A Collaborative Intelligent Tutoring System for Medical Problem-Based Learning

Siriwan Suebnukarn, Peter Haddawy
Computer Science and Information Management Program
Asian Institute of Technology
PO Box 4, Pathumthani, Thailand
+66 2524 5700

ssiriwan@cs.ait.ac.th, haddawy@ait.ac.th

ABSTRACT

This paper describes COMET, a collaborative intelligent tutoring system for medical problem-based learning. The system uses Bayesian networks to model individual student knowledge and activity, as well as that of the group. It incorporates a multi-modal interface that integrates text and graphics so as to provide a rich communication channel between the students and the system, as well as among students in the group. Students can sketch directly on medical images, search for medical concepts, and sketch hypotheses on a shared workspace. The prototype system incorporates substantial domain knowledge in the area of head injury diagnosis. A major challenge in building COMET has been to develop algorithms for generating tutoring hints. Tutoring in PBL is particularly challenging since the tutor should provide as little guidance as possible while at the same time not allowing the students to get lost. From studies of PBL sessions at a local medical school, we have identified and implemented eight commonly used hinting strategies. We compared the tutoring hints generated by COMET with those of experienced human tutors. Our results show that COMET's hints agree with the hints of the majority of the human tutors with a high degree of statistical agreement (McNemar test, p = 0.652, Kappa = 0.773).

Categories and Subject Descriptors

I.2.1 [**Artificial Intelligence**]: Applications and Expert Systems; J.3 [**Life and Medical Sciences**]: Medical Information Systems; K.3.1 [**Computers and Education**]: Computer Uses in Education – *Collaborative Learning*.

General Terms

Human Factors, Design.

Keywords

Intelligent Tutoring Systems, Problem-based Learning, Bayesian Networks, Medicine, Collaborative Environments, Multi-modal Interfaces

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
IUI'04, January 13-16, 2004, Madeira, Funchal, Portugal.
Copyright 2004 ACM 1-58113-815-6/04/0001…$5.00.

1. INTRODUCTION

Today a great many medical schools have turned to a problem-based learning (PBL) approach to teaching. PBL instructional models vary but the general approach is student-centered, small group, collaborative problem-based learning activities [1]. A major goal of the PBL approach in medicine is to produce practitioners who can function collaboratively in real-world problem-solving situations during their medical careers. While PBL has many strengths, effective PBL requires the tutor to provide a high degree of personal attention to the students. In the current academic environment where resources are becoming increasingly scarce and costs must be reduced, providing such attention becomes increasingly difficult. This is exacerbated by the fact that medical school faculty, in particular, often have limited time to devote to teaching. As a consequence, medical students often do not get as much facilitated PBL training as they might need or want. Fortunately, the field of Intelligent Tutoring Systems (ITS) has begun to address this problem by developing systems that can provide one-to-one tutoring with essentially no incremental cost per additional student. While work in this area has traditionally focused on tutoring of individual students, medical PBL is typically carried out in a group setting. Work in the area of computer supported collaborative learning (CSCL) has addressed the problem of creating shared learning environments where students from diverse locations can participate in group learning.

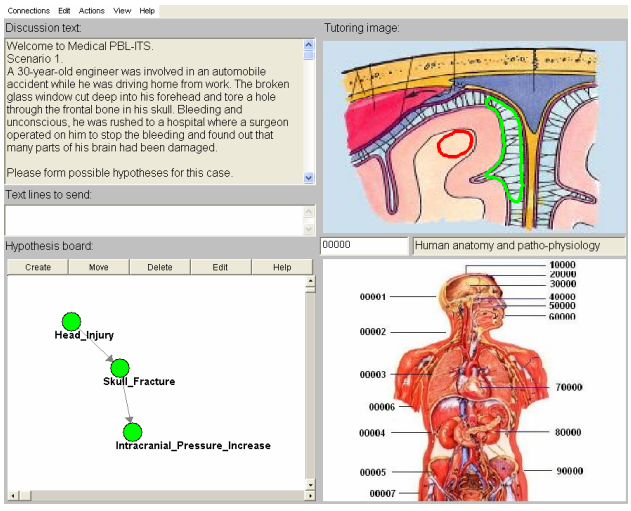

Figure 1. COMET student interface.

In the current work, we combine concepts from ITS with those from CSCL to develop an intelligent group-based medical PBL system. One of the major challenges here is to automate the tutoring process in the context of group activity.

2. RELATED WORK

Researchers in the area of computer supported collaborative learning have developed numerous computer supported PBL environments over the past decade [2] [3]. This work has focused mainly on providing and maintaining shared information resources and shared workspaces rather than on intelligent tutoring. Work on intelligent tutoring systems for medical PBL has generally addressed tutoring of individual students, e.g., in the area of cardiovascular physiology [4]. More recent work has attempted to integrate computer-supported collaborative environments with some form of user modeling. The Docs 'n Drugs project [5] supports intelligent tutoring for group-based medical PBL by including collaborative work and intelligent tutoring capabilities in one system. But the tutoring module in Docs 'n Drugs is still focused on guiding individual students rather than the group as a whole. Jameson et al [6] propose a generative model of individual group members, which is a computational model of relevant beliefs, preferences, motivation and other relevant properties. The work focuses on supporting asynchronous collaboration, with the models being used to predict member's responses to proposed solutions during discussion sessions when they are not present. Lock and Kudenko [7] propose a multi-component user modeling approach in which each user model contains an explicit team profile in addition to other distinct components. The models are developed in the context of personalized information briefing for military decision-making. The models represent each user's information interests, as well as an aggregate interest for the group as a whole. Building upon results from Social Choice Theory, Masthoff [8] addresses the issue of combining models of individuals' preferences in order to infer group preferences in a more general framework. The work is illustrated with the problem of selecting appropriate television programming for a group and does not address problem solving.

Our work departs from previous efforts to incorporate user modeling into computer supported collaborative environments by focusing on modeling individual and group problem solving behavior. Since problem solving in group PBL is a collaborative process, modeling individuals and the group is necessary if we wish to develop tutoring algorithms that can do things like focus the group discussion, promote collaboration, and suggest peer helpers. In this paper we present COMET, a collaborative intelligent tutoring system for medical problem-based learning. The system uses Bayesian networks to model individual student knowledge and activity, as well as that of the group. It uses the models to generate tutorial hints to guide group problem solving activity. It incorporates a multi-modal interface that integrates text and graphics so as to provide a rich communication channel between the students and the system, as well as among students in the group. Students can sketch directly on medical images (a common practice in medical group consultation), search for medical concepts, and sketch hypotheses on a shared workspace. In addition, the system can choose to display medical images to the students as a form of guidance in problem solving.

3. MEDICAL PROBLEM-BASED LEARNING

In PBL it is the problem that drives the learning. In medical PBL settings students are presented with "real-life" cases that require defining the problem, creating hypotheses, gathering and analyzing data, and evaluating or justifying solutions collaboratively [1]. The patient's problem is designed to challenge students to develop reasoning, problem solving, and team skills. Students are actively involved in the creation of solutions for the problem. They integrate and organize learned information as they encounter the problem. The problem becomes a stimulus for learning and for the recall and application of their knowledge. Consider, for example, the scenario in the upper left of Figure 1, taken from a PBL session in the brain course at Thammasat University Medical School. Here students must come up with possible hypotheses to explain why the patient has become unconscious.

One of the main issues in PBL is the role of the tutor. A number of studies [9] [10] have examined the effectiveness of student learning with different types of tutors. Like a good coach, a tutor needs enough command of whatever the learners are working on to recognize when and where they most need help. Content experts, however, who do not understand the importance of being a guide and facilitator rather than a purveyor of information can be tempted to give answers rather than help learners find their own answers. In such cases, the content expert may be prone to reverting to a lecture mode, destroying the PBL process. So the ideal tutor should be an expert in both learning content and process, which is rare to find among human tutors. The tutor intervenes to as small an extent as possible, posing open-ended questions and giving hints only when the group appears to be getting stuck or off track. In this way, the tutor avoids making the students dependent on him for their learning [9].

4. SYSTEM OVERVIEW

COMET is designed to provide an experience that emulates that of live human-tutored medical PBL sessions as much as possible while at the same time permitting the students to participate from disparate locations. The system is implemented as a Java client/server combination, which can be used over the Internet or local area networks and supports any number of users. Our prototype incorporates substantial domain knowledge about head injury diagnosis. The system contains four primary components: student multi-modal interface, medical concept repository, student clinical reasoning model, and tutoring module. We describe each component in turn.

4.1 Multi-modal Interface

The multi-modal interface (Fig. 1) supports communication between students and the system and among students in the group. It contains three components: hypothesis board, chat pane, and image pane.

4.1.1 Hypothesis Board

The hypothesis board (Fig. 1, lower left pane) provides the central shared group workspace. It records the group's collective thinking and serves as a stimulus for additional ideas by group members. Students can create hypothesis nodes and create causal

links between nodes. Hypothesis node labels are created by either retrieving them from the medical concept repository or by sketching on the medical image (see discussion of these below). Nodes and links can also be deleted. Any student may make changes to the board contents and all students have the same view of the board. The system has access to all information entered on the board.

4.1.2 Chat Pane

Students may communicate with others in the group by typing into the chat pane (Fig. 1, middle left pane). COMET has no language processing capabilities, so the text in the chat pane is not taken as input to the system.

4.1.3 Medical Image Pane

In medical consultation it is common for physicians to discuss medical images and to communicate by pointing to or drawing on the images. Since this is such a common and useful means of communication among physicians, it is important for the system to support this mode of communication. The interface includes an image pane (Fig. 1, upper right pane) in which COMET displays images that are relevant to the current focus of the group discussion. All students see the same image and see anything that other students sketch or point to on the image. If a student circles or otherwise annotates a region of the image representing a valid hypothesis, it is taken as input to the system and the corresponding hypothesis is added to the hypothesis board.

4.1.4 Hierarchical Medical Concept Repository

The system includes a hierarchical medical concept repository (Fig. 1, lower right pane) to help students better understand the relationships between domain concepts, as well as to facilitate system input. Concepts from human anatomy and patho-physiology are indexed by a five-digit number that is mapped to a specific anatomic part of the human body. The numeric scale indicates relationships along four dimensions: general to specific, structural dependence, circulation paths, and chronology. Students can search through a specific system and search deeply to get the more specific body part. For example, 20000 represents the Musculoskeletal system, 21000 represents Bones of the Cranium and face, and 21110 represents the Frontal bone. The physiological, biochemical or behavioral functions or dysfunctions relevant to an anatomic part are displayed for each search level.

All valid hypotheses are stored in the repository, relieving the system of the need to process typed text input. This is a great simplification since there may be many ways to refer to the same medical concept. Since the repository contains a very large number of concepts, storing the hypotheses in the repository does overly simplify the students' task. In fact, by requiring the students to search through the repository to find the appropriate hypothesis, the system helps the student to better understand the relationship between the hypothesis and the larger anatomic and patho-physiological context.

5. STUDENT CLINICAL REASONING MODEL

Generating appropriate tutorial actions requires a model of the students' understanding of the problem domain. This modeling task is necessarily wrought with uncertainty since we have only a limited number of observations from which to infer each student's level of understanding. Thus we have chosen to use Bayesian networks as our modeling technique. We model each student with an instance of our general Bayesian network student model. The group is reasoned about by combining information from the models of the individual students, as described below. Figure 2 shows a portion of the student model corresponding to the head injury scenario in Figure 1. The model contains two types of information; (1) the hypothesis structure based on the differential diagnosis of the scenario (the right group of nodes in Fig. 2); and (2) the application of medical concepts in terms of anatomy and patho-physiology (the left group of nodes in Fig. 2) to derive the hypotheses.

We represent the hypothesis structure following the model of Feltovich and Barrows [11], which defines three categories of illness features: enabling conditions, faults, and consequences. Enabling conditions are illness features associated with the acquisition of illness. They include more specific categories such as predisposing factors (e.g., compromised host factors, unusual travel, hereditary factors), which can make an individual more susceptible than usual to illness in general or to particular illnesses. Faults are the major real malfunctions in illness and are characterized abstractly as major different types, e.g., direct trauma, invasion of tissue by pathogenic organisms, inadequate blood supply, or inability of tissue to survive/ thrive. Consequences are the secondary consequences of faults within the organism, and generally comprise different types of signs and symptoms, e.g., unconsciousness, brain damage, or intracerebral hemorrhage. In Figure 2 (right half), we have seven possible faults associated with the single enabling condition car accident: *Head_Injury*, *Brain_Moving*, *Scalp_Lacerate*, *Skull_Fracture*, *Brain_Contusion*, *Diffuse_Axon_Injury*, and *Brain_Stem_Damage*. The remaining hypothesis nodes are consequences of these faults. Each hypothesis node has parent nodes, which have a direct casual impact on it. For example, *Brain_Damage* has parents *Brain _Infection* and *Intracerebral_Hemorrhage*. All hypothesis nodes have two states, indicating whether or not the student knows that the hypothesis is a valid hypothesis for the case.

The application of medical concepts is represented in terms of three kinds of nodes: goals, general medical knowledge, and apply actions. Every hypothesis node (except the root, which represents the scenario itself) has a unique *Apply* node as parent. The *Apply* node represents the application of a medical concept to a goal in order to derive the hypothesis. For example the *Apply3* node indicates that the student is able to use knowledge of the *Blood_Flow_Decrease* medical concept to infer that *Brain_ Damage* is a consequence of *Brain_Infection*. Each hypothesis node thus has a conditional probability table specifying the probability of the hypothesis being known conditioned on whether the parent hypotheses are known and whether the student is able to apply the appropriate piece of knowledge to determine the cause-effect relationship. The conditional probability tables for the *Apply* nodes are simple AND gates. Our notion of applying a concept to a goal to generate a hypothesis is similar to the student models used by Conati, et al [12]. Our model differs from theirs in the way we represent the internal causal relationships of the hypothesis structure (Fig. 2, right half). The hypothesis nodes in our model correspond to the fact nodes in

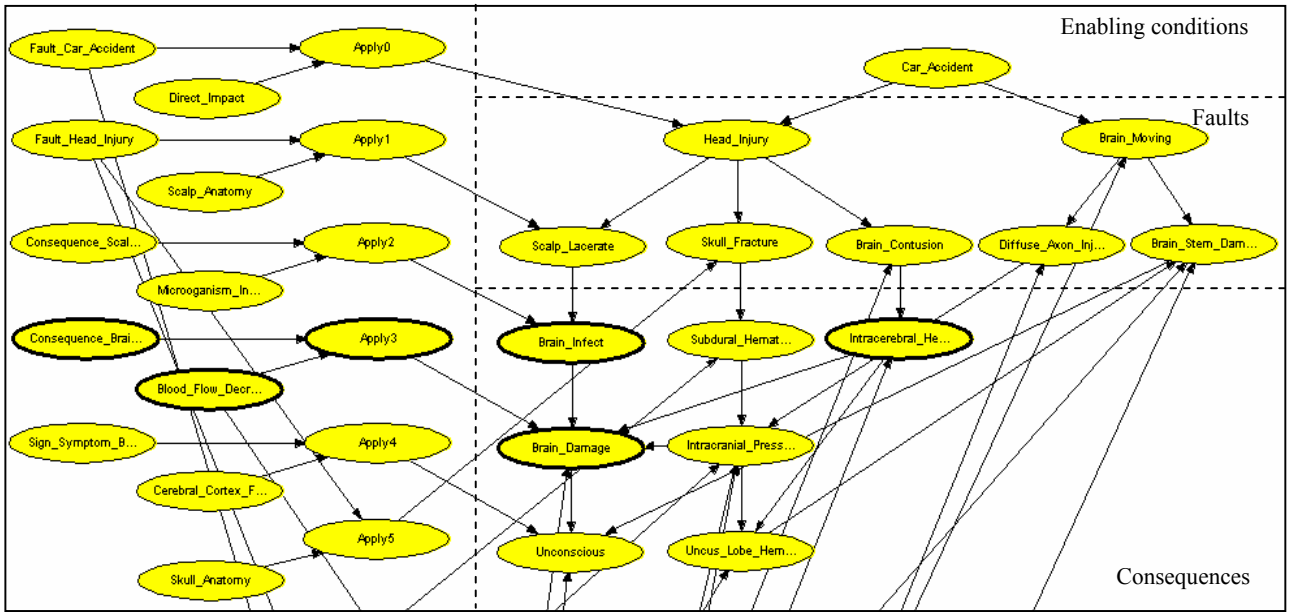

Figure 2. Part of the Bayesian network student model. The complete network contains 66 nodes.

their model. In their model, each fact node has a goal node as parent, while in our model each hypothesis node has an apply node and another hypothesis as parents. In their model, a fact node can be a parent of a goal node, whereas in our model goals are root nodes.

To construct the model which took about one person-month, we started from an initial Bayesian network built based on information extracted from medical textbooks and from interviews with a neurosurgeon and a physiologist from Thammasat University Medical School. The initial model was refined by learning structure and parameters using data collected from a medical problem-based learning tutorial. The data for this study consisted of tape recordings of tutorial sessions that occurred in the brain course at Thammasat University Medical School. A total of 15 groups of third year medical students were involved in this study. Each group consisted of eight students with different backgrounds. They were presented with the head injury case and asked to construct possible hypotheses for the case, under the guidance of a tutor. After the sessions the tape was analyzed to determine whether or not each hypothesis was mentioned. Since the data we were able to gather was rather sparse (15 data set), we used the constraint-based PC learning algorithm [13], which allowed us to incorporate the prior knowledge of structure and of parameters.

The student model is used to reason about the state of knowledge of each student, as well as about problem solving behavior of each student and of the group. The general student model is instantiated for each student by entering that student's medical background knowledge as evidence. For example, if a student has a background in anatomy, we would instantiate the *Skull_Anatomy* and *Scalp_Anatomy* nodes. We make the assumption that once a hypothesis in the domain model is mentioned by one student in the group, every student knows that hypothesis. So as hypotheses are mentioned, they are instantiated in each student model. Following commonly accepted practice in medical PBL [11], we assume that students should and generally do enumerate the possible hypotheses by focusing sequentially on the various causal paths in the domain, linking enabling conditions with faults and consequences. So for each student, we must determine what causal path he is reasoning along, which we do by identifying the path of highest probability. This is computed as the joint probability of the nodes along the path. It is also necessary to identify a causal path that can be used to focus group discussion, particularly when the discussion seems to be diverging in different directions. This is done as follows. We take the most likely path for each student and compute the sum of the probabilities of that path over all the students. We then take the path with the highest sum. The idea is that we want to identify the path that has much of the attention of much of the group and has at least one member whose attention is focused on that path. Note that if we only identify the path with highest sum of probabilities over all members of the group, it is possible that no single member of the group is actually focused on that path.

6. TUTORING MODULE

Our automated tutor takes on the role of guiding the tutorial group to construct possible hypotheses for the case by the use of specific open-ended questions. From our study of the tutoring session transcripts, we identified eight hint strategies commonly used by experienced human tutors: 1) focus group discussion using general hint, 2) focus group discussion using specific hint, 3) promote open discussion, 4) deflect uneducated guessing, 5) avoid jumping critical steps, 6) address incomplete information, 7) refer to experts in the group, and 8) promote collaborative discussion. We developed algorithms to generate each of these types of hints, using as input the interaction log and the Bayesian network student models. Each hint strategy is discussed in turn below.

Strategy 1, 2: Focus Group Discussion

At the beginning of a session, members of the group may suggest various different valid hypotheses, without focusing on any given causal path. When such a lack of focus becomes apparent, the tutor should intervene by directing the students to focus on one of the hypotheses. The hypotheses that the students are asked to develop should be those of underlying basic mechanisms. Generally, the tutor guides with open-ended general hints related to the ideas they have suggested. Then more specific hints are aimed at identifying the correct mechanism responsible for the patient's problem.

Students: car accident, skull fracture, brain damage
Tutor: Should we discuss each of them in more detail? What is the consequence of skull fracture?

Students: Silent

Tutor: What happens to the structure underlying the skull?

In our study, the tutor typically gave the first hint after the students had mentioned about three or four hypotheses without focusing on a particular causal path. This number represents about twenty percent of the hypothesis nodes in the domain. Thus we use this percentage as the point at which the tutor should intervene if the students have not yet focused their discussion. The system selects the hypothesis node which has not been mentioned and has highest probability (the most likely hypothesis that the students can think about) from the group reasoning path as a hint node. Its parent nodes will be used to generate the tutoring dialogue. The system gives the first general hint using the parent goal node, and if there is no student response observed, the more specific parent medical concept node will be used. The medical image which is relevant to the parent nodes will be displayed along with the tutoring dialogue.

Strategy 3: Create Open Environment for Discussion

While in the process of guiding the group to elaborate a particular hypothesis and its consequences, the tutor may be presented with a new proposed idea, which is not directly related to the current path. One option is for the tutor to ask the students to delay discussion of this concept, but such feedback is usually considered too negative and not conducive to encouraging student contributions. A better approach is for the tutor to incorporate this concept in the current discussion but retain focus by encouraging the students to relate the new concept to the current hypothesis under discussion.

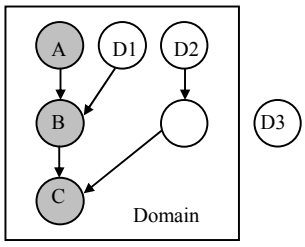

Figure 3. Three types of divergence

Students: skull fracture → subdural hematoma → intracranial pressure increase

Student: brain contusion

Tutor: "Can you relate its mechanism to what we have discussed?"

For any new hypothesis mentioned by the group, the system checks whether it is in the causal path on which the group is currently focused. If it is not in the path, the system will determine the degree of divergence and respond to it differently depending on its relation to the current path. We distinguish three different degrees of divergence: mild, moderate, and severe. As shown in Figure 3, a hypothesis (node D1) is mildly divergent if either its parent or child node is in the current path (A → B → C). In this case, the system first asks the group to find the relationship between node D1 and the current path. If the students cannot suggest a relationship or suggest the wrong one, COMET generates a more specific hint, focusing on node B (the hint node). A hypothesis is moderately divergent (node D2) if it is linked through one or more intermediate nodes to the current path. The system first comments that the new hypothesis is somewhat divergent from the path the group is focusing on but that it is useful if they can find its relationship to the path. In this case Node C serves as the hint node. A hypothesis is severely divergent (node D3) if the domain model does not contain a node corresponding to that hypothesis. The system responds by asking the other group members to comment on whether this hypothesis is relevant, which is a subtle way of suggesting that it may not be.

Strategy 4: Deflect Uneducated Guessing

From time to time, a student will contribute erroneous information or an erroneous interpretation of data. The tutor should attempt to turn uneducated guessing into a search for mechanisms.

Students: intracranial pressure increase → brain contusion
Tutor: "I feel there is something not quite right about the matter."

In this example the student proposed an incorrect causal link between hypotheses. Whenever a student draws an arc between two hypothesis nodes, the system checks to see if the domain model contains a directed path from the node at the tail to the node at the head. If not, then the student has drawn an incorrect causal link and the tutor intervenes. In this case, the hint node is the child of the node at the tail.

Strategy 5: Avoid Jumping Critical Steps

The tendency to jump over intermediate hypotheses in a causal path is quite common and must be resisted. If this is allowed to happen the critical thinking stage of the group process is lost. If the students jump directly from the patient enabling conditions to the effects or signs and symptoms without explaining their consequences, they miss exploring the problem situation in depth. The tutor might point to what they have mentioned by asking generally for its consequence or giving them a hint to the related medical concepts.

Students: skull fracture → unconsciousness

Tutor: "Can you think of the mechanism underlying why skull fracture causes unconsciousness?"

Figure 4 illustrates a situation in which a student creates a direct arc from hypothesis A to hypothesis E, missing the intermediate

hypotheses B, C, and D. The system determines whether there are intermediate nodes between the head and tail of the new arc. If so, the system gives a hint by pointing the group to the highest probability intermediate node.

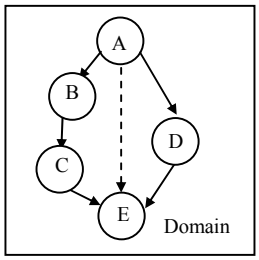

Figure 4. Jumping critical steps

Strategy 6: Address Incomplete Information

One objective of PBL is to have the students enumerate all possible causal paths between likely hypotheses in the domain. If the students miss some hypothesis, the tutor must indicate to the students that some possible hypotheses have not yet been explored.

Students: skull fracture → subdural hematoma → intracranial pressure increase → brain damage → unconsciousness

Tutor: "Would there be any alternative explanation of unconsciousness?"

This happens when all hypothesis nodes in the current reasoning path have been mentioned. The system will now identify a new reasoning path and new hint node.

Strategy 7: Refer to Expert in the Group

PBL is typically carried out in a group setting in order to teach students from different backgrounds how to collaborate to solve a given problem. Referring to an expert in the group might help to continue the group learning process when the group gets stuck even when the tutor tries to provide more specific hints.

Students: Acceleration –deceleration force → brain stem damage → unconsciousness

Tutor: "Can you think of the mechanism underlying why brain stem damage causes unconsciousness?"

Students: Silent

Tutor: "What is the function of the brain stem?"

Students: Silent

Tutor: "It seems like we have a problem. Val, you are good at physiology; could you help the group?"

One possible way to continue the learning process is by asking someone who is likely to know the answer in the group. The system examines each individual student model, trying to find the student who has the highest probability of knowing the hint node. The differences in background knowledge, represented in the individual student models, result in differences in the estimated likelihoods that the various students know the hypotheses.

Strategy 8: Promote Collaborative Discussion

Collaboration is an important medical career skill. In group PBL discussions it is not unusual for one member to take a dominant role in the group, always pushing his ideas and frequently interrupting other students. Even if the ideas this leader comes up with are good, the tutor must make a point of asking other members for their ideas in order to try to promote participation by the less dominant members.

We implement this strategy by using the log file of students' activity. Any actions consisting of drawing on the hypothesis board or annotating the tutoring image are saved as the activity log file in the server. Suppose there are 6 students in the group. If there is no input from a student after 6 hypotheses have been mentioned, the system should point to that student.

7. SAMPLE INTERACTION

We show an example of running the system with a group of five students. One of them (Val) is an anatomist. Figure 5 shows a screen shot of the interface for one student. The following transcript shows the interaction with the system after the students read the problem scenario.

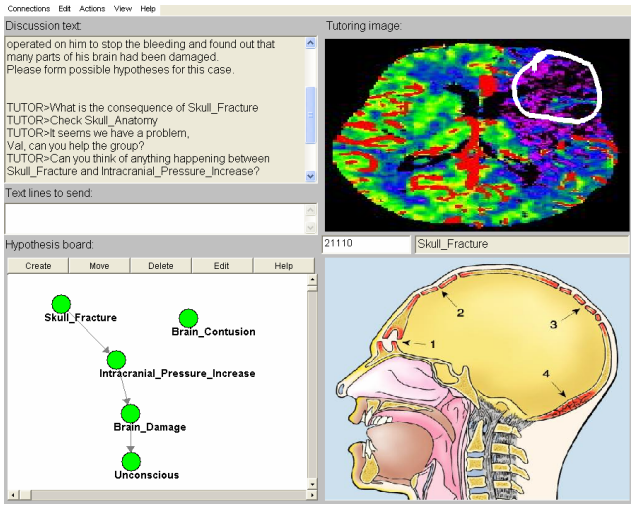

Figure 5. Screen shot of system at end of sample interaction transcript.

Students: Intracranial_Pressure_Increase, Skull_Fracture, and Brain_ Damage.

Tutor: (Stratagy 1) What is the consequence of Skull_Fracture?

Student: Unconsciousness

Tutor: (Strategy 2) Check Skull_Anatomy

Student: Silent

Tutor: (Strategy 7) It seems we have a problem, Val, can you help the group?

Val: Brain_Contusion

Students: Skull_Fracture → Intracranial_Pressure_ Increase → Brain_Damage → Unconsciousness

Tutor: (Strategy 5) Can you think of anything happening Between Skull_Fracture and Intracranial_Pressure_ Increase?

8. EVALUATION

In order to evaluate the appropriateness and quality of the hints generated by our system, we compared its responses to those of experienced human tutors. For each of the eight hinting strategies discussed in the previous section, we created three scenarios under which COMET would generate that hint, giving a total of 24 test scenarios. Ten tutors with at least five years experience in conducting the brain course at Thammasat University Medical School were asked to analyze each situation and indicate how they would manage the PBL group: whether they would provide a hint and what hint they would provide. This gave us a total of 240 data points to compare to the system's automated responses. Figure 6 shows the percentage of human tutors (y-axis) using various strategies for each strategy used by COMET (x-axis). For example, the first bar indicates that in the situations in which COMET used strategy 1, 60% of the human tutors also used strategy 1, while 7% used strategy 2, 20% used strategy 4, and the remaining 13% gave no hint. The degree of agreement between COMET and the human tutors differed for the various hint strategies. The greatest agreement occurred for strategies 3, 4, 5, and 8. For strategy 7, when the group got stuck even after receiving a very specific hint from the tutor, about one third of the human tutors tended not to refer to the expert student in the group, but rather to deflect uneducated guessing or avoid jumping a critical step. On average, 74.17% of the human tutors used the same hint strategy as COMET. The most similar human tutor agreed with COMET 83% of the time and the least similar tutor agreed 62% of the time.

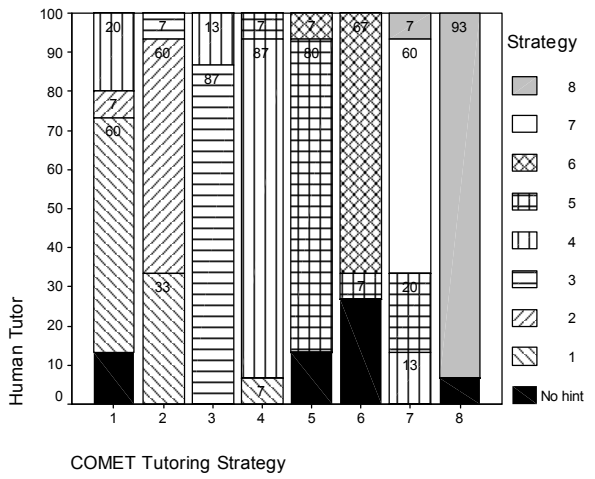

Figure 6. Results comparing system and human tutor responses. The x-axis indicates the strategy used by COMET and each bar indicates the percentage of the human tutors using the various strategies for that scenario.

To test the statistical significance of the agreement between the system and the human tutors, we used the McNemar test and Kappa statistic [14], which are commonly used in medicine to determine the degree of agreement between two alternative testing procedures. There were no statistical differences between the human tutors and COMET (McNemar test, $p = 0.652$). The results showed high degree of agreement between the hints generated by COMET and by the human tutors (Kappa index = 0.773). Interestingly, for those human tutor responses that differed from the system's there was generally little agreement among the responses. The exception is perhaps for COMET strategy 2 but here the dissenting human response was to choose strategy 1, which is very similar to 2. Our results indicate that the responses generated by COMET are in line with the majority consensus responses of the human tutors tested.

9. CONCLUSIONS AND FUTURE WORK

In this paper we have described an initial examination of tutoring strategies in PBL tutorial group settings. Examples from classroom discourse were analyzed and a number of specific types of tutorial actions identified. We have described a Bayesian network clinical reasoning model that integrates the hypothesis structure based on the differential diagnosis of a head injury case and the application of the corresponding medical concepts in the problem solving process. The proposed representation can be used as a guideline for developing new scenarios. We then implemented human tutoring strategies in our collaborative intelligent tutoring system. Results from our evaluation show a high degree of agreement of tutoring dialogues between our system and human tutors.

We plan to more extensively test the current system with medical students to improve the system's robustness. We also plan to incorporate all the essential domains in the pre-clinical medical PBL course. The current version of COMET supports single session group PBL. But PBL typically occurs over a period of several days, with students carrying out individual learning tasks and bringing their learned knowledge back to the group. We intend to add support for this aspect of PBL. Finally, the ultimate test of the effectiveness of our work is how it impacts student learning. So we intend to compare the effectiveness of student learning with COMET versus student learning with human tutors.

10. ACKNOWLEDGMENTS

We thank Hugin Expert for providing us the use of the Hugin Reseacher software. Thanks to Fabio Cozman for the use of the JavaBayes interface libraries in building the hypothesis board. Thanks to Thammasat University Medcial School for their participation in the data collection and system evaluation, and to Dr. Kesorn Jongjarern and Dr. Pornchai Yodvisitsak for their helpful suggestions in designing the student clinical reasoning model.

11. REFERENCES

[1] Barrows, H. S. A taxonomy of problem-based learning methods. *Medical Education*, 20, 481-486, 1986.

[2] Mahling, D. E., Sorrows, B. B., and Skogseid, I. A Collaborative Environment for Semi-Structured Medical Problem Based Learning, Proceedings of Computer-Supported Collaborative Learning, 1995.

[3] Suthers, D., Toth, E., and Weiner, A. An Integrated Approach to Implementing Collaborative Inquiry in the Classroom. Proceedings of Computer-Supported Collaborative Learning, 272-279, 1997.

[4] Martha, W., et al. CIRCSIM-Tutor: An Intelligent Tutoring System Using Natural Language Dialogue. Proceedings of Midwest AI and Cognitive Science Conference, 16-23, 2001.

[5] Martens, A., Bernauer, J., Illmann, T., and Seitz, A. Docs 'n Drugs - The Virtual Polyclinic. An Intelligent Tutoring System for Web-Based and Case-Oriented Training in Medicine. Proceedings of the AMIA Fall Symposium, 2001.

[6] Jameson, A., Baldes, S., and Kleinbauer, T. Generative Models of Group Members as Support for Group Collaboration. Workshop on User and Group Models for Web-Based Adaptive Collaborative Environments Proceedings of the International Conference on User Modeling, 1-14, 2003.

[7] Lock, Z., and Kudenko, D. Multi-component User Models of Team Members. Workshop on User and Group Models for Web-Based Adaptive Collaborative Environments, in International Conference on User Modeling, 25-34, 2003.

[8] Masthoff, J. Modeling a group of television viewers. Proceedings of the Workshop Future tv, in Intelligent Tutoring Systems Conference, 34-42, 2002.

[9] Diana, H. J. M., et. al. Relationship of tutors' group-dynamics skills to their performance rating in problem-based learning. *Academic Medicine*, 76, 473-476, 2001.

[10] Das, M., Mpofu, D. F. S., Hasan, M. Y., and Stewart, T. S. Student perceptions of tutor skills in problem-based learning tutorials. *Medical Education* 36, 272-278, 2002.

[11] Feltovich, P. J., and Barrows, H. S. Issues of generality in medical problem solving. In H. G. Schmidt and M. L. De Volder (Eds.) *Tutorials in problem-based learning: A new direction in teaching the health professions*. The Netherlands: Van Gorcum, 1984.

[12] Conati C., Gertner A. and VanLehn K. Using Bayesian Networks to Manage Uncertainty in Student Modeling. *Journal of User Modeling and User-Adapted Interaction*, 12, 371-417, 2002.

[13] Spirtes, P., Glymour, C., and Scheines, R. *Causation, prediction, and search*. Cambridge, MA: MIT Press, 2000.

[14] Dawson, B., and Trapp, R. G. *Basic & Clinical Biostatistics*. 3rd ed. McGraw-Hill, 2001.