment of the student's state. Specifically, it is assumed that the student has acquired most of the necessary declarative knowledge by attending lectures and reading assigned material. We also assume that it is unlikely that the student has integrated the material into a mental model that can be used in problem solving. The global assessment is updated after predictions are analyzed and after topics are tutored. Local assessments are formed as individual topics are tutored. A tutor's local assessment of a student will rise when a student correctly answers a question. Also, the local assessment may also rise somewhat after a partially correct answer or after the student asks a relevant question. In other words, the local assessment will increase if the student understands the intent of the tutor even if a completely correct answer is not immediate. Repeated incorrect, or otherwise inappropriate, answers will lower the local assessment.

Determining When to Hint in CST

We have been unable to observe a regular pattern for the type of hint (PT or CI) initially generated in the discussion of a particular topic. We attribute this to the individual style of the tutor. Our study of student responses and follow up tutorial tactics [Hume et al., 1996b] has revealed some regular hinting patterns. From our observations, we have generated the following algorithm for determining when to hint:

1. When prediction errors are initially made, try hinting unless the global assessment is very low.
2. If the global assessment is sufficiently high, try a second hint if the first hint is not successful.
3. Continue to provide hints on a topic as long as:
    a) the global and local assessment are sufficiently high, and
    b) the number of hints while tutoring one topic is sufficiently low.
4. If a follow up hint is to be provided then:
    a) use a PT-Hint when the local assessment is high, and
    b) use a CI-Hint when the local assessment is low.

When the decision is made not to hint, the tutor often attempts to maintain active learning by generating a DLR. Explanations are, in a sense, a last resort.

Discussion

Why is there so little substantive discussion of hinting in the ITS literature? Not every ITS developer has attempted to explicitly simulate a human tutor working the domain, and it is possible that others have simply not considered hinting as a tutorial tactic. Even when information about human tutors is available, it is possible that developers have either not noticed the occurrence of hints, or at least have not directed much attention to this phenomenon. The occasional reference to "hints" (or "hinting"), usually with only isolated examples presented and no rules for when and how to generate them, would support this hypothesis. From our observations and a review of the human tutoring literature, we conclude that hinting is a more important topic than is evident from the ITS literature.

We are aware that our particular domain, student audience, and tutoring protocol all affect our tutor's choice of tactics. For example, our students rarely interrupt the tutor's agenda by asking an explicit question. However, from the existing literature on ITSs and hinting, we notice that the student, not the tutor, often dictates when hints are provided. For example, Sherlock II waits for the student to reach an impasse [Lesgold et al., 1992]. GIL [Reiser et al., 1992] provides hints to students when (1) the student reaches an impasse or (2) the student requests help.

ITSs have not produced natural and effective hints. We believe that there are many questions about hinting that must be addressed by the ITS community. We are interested in the following questions addressing the issues of when and how to hint, the relationship between hinting and the student model and how the tutoring protocol affects the hinting phenomenon.

1. Are the decisions about when and how hints that are made in tutoring settings in which the student controls the agenda different from the decisions that were made when the tutor controls the agenda?

2. What differences, if any, are there in the roles of the student model among these different tutoring situations?

3. What effect does the age and maturity of the student have on the tutor's pattern of hinting? Most studies on tutoring have looked at a student population that is younger than first year medical students.

4. Our protocol requires that the student make a sequence of predictions, without guidance, so that the tutor can determine what is to be tutored. However, human tutoring often commences with an interactive dialogue before any errors have been committed. What consequences are imposed by these two quite different protocols on the generation of hints?

5. What is the difference between the student knowing he/she has a problem and the tutor determining that the student has a problem? For example, a student working on a mathematics problem often knows when a mistake has been made. This is not the case in our domain and we assume this is one of the reasons why our students ask few questions. Are there different types of hints for when the student asks for help as opposed to the tutor determining that help must be provided?

6. What is the effect of non-verbal communication present in the usual face-to-face tutoring situation? We assume it is natural for a tutor to gauge the progress of a student. The correctness of answers to domain questions is certainly the most obvious input to the tutor's evaluation of the student. However, there are certainly more subtle cues that human tutors utilize (intonation and pauses are obvious examples). How does the lack of such cues in our tutoring (human or by computer) affect student modelling and hinting?

7. Is a human tutor's evaluation of the student optimal? A human has a limited memory, while a computer's memory is practically limitless. Thus, a computer tutor can, in principle, save and use vastly more information about the student's responses than a human tutor. Can a denser, more fine-grained student model in an ITS be an asset? Or, is a coarse grained student model sufficient, perhaps even preferable?

8. What is the relationship between the ambiguous surface forms of hints and the intention of the tutor? For example, our tutors regularly pose questions that are intended to be interpreted as statements. Likewise, they regularly make statements that are intended to be questions.

9. Are hints effective (whether generated by human or computer tutors)? That is, do they actually enhance student learning? We believe so but we have only anecdotal evidence.

10. Is student modelling effective and can an ITS's student model be used to help determine *how* to tutor?

There is some literature on human tutoring that alludes to the relationship between hinting and student modelling. Fox comments that "the tutor and student both make use of strategies which maximize the student's opportunities to correct his or her own mistakes" [Fox, 1993] (p. 122). She also suggests that student behavior that may be useful in diagnosis. Reiser [Reiser, 1989] observed that human tutors "moderate their control of the interaction to provide sufficient assistance for the student to solve the problem." He also observed that hints, and the feedback from hints, are the tools human tutors use for this monitoring. We find these references significant as they support our claim that student modelling is essential for hinting. More research is needed to understand how to make ITS hint effectively and how hinting should be altered to accommodate different domains and student populations.

Conclusion

CST's (version 3) knowledge base and student modeller are, essentially, fully implemented. We are enhancing CST's natural language capabilities. When CST is operational, we plan to conduct experiments that address our questions about student modelling and hinting. We acknowledge, however, that many of our findings (current and future) only pertain the context of our environment. Specifically:

1. Our students (first year medical students) are older and more mature than students in most tutoring experiments.

2. Our human tutors (medical school professors) are domain experts and experienced tutors. Many human tutors are students with slightly more domain knowledge than their tutees. Many human tutors also have little tutoring experience.

3. Our tutoring goal (human and ITS) is assist in the acquisition of procedural knowledge (the ability to solve problems). While declarative knowledge is necessary, some tutoring environments focus on declarative knowledge more than procedural knowledge.

4. Our domain requires that the student use qualitative reasoning. There is no quantitative manipulation.

5. Our protocol (human and ITS) dictates that students complete a problem solving activity where the tutor can observe errors. The student is not allowed to ask questions until after a problem is attempted.

6. Our protocol (human and ITS) relies on a natural language, via keyboards and scrolling text, interface. Many human tutoring experiments are conducted in a face-to-face manner. Many ITSs use menus. Soon, hypertext and other multimedia interfaces will be employed.

7. Our students often do not know when they have a misconception.

We have done a careful search of the ITS literature and have concluded that many of our colleague's findings are also specific to their particular domains, protocols and student populations. ITS researchers need a forum to coordinate their efforts in order to generalize their findings.

References

[Fox, 1993] Fox, B. (1993). *The Human Tutorial Dialogue Project: Issues in the Design of Instructional Systems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

[Hume et al., 1993] Hume, G., Michael, J., Rovick, A., & Evens, M. (1993). The use of hints as a tutorial tactic. *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (pp. 563-568). Boulder, CO.

[Hume et al., 1995] Hume, G., Michael, J., Rovick, A., & Evens, M. (1995). Controlling active learning: how tutors decide when to generate hints. *Proceedings of the 8th Florida Artificial Intelligence Research Symposium* (pp. 157-161). Melbourne Beach, FL.

[Hume et al., 1996a] Hume, G., Michael, J., Rovick, A., & Evens, M. (1996). Hinting as a tactic in one-on-one tutoring. *The Journal of Learning Sciences*, 5 (1), 23-47.

[Hume et al., 1996b] Hume, G., Michael, J., Rovick, A., & Evens, M. (1995). Student Responses and Follow Up Tutorial Tactics in an ITS. To appear in the *Proceedings of the 9th Florida Artificial Intelligence Research Symposium* Key West, FL.

[Khuwaja et al., 1992] Khuwaja, R., Evens, M., Rovick, A., & Michael, J. (1992). Knowledge representation for an intelligent tutoring system based on a multilevel causal model. In C. Frasson, G. Gauthier, & G. I. McCalla (Eds.), *Intelligent Tutoring Systems: Proceedings of the Second International Conference, ITS '92* (pp. 217-224). Montreal, Canada.

[Kim et al., 1989] Kim, N., Evens, M., Michael, J., & Rovick, A. (1989). CIRCSIM-TUTOR: An intelligent tutoring system for circulatory physiology. In H. Maurer (Ed), *Computer Assisted Learning, Proceedings of the International Conference on Computer-Assisted Learning* (pp. 254-266). Dallas, TX. Berlin: Springer-Verlag.

[Lesgold et al., 1992] Lesgold, A., Eggan, G., Katz, S., & Rao, G. (1992). Possibilities for assessment using computer-based apprenticeship environments. W. Regian & V. Shute (Eds.), *Cognitive approaches to automated instruction* (pp. 49-80). Hillsdale, NJ: Lawrence Erlbaum Associates.

[Li et al., 1992] Li, J., Seu, J., Evens, M., Michael, J., & Rovick, A. (1992). Computer dialogue system (CDS): a system for capturing computer-mediated dialogue. *Behavior Research Methods, Instruments*, & Computers, 24, 535-540.

[Reiser, 1989] Reiser, B. (1989). Pedagogical strategies for human and computer tutoring. *Proceedings of the American Educational Research Association*. San Francisco, CA.

[Reiser et al., 1992] Reiser, B., Kimberg, D., Lovett, M., & Ranney, M. (1992). Knowledge representation and explanation in GIL, an intelligent tutor for programming. In J. Larkin, & R. Chaby (Eds.), *Computer Assisted Instruction and Intelligent Tutoring Systems: Shared Goals and Complementary Approaches*. (pp.111-149). Hillsdale, NJ: Lawrence Erlbaum Associates.

Acknowledgments

This work was supported by the Cognitive Science program, Office of Naval Research under Grant No. N00014-941-0338 to the Illinois Institute of Technology. The content does not reflect the position or policy of the government and no official endorsement should be inferred.