



A Bayesian approach to generating tutorial hints in a collaborative medical problem-based learning system

Siriwan Suebnukarn^{*}, Peter Haddawy

Computer Science and Information Management Program, Asian Institute of Technology,
Pathumthani 12120, Thailand

Received 26 July 2004; received in revised form 31 March 2005; accepted 20 April 2005

KEYWORDS

Intelligent tutoring systems;
Computer-supported collaborative learning;
Bayesian networks;
Medicine;
Problem-based learning

Summary

Objectives: Today a great many medical schools have turned to a problem-based learning (PBL) approach to teaching. While PBL has many strengths, effective PBL requires the tutor to provide a high degree of personal attention to the students, which is difficult in the current academic environment of increasing demands on faculty time. This paper describes intelligent tutoring in a collaborative medical tutor for PBL. The main contribution of our work is the development of representational techniques and algorithms for generating tutoring hints in PBL group problem solving, as well as the implementation of these techniques in a collaborative intelligent tutoring system, COMET. The system combines concepts from computer-supported collaborative learning with those from intelligent tutoring systems.

Methods and materials: The system uses Bayesian networks to model individual student clinical reasoning, as well as that of the group. The prototype system incorporates substantial domain knowledge in the areas of head injury, stroke and heart attack. 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. In order to evaluate the appropriateness and quality of the hints generated by our system, we compared the tutoring hints generated by COMET with those of experienced human tutors. We also compared the focus of group activity chosen by COMET with that chosen by human tutors.

Results: On average, 74.17% of the human tutors used the same hint as COMET. The most similar human tutor agreed with COMET 83% of the time and the least similar tutor agreed 62% of the time. 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

^{*} Corresponding author. Tel.: +66 25246577; fax: +66 25245721.
E-mail address: Siriwan.Suebnuakarn@ait.ac.th (S. Suebnukarn).

(McNemar test, $p = 0.652$, $\kappa = 0.773$). The focus of group activity chosen by COMET agrees with that chosen by the majority of the human tutors with a high degree of statistical agreement (McNemar test, $p = 0.774$, $\kappa = 0.823$).

Conclusion: Bayesian network clinical reasoning models can be combined with generic tutoring strategies to successfully emulate human tutor hints in group medical PBL.

© 2005 Elsevier B.V. All rights reserved.

1. Introduction

Over the past few decades, Problem-based learning (PBL) has been introduced as an alternative to the traditional didactic medical education. Traditional pedagogy is subject-based where the students are told what they need to know, a course outline is provided, often a required text is specified, the teacher lectures, and the students solve problems afterwards [1]. In contrast, in the PBL paradigm a problem situation is posed first; the students must identify what they need to know to solve the problem; they learn it and use the new knowledge to solve the problem [2]. Problem-based learning instructional models vary but the general approach is student-centered, small group, collaborative problem-focused learning activities. 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. A number of studies have been conducted comparing the effectiveness of medical education programs based on PBL and traditional models of instruction. In many instances, the conclusion drawn is that PBL has established itself as an effective alternative to traditional instructional models [3,4].

While PBL has many strengths, effective PBL requires the tutor to provide a high degree of personal attention to the students to recognize when and where they most need help, and to guide them to find their own answers [5,6]. Whereas an instructor can deliver a lecture to a large lecture hall of students, a typical PBL tutorial session includes only six to eight students and lasts around 2 h. 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. In contrast to traditional com-

puter aided educational programs, ITS have the ability to adapt to each individual student. Such adaptation requires a student model where all the information about the student is stored, including his/her cognitive problem solving state [7]. The ITS generate tutoring dialogue, and select proper topics or questions based on the student's cognitive state, inferred from student behavior during interaction with the system. Work in the area of ITS has focused on tutoring of individual students, e.g. ANDES [8], SQL-tutor [9], CIRCSIM [10], Slide tutor [11], and thus on modeling the cognitive state of individual students. In contrast, intelligent tutoring for PBL requires the ability to model the dynamics of the group as well.

Researchers in the area of computer supported collaborative learning (CSCL) have developed numerous computer supported PBL environments over the past decade [12,13]. This work has focused mainly on providing and maintaining shared information resources and shared workspaces rather than on intelligent tutoring. More recent work has attempted to integrate CSCL with some form of user modeling. For example, the Docs 'n Drugs project [14] 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.

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. 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 the user modeling techniques and tutoring algorithms used in COMET, a collaborative intelligent tutoring system for medical

problem-based learning. This work was first reported in the Ninth International Conference on Intelligent User Interfaces (IUI-04), 2004 [15]. The system uses Bayesian networks (BNs) to model individual student knowledge and activity, as well as that of the group. The BN models are used by the tutoring module to generate tutorial hints that guide group problem solving activity.

Medical PBL is challenging due to the complexity of the knowledge involved, the lack of standard, commonly accepted student problem-solving techniques, and the lack of standards for tutoring. Thus, one objective of the work presented in this article has been to study human-tutored PBL sessions in order to identify prototypical patterns of student clinical reasoning, as well as tutoring strategies. This knowledge was then used to create individual and group student models, as well as tutoring algorithms that use the models along with the identified tutoring hint strategies to generate the various tutoring hints.

The rest of the paper is organized as follows. Section 2 discusses the medical PBL process and the role of the students, the scenario, and the tutor. Section 3 provides a brief overview of COMET, describing its various components. Section 4 describes the acquisition of and inference with the BN clinical reasoning model. Section 5 describes the tutoring strategies and their implementation. Section 6 presents results of a comparison study of tutoring hints generated by COMET with those of experienced human tutors, as well as a validity study of the student model in determining group activity. Section 7 discusses related work and Section 8 presents conclusions and directions for future work.

2. Medical PBL

According to Barrows [2], PBL can be described as “the learning that results from the process of working toward the understanding or resolution of a problem”. In PBL it is the problem that drives the learning. In medical PBL settings students are presented with real-life cases that require defining and analyzing the problem, creating hypotheses, gathering and analyzing data, and evaluating or justifying solutions collaboratively. Since Barrows originally introduced the concept of PBL in medical education, a number of variations of the classical model have emerged. Below we describe a distillation that applies to almost all forms of PBL.

Problem-based learning models vary but the general approach in this type of learning is case-based, student-centered, small group (of no more than eight), collaborative problem-focused learning

activities. The group has a tutor who facilitates and guides the group, rather than an expert imparting knowledge [16,17]. Problem-based learning is typically carried out in three phases:

1. *Problem analysis*: This is the first basic phase in the PBL approach where the students develop the cognitive skills necessary for clinical reasoning. In group discussion PBL settings the students evaluate the patient problem presented to them exactly as they would a real patient, attempting to determine the possible underlying anatomical, physiological, biochemical, or behavioral dysfunctions responsible for the patient’s problem. The task of the students is to enumerate all possible causal paths (hypotheses and their causal relations) that would explain the progression from the enabling conditions to the signs and symptoms in the given problem scenario.
2. *Self-directed study*: In this phase, students work outside the tutorial session, using any relevant learning resources, e.g. literature, laboratories, specialists, to address any open issues identified in the first phase.
3. *Synthesis and application of newly acquired information*: The third phase begins when the students return from their self-study period. They analyze data and evaluate or justify solutions collaboratively, and wrap up the problem.

The most important phase is problem analysis where the patient problem is first encountered by the students, before any study occurs in the area of the problem. The scenario describing history and examinations is designed to challenge students to develop reasoning, problem solving and team skills. A group of students with different background knowledge is actively involved in the creation of solutions for the problem. Since the main argument for using collaborative problem solving in medical PBL is the wider range of hypotheses generated, the richer variety of suggestions made and the quality of discussion that ensues; the varied backgrounds, knowledge and skills of the students can be of considerable help in peer group education, and in ensuring that alternative aspects of each issue are discussed.

Consider, for example, the following scenario taken from a PBL session in the brain course at Thammasat University Medical School.

“A 30-year-old male was involved in a car accident while he was driving home. The broken glass window cut deep into his forehead and tore a hole through his skull. Bleeding and unconscious, he was rushed to a hospital where a surgeon operated on him to stop the bleeding.”

The work-up questions on history and examinations are aimed at verifying the correct mechanism responsible for the patient's problem. The students 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. Here students must enumerate possible hypotheses to explain why the patient has become unconscious. Fig. 1 is a photograph of the white board at the end of the group PBL session, showing the hypotheses identified and the causal relations among them.

One of the main issues in PBL is the role of the tutor. A number of studies [5,6] 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 learning 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 [5].

3. COMET environment—overview of components

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 current version of COMET supports the initial problem analysis phase of PBL where there is intensive interaction between students and the system as well as among students in the group. COMET contains four primary components similar to any typical ITS [18]: domain clinical reasoning model (or domain model), student clinical reasoning model (or student model), pedagogic module, and interface which includes student multi-modal interface, medical concept repository. The architecture of COMET differs from that of most ITS's in that the domain model and student model are embodied in one representation.

In this section we describe the aspects of multi-modal interface and medical concept repository that are necessary to understand the design and application of the domain clinical reasoning model, student clinical reasoning model, and pedagogic module which are described in the next section. The system is implemented as a Java client/server combination that can be used over the Internet or local area networks and supports any number of simultaneous PBL groups. Although there is no technical limit on the number of students in a COMET PBL group, for practical pedagogical reasons a PBL group typically consists of no more than eight students.

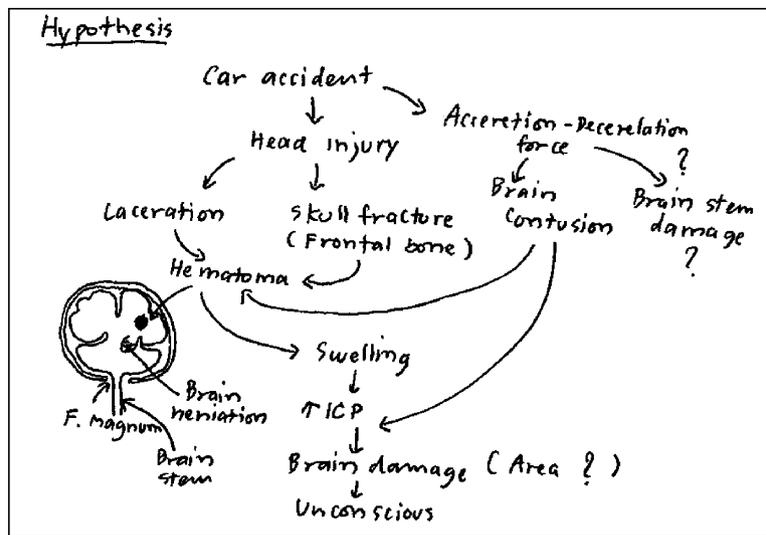


Fig. 1 A depiction of the white board after a PBL session at Thammasat University Medical School. Terms originally written in Thai have been translated into English. The graph shows hypotheses with arrows indicating cause-effect relations among them.

Our prototype incorporates substantial domain knowledge about head injury diagnosis, stroke and heart attack. The system implementation is modular and the tutoring algorithms are generic so that adding a new scenario requires only adding the appropriate model representing how to solve a particular case (domain clinical reasoning model). The student clinical reasoning model, which is a probabilistic overlay of the domain clinical reasoning model is then constructed during runtime by instantiating the nodes that define knowledge and activity of an individual student.

3.1. Multi-modal interface

Live human-tutored medical PBL sessions are facilitated by the use of a whiteboard or similar device where problem analysis and hypotheses can be sketched (Fig. 1). The whiteboard allows these ideas to be edited and expanded as the problem solving process continues. The students are encouraged to express all their ideas in this process and to challenge each other's ideas as they proceed under the guidance of the tutor. The students attempt this on the basis of the knowledge and skills they already

have, aided by handy references such as a medical dictionary and a few appropriate textbooks on anatomy, physiology and the like. COMET emulates this PBL environment by incorporating a multi-modal interface (Fig. 2) that supports communication between students and the system, as well as among students in the group.

The hypothesis board (Fig. 2, lower 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 retrieving them from the medical concept repository. 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.

Students may communicate with others in the group by typing into the chat pane (Fig. 2, middle left pane). COMET has no language processing capabilities, so the text in the chat pane is not taken as input to the system. To support collaborative interaction, the discussion pane (Fig. 2, upper left pane)

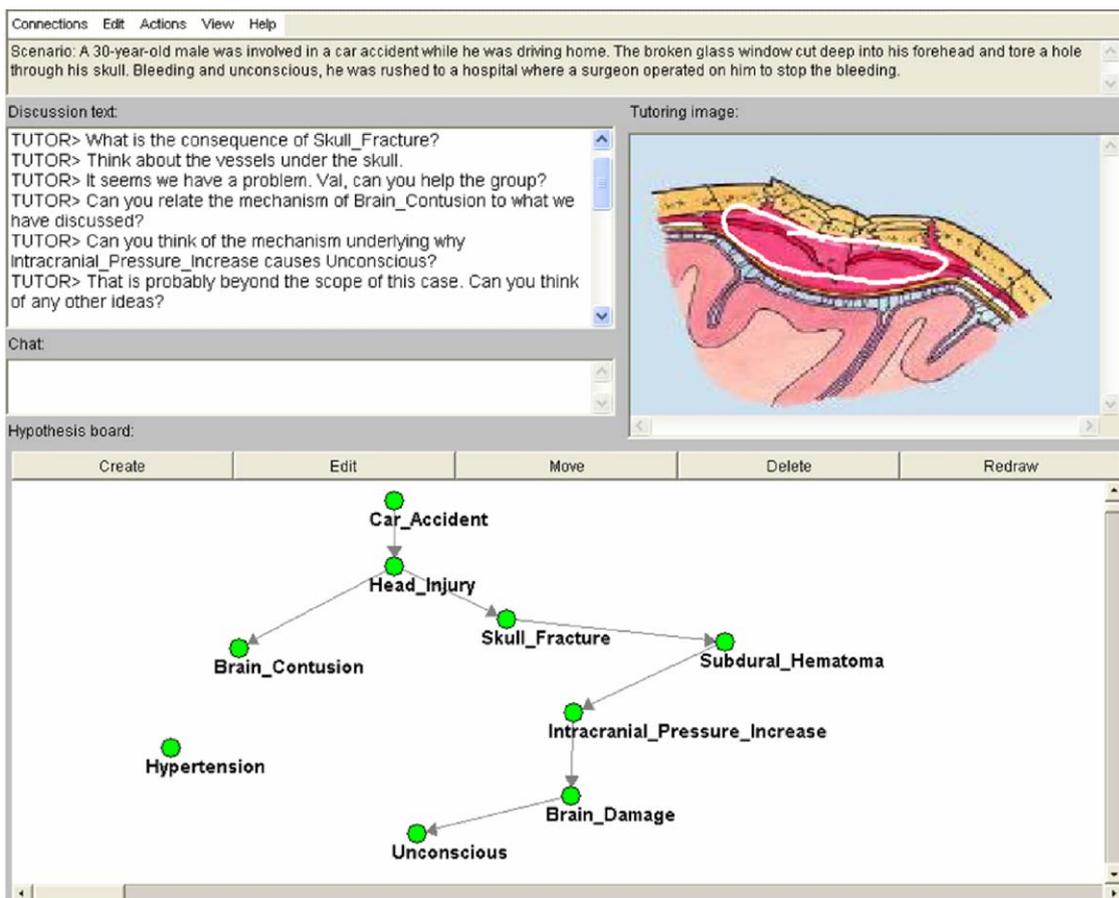


Fig. 2 COMET student interface.

is the place for displaying tutoring hints, and student chat dialogues. The algorithms used to generate the tutoring hints are described in Section 5.

In collaborative learning, a shared workspace plays the important role of supporting learners' construction of knowledge representations. Study of the role of external representations in face-to-face and online collaboration has indicated that direct manipulation of a shared representation supports rich conversation by making it easy to identify the group attention [19]. 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. 2, 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.

3.2. Hierarchical medical concept repository

The system includes a hierarchical medical concept repository (Fig. 3) to help students better under-

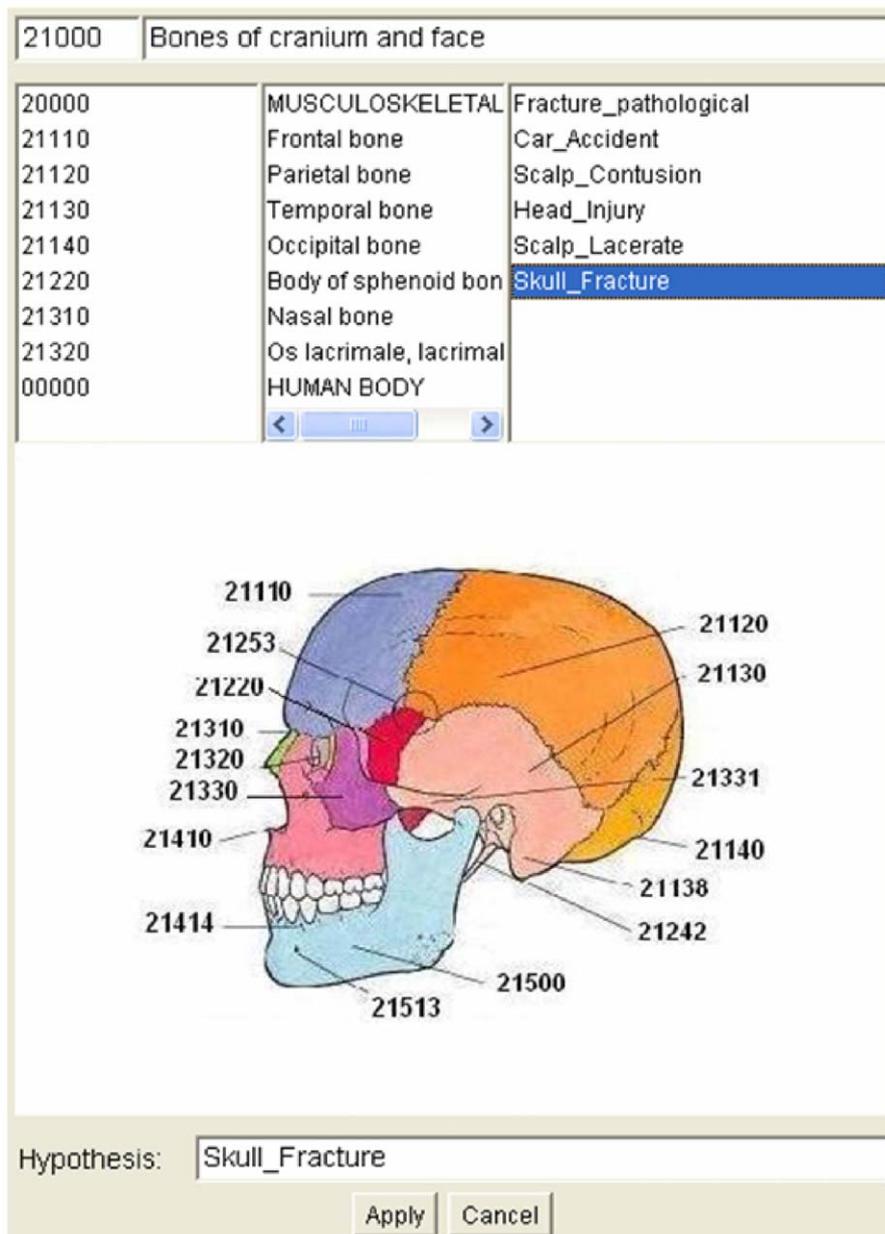


Fig. 3 Medical concept repository showing the result of searching from the whole body (00000) to musculoskeletal system (20000) and bones of cranium and face (21000).

stand the relationships between domain concepts, as well as to facilitate system input. The clinical terminology used in our repository is based on SNOMED version 3.0 [20] and clinically oriented anatomy [21]. Concepts from human anatomy 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. Each five-digit number is connected to the corresponding anatomic part on the image. Students can search through a specific system by clicking on the numeric label of each anatomic part on the image, and drill down to more specific body parts. For example, 00000 refers to the whole body, 20000 refers to the musculoskeletal system, 21000 refers to bones of cranium and face, and 21110 refers to the frontal bone. Fig. 3 shows the result of searching from 00000 → 20000 → 21000. The anatomical image, physiological, biochemical and behavioral dysfunctions relevant to a selected 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 not 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 students to better understand the relationship between the hypothesis and the larger anatomic and patho-physiological context.

4. Domain and student clinical reasoning models

Generating appropriate tutorial actions requires a model of the students' understanding of the problem domain and of the problem solution. However, as in human-tutored PBL sessions, COMET must provide an unrestricted interaction style, which gives the students the freedom to solve the patient case without requiring them to explain the rationale behind their actions. This complicates the modeling task since we have only a limited number of observations from which to infer each student's level of understanding. The resulting need to deal with uncertainty has lead us to use BNs as our modeling technique.

In the traditional ITS architecture, the domain model and student are embodied in separate repre-

sentations [18], while in COMET the BN represents both the domain model and the student model. The domain model is contained in the part of the structure of the network that represents the hypotheses and the cause—effect relations among them. The student model is contained in the part of the network that represents how the hypotheses are derived and in the network's probabilities. The probabilities do not represent the likelihood of occurrence of the hypotheses but rather the likelihood that a student will be able to create the hypotheses.

The following sections describe the BN structure of the domain clinical reasoning model, how the conditional probabilities in the networks are obtained, and how the models are used for individual and collaborative student modeling.

4.1. The BN structure

To construct the model, we used information from various sources. From research papers and textbooks on clinical reasoning and PBL, we obtained the generic structure of the network. This structure includes the application of medical concepts to goals to derive hypotheses, and the classification of hypotheses into enabling conditions, faults and consequences. For each specific scenario, we consulted medical textbooks and experts to obtain the hypotheses, the causal relations among them, the goals and the medical concepts used to derive the hypotheses. The model for each problem scenario required about one-person month to build.

We investigate issues of generality in clinical reasoning, which will serve as a foundation in developing our domain-general structure. The classic model of clinical reasoning is the hypothetico-deductive model [22]. It is characterized by the generation of multiple competing hypotheses from initial patient cues and collection of data to confirm or refute each hypothesis. The steps described in this model are incorporated in the PBL approach necessary to teach clinical reasoning skills to the students [2].

Fig. 1 shows a portion of the hypothesis structure created by one PBL group. It shows a directed acyclic graph representing cause—effect relationships among hypotheses. Since we assume that each student is participating in the process of creating this graph, the graph forms the basis of our student model. The hypothesis graph can be conveniently represented as a BN since BNs are also directed acyclic graphs. In addition, Bayesian networks can represent our uncertainty about the state of knowledge of the students.

To come up with hypotheses explaining the case, the clinical reasoning process involves the following iterative three steps:

1. *Problem identification* is done through a process of selecting and prioritizing problems obtained from studying the case. This process is similar to “subgoaling” in means-ends problem solving [23]. We represent each problem identified during the clinical reasoning process with a “goal” node in the BN model. These “goals” may be diagnostic, treatment, preventative, or psychosocial issues that require attention (Fig. 4).
2. *Problem analyses* are developed for each problem identified in step 1. Students are encouraged to use their previous knowledge to explain the problem. The previous knowledge includes anatomy, patho-physiology, microbiology, pharmacology and clinical experience in diagnosis and treatment planning. We represent this knowledge with “concept” nodes in the BN model (Fig. 4).
3. The *hypotheses* are derived by applying medical concepts from the problem analyses. “Apply” nodes represent the student’s action in applying a “concept” to a “goal” to derive a “hypothesis”. Hypotheses are additionally linked with directed arcs representing the causal relationships among them (Fig. 4).

Although some consistent characteristics of the clinical reasoning process can be identified based on the hypothetico-deductive reasoning model, they

are not particularly satisfying for understanding the process of reasoning, or useful for communicating it when training future clinicians. They are deficient in that they leave critical components of reasoning unexplained. For example, how are good hypotheses generated, and what is the nature of a good hypothesis set? We have explored the clinical problem representation called “illness script” proposed by Feltovich and Barrows [24] and incorporated this approach in the design of our system. At its most general level of description, the script proposes that an illness can be characterized by three component parts: enabling conditions, faults, and consequences. These are abstract categories of illness features, and each can be further specified into more specific categories. 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.

Fig. 4 shows a portion of the BN student model built corresponding to the head injury scenario in

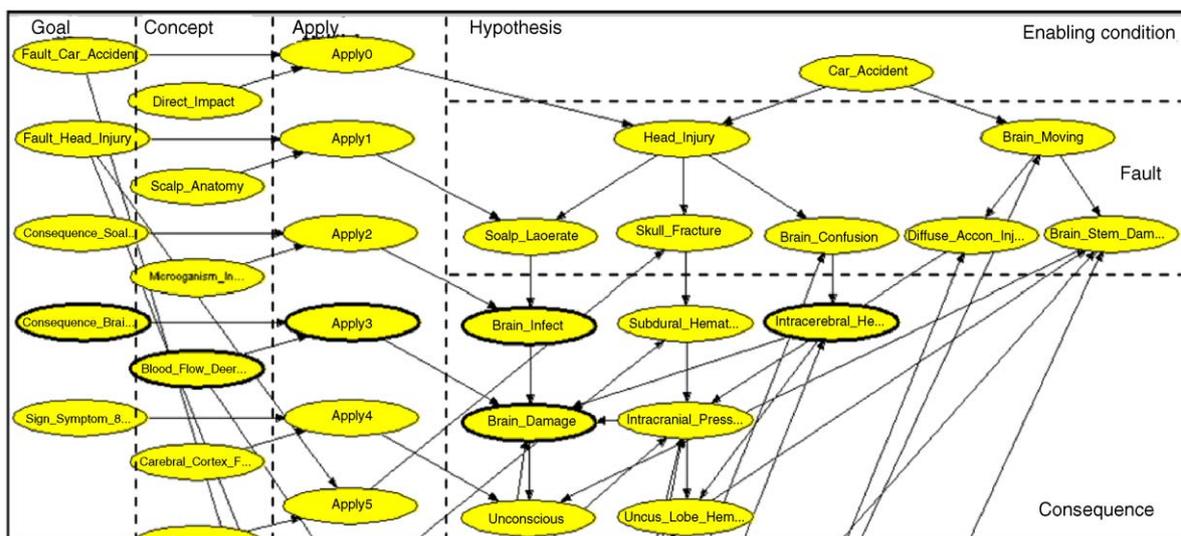


Fig. 4 Part of the Bayesian network student model. The complete network contains 66 nodes. Each hypothesis node has parent nodes, which have a direct causal impact on it. For example, Brain_Damage has parents Brain_Infection and Intracerebral_Hemorrhage. The Apply3 node indicates that the student is able to use knowledge of the Blood_Flow_Decrease to infer that Brain_Damage is a consequence of Brain_Infection.

Fig. 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) and (2) the application of medical concepts in terms of anatomy and patho-physiology (the left group of nodes) to derive the hypotheses. We represent the hypothesis structure following the model of Feltovich and Barrows. In Fig. 4 (right half), we have seven possible faults associated with the single enabling condition car accident: 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. Notice that this means that continuous lab data such as hematocrit or blood pressure must be represented in terms of binary variables. Part of the modeling process is to distill such continuous case data into discrete concepts that build a generalized conceptual understanding.

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 one of its parents. 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. [8]. Our model differs from theirs in the way we represent the internal causal relationships of the hypothesis structure (Fig. 4, right half). The hypothesis nodes in our model correspond to the fact nodes in their model. In their model, each hypothesis 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 hypothesis node can be a parent of a goal node, whereas in our model goals are root nodes.

4.2. The model conditional probabilities

The conditional probability tables for each network were obtained by learning from data obtained from transcripts of PBL sessions. The data for this study consisted of tape recordings and photographs of tutorial sessions for the head injury, stroke, and heart attack cases at Thammasat University Medical School. A total of 15 groups of 3rd year medical students were involved in this study. Each group, consisting of eight students with different backgrounds, was presented with the head injury, stroke and heart attack cases and asked to construct possible hypotheses for the case, under the guidance of a tutor. After the sessions the tapes and the results on the whiteboard were analyzed to determine whether or not each goal, concept, and hypothesis was mentioned. We used the EM learning algorithm [25] provided by the HUGIN Researcher software to learn the conditional probabilities of each node. The validity study of the models is described in Section 6.

4.3. Individual and collaborative student clinical reasoning models

The student clinical reasoning model, which is a probabilistic overlay of the domain clinical reasoning 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 domain clinical reasoning model is instantiated for each student prior to group discussion 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. During group discussion, we make the assumption that once a hypothesis or link in the domain model is created by one student in the group and is accepted by the tutor, every student knows that hypothesis or link is correct and relevant to the current scenario. This assumption rests on two facts. First, all students have enough basic knowledge in anatomy, physiology and pathology to understand ideas suggested by other members. Second, each student knows that if a hypothesis or link is incorrect or a hypothesis is irrelevant, the tutor will intervene. So as hypotheses in the domain are created, they are instantiated in each student model.

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. Work by Jameson et al. [26] and by Lock and Kudenko [27] suggests that a group model is more

than the sum of its parts. They suggest that group models can be created by extracting and combining information from individual user models or by integrating group information into each individual user model. We follow the first approach to model group clinical reasoning in the PBL process by using the individual models to identify the reasoning path that the group is currently focusing on.

Following commonly accepted practice in medical PBL, 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 in that student's model. This is computed as the joint probability of the nodes along the path. For example, suppose we have the following hypotheses entered into the student model: Car_Accident, Head_Injury, Intracranial_Pressure_Increase, unconscious, as shown in Fig. 5. The evidence is entered and propagated, and new beliefs are retrieved. Here we have six candidate paths.

- Path 1: Unconscious ← Brain_Damage ← Brain_Infection ← Scalp_Laceration ← Head_Injury ← Car_Accident

- Path 2: Unconscious ← Brain_Damage ← Intracranial_Pressure_Increase ← Subdural_Hematoma ← Skull_Fracture ← Head_Injury ← Car_Accident
- Path 3: Unconscious ← Brain_Damage ← Intracranial_Pressure_Increase ← Diffuse_Axon_Injury ← Brain_Moving ← Car_Accident
- Path 4: Unconscious ← Brain_Damage ← Intracerebral_Hematoma ← Brain_Contusion ← Head_Injury ← Car_Accident
- Path 5: Unconscious ← Brain_Stem_Damage ← Brain_Moving ← Car_Accident
- Path 6: Unconscious ← Brain_Stem_Damage ← Uncus_Lobe_Herniate ← Intracerebral_Hematoma ← Brain_Contusion ← Head_Injury ← Car_Accident

The most likely current reasoning path for this student is path 2 since it has the maximum joint probability.

Since the students work in a group, 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. We would like to identify a 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. This is done as follows. We identify a set of candidate paths by taking the most likely path for each student. This guarantees that each candidate

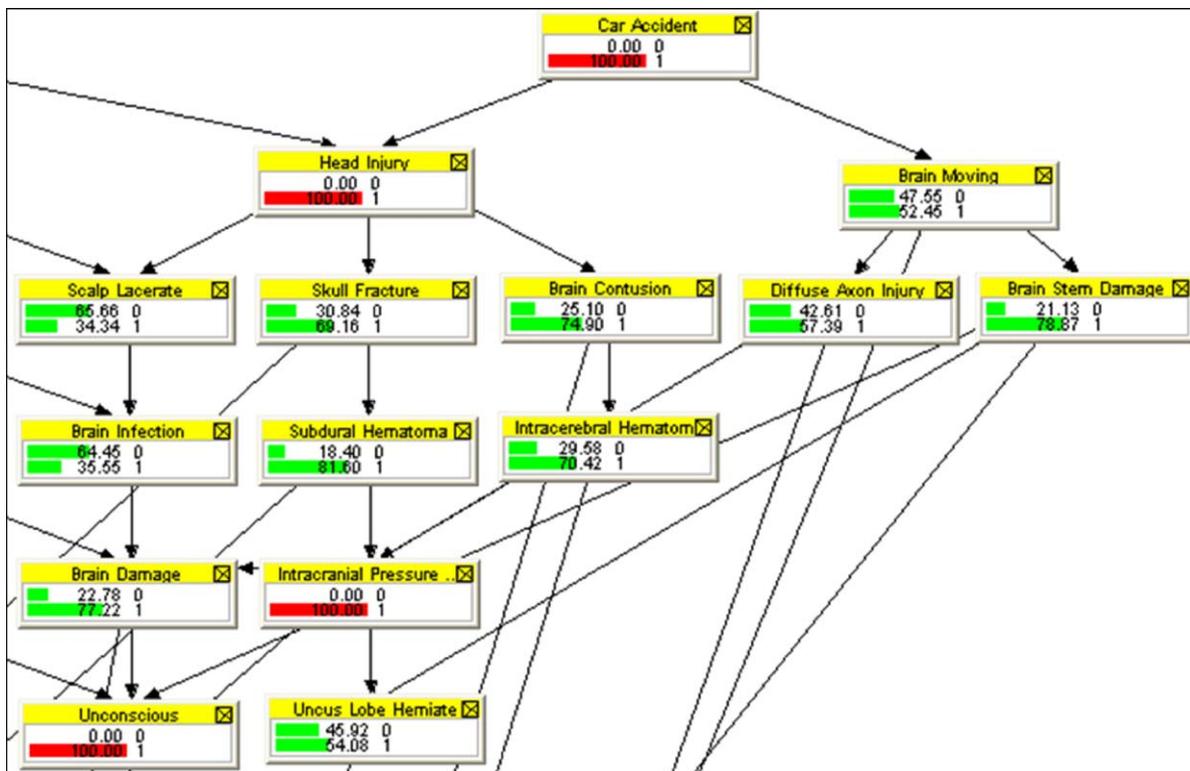


Fig. 5 Status of a student model after observing actions.

path has at least one student currently focused on it. We then compute the sum of the probabilities of each candidate path over all students and select the path with the highest sum. This gives us the candidate path with the highest average attention over all students. In Section 6.2 we compare the group paths generated by this algorithm with those chosen by human tutors.

5. Pedagogic module

Our automated tutor needs to take on the role of guiding the tutorial group to construct possible hypotheses for the case by the use of specific open-ended questions. The tutor should give hints when the group appears to be getting stuck, off track, collaborating poorly, or producing erroneous hypotheses. To do this, the tutor requires knowledge of both the problem domain and the problem solving process.

5.1. Tutoring strategies

Our first step was to identify the strategies used by human tutors. We did this by analyzing tape recordings of tutorial sessions at Thammasat University Medical School as described in Section 4.2. We sought to determine the types of student behavior that triggered tutor intervention and the types of hints that were given in these situations. 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 BN student clinical reasoning models. All strategies except strategies 4 and 8 use both the structure and the probabilities of the BN models. Strategy 4 uses only the structure of the model, while strategy 8 uses only a count of the number of inputs from each student. Strategies 1, 2, 3, 6 make use of the group reasoning path discussed in Section 4.3. Strategies 1–6 have general and specific versions. COMET first gives a general hint using the parent goal node of the hypothesis that it has determined the students should focus on, and if there is no student response or an incorrect response is given, the more specific parent medical concept node is used. This policy is consistent with research on human tutoring [28] that shows human

tutors often start with a general hint and then provide a more specific hint if the student response is still inadequate. Each hint strategy is discussed in turn below.

5.1.1. 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 causal path. The hypotheses that the students are asked to develop should be those of underlying basic mechanisms. Tutors typically first intervene by giving a general hint. If the students do not respond, a more specific hint is given, aimed at helping the students identify the correct mechanism responsible for the patient's problem.

- *Students*: Car accident, skull fracture, brain damage (students in the group started generating the hypotheses by mentioning car accident, skull fracture, and brain damage without mentioning the relationships among them.)
- *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 (e.g., H1, H2, and H3 in Fig. 6). This number represents about 20% 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.

COMET implements this particular strategy as follows. The system first determines the group-reasoning path on which to focus discussion, as described in Section 4.3. The students are considered to have focused if all hypotheses created lie along a particular path, otherwise they have not yet

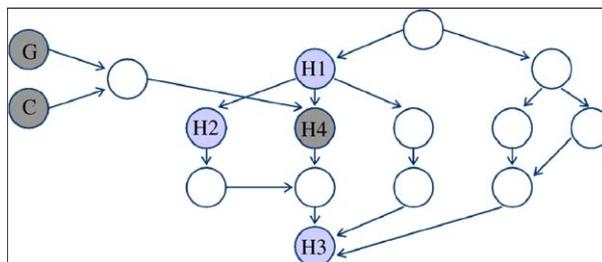


Fig. 6 Focus group discussion.

focused their discussion. The system then selects the hypothesis node from the group path that has not been mentioned and has the highest probability (e.g., H4). Its parent nodes are then used to generate the tutoring hint. The system gives the first general hint using the parent goal node (e.g., G), and if there is no student response observed, the more specific parent medical concept node will be used (e.g., C). The medical image relevant to the parent nodes is displayed along with the tutoring hint.

5.1.2. Strategy 3: create open environment for discussion

While in the process of guiding the group to elaborate on a particular hypothesis and its consequences, the tutor may be presented with a new proposed idea that 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 into the current discussion but to retain focus by encouraging the students to relate the new concept to the current hypothesis under discussion, e.g.

- *Students:* Skull fracture → subdural hematoma → intracranial pressure increase (the notation “skull fracture → subdural hematoma” means that a student draws a link from skull fracture to subdural hematoma on the hypothesis board.)
- *Student:* Brain contusion
- *Tutor:* “Can you relate its mechanism to what we have discussed?”

This strategy is implemented as follows: 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 determines the degree of divergence and responds differently depending on its relation to the current path. We distinguish three different degrees of divergence: mild, moderate and severe. As shown in Fig. 7, a hypothesis (H4) is mildly divergent if either its parent or child node is in the current path (H1 → H2 → H3). In this case, the system first asks the group to find the relationship between the divergent node (H4) and the current path. A hypothesis (H5) is moderately divergent if it is linked through one or more intermediate nodes to the current path. In this case, 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. If the

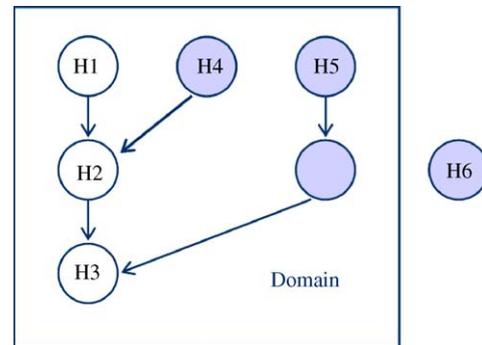


Fig. 7 Three types of divergence.

students cannot suggest a relationship or suggest the wrong one, COMET again generates a more specific hint, focusing on the medical concept node (H3) of the divergent node. A hypothesis (H6) is severely divergent if the domain model does not contain a node corresponding to that hypothesis. In this case, 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.

5.1.3. 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”
- *Students:* Intracranial pressure increase → intracerebral hematoma
- *Tutor:* “Think about the effect on the blood flow”.

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 system first comments to the group that there may be something wrong with their input. If the students still suggest the wrong link, COMET generates a more specific hint, focusing on the medical concept node (C) of the highest probability child node (H3) of the node at the tail (H1) (Fig. 8).

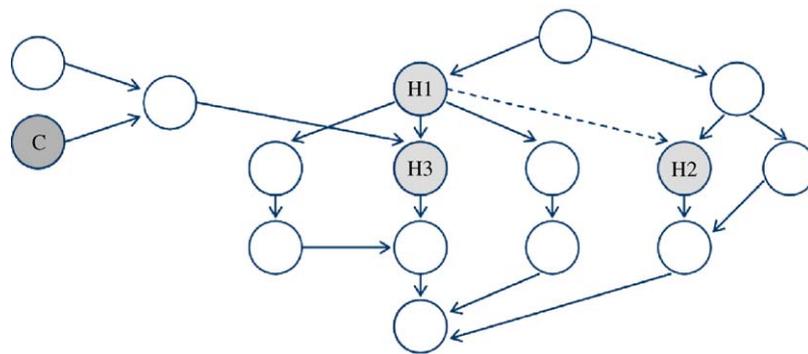


Fig. 8 Uneducated guessing.

5.1.4. Strategy 5: avoid jumping critical steps

According to our observations of PBL sessions, 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 the consequences, they miss exploring the problem scenario in depth. When an intermediate step is left out, the tutor typically first reacts by asking for the consequence of the cause. If the group does not respond, the tutor then gives a more specific hint by pointing the group to its related medical concepts.

- *Students:* Skull fracture → unconsciousness
- *Tutor:* “Can you think of the mechanism underlying why skull fracture causes unconsciousness?”
- *Students:* Skull fracture → brain damage
- *Tutor:* “Think about blood vessel damage under the skull.”

When a new arc is created, COMET first determines whether there are intermediate hypotheses between the head and tail of the arc. For example, in Fig. 9 a student has created a direct arc from hypothesis H1 to hypothesis H4, missing the intermediate hypotheses H2, H3 and H5. If so, the system asks the group to find the mechanism underlying why the hypothesis at the head (H1) causes the hypothesis at the tail (H4). If the students suggest the wrong mechanism, COMET generates a more specific hint by pointing the group to the medical concept node of the highest probability intermediate node.

5.1.5. 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 gives a hint asking the group to discuss new hypotheses explaining the case. Similar to strategies 1 and 2, after the students suggest new hypotheses, the system identifies a new highest probability group-reasoning path from the rest of candidate paths to focus the group discussion.

5.1.6. Strategy 7: refer to expert in the group

Problem-based learning is 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 can help to continue the group learning process if the group gets stuck and remains stuck even when the tutor tries to provide more specific hints.

- *Students:* Car accident → brain moving → unconsciousness

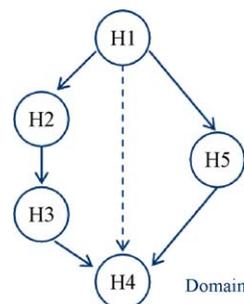


Fig. 9 Jumping critical steps.

- *Tutor*: “Can you think of the mechanism underlying why brain moving causes unconsciousness?”
- *Students*: Silent
- *Tutor*: “What happens to the brain stem?”
- *Students*: Silent
- *Tutor*: “It seems like we have a problem. Val, you are good at physiology; could you help the group?”

This strategy is implemented by examining each individual student model, to find the student who has the highest probability of knowing the hypothesis node that was the focus of the previous hint. The differences in background knowledge, represented in the individual student models, result in differences in the estimated likelihoods that the various students know the hypothesis.

5.1.7. 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 [29]. 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 student activity. Any actions consisting of drawing on the hypothesis board or annotating the tutoring image are saved in the activity log file on the server. If we have a group of n students and there is no input from a particular student after n hypotheses have been created, the system should prompt that student. To avoid having one student dominate the discussion, each student is allowed to suggest up to three hypotheses consecutively.

5.2. Selecting strategy when multiple strategies apply

There are various situations in which two hint strategies may apply. In the literature on automated tutoring, this is dealt with in various ways. For example, Heffernan and Koedinger [30] considered student errors in the order they would be encountered in a post-order traversal of the parse tree of the correct answer. Therefore, multiple questions are added to the tutorial agenda depending upon the tutorial strategy selected for each error. In our case, after determining the most appropriate response in every situation in which two of COMET’s hint strategies apply, we found that the responses fall into two categories of policies: ordering and

precedence. An ordering policy gives both hints, determining which should be given before the other. This is similar to meta-strategy used by Major [31] to determine pedagogical parameters (such as severity of errors, or whether prerequisites are required) under which each strategy is prioritized. We now give two examples of ordering policies. In our discussion below, we will distinguish between the general hint for a strategy and the specific hint by using the notation $x(g)$ for strategy x general hint and $x(s)$ for strategy x specific hint. Strategy 3 (create open environment for discussion) covers the cases of mild, moderate and severe divergence, which we will refer to as 3.1, 3.2 and 3.3, respectively. So, for example, the general hint for moderate divergence would be 3.2(g).

- There are many situations in which strategy 3.3(g) may be triggered at the same time as other strategies. Since 3.3(g) is triggered when a student creates a node outside the scope of the problem, this must be pointed out whenever it happens. Therefore, in all cases the hint for 3.3(g) is given first, followed by the hint for the other applicable strategy.
- Strategy 8 (promote collaborative discussion) can be triggered in conjunction with any of the other strategies. When this happens, the hint for the other strategy is given first indicating the content of the feedback, followed by referring to a specific group member using strategy 8.

In situations where a general hint has been given and then a more specific hint or a different hint could be given, the more specific hint generally takes precedence. This is in line with the tutor-centered pedagogic principle [28] in which the tutor continues his plan by using the follow-up (specific) hint of the previous (general) hint. Any new error that a student makes is considered as an incorrect response to the previous general hint. The following are three examples of such precedence policies.

- When strategy 1 is given and the student creates a node, besides 3.3 and 8, the only other possible strategies that the new node could triggered are 3.1(g) and 3.2(g). In this case, strategy 2 (the follow-up to 1) will take precedence over 3.1(g) and 3.2(g).
- When strategy 3.1(g) is given, the new node could trigger a different application of 3.1(g), call it 3.1(g)’, or it could trigger 3.2(g) or 3.3(g). In this case, 3.1(s) (the follow-up to 3.1(g)) takes precedence over 3.1(g)’, 3.2(g), and 3.3(g).
- When students create new node or link, it is possible for 4(g) or 5(g) to be triggered along with

the specific version of the previous general hint. In these cases, the specific hint takes precedence.

5.3. Sample interaction

We show an example of five students interacting with COMET. One of them, Val, is an anatomist. Fig. 1 shows a screen shot of the interface for one

student. Although the current version of COMET has no language processing capabilities, and the system has access to only information entered on the hypothesis board, it can generate a variety of hints that help the students to generate hypotheses by themselves. Observations of group interactions with the system showed that the students gradually created hypotheses while they identified and analyzed each problem in the chat pane. The feedback given by the system influenced the ensuing discussion and

Table 1 Sample interaction (student input from the hypothesis board and tutor dialogues from the discussion pane)

	Student input and tutor dialogues	Explanation of student actions and tutoring strategies
Students	Car_Accident, Skull_Fracture, Intracranial_Pressure_Increase	Students in the group gradually create the hypotheses car_Accident, Skull_Fracture and Intracranial_Pressure_Increase on the board, while analyzing the problem
Tutor	What is the consequence of Skull_Fracture?	<i>Strategy 1:</i> COMET focuses group discussion by identifying which causal path the group should focus on, finding the highest probability non-mentioned node along the path (Subdural_Hematoma), and providing a hint using its parent goal node (Consequence_Skull_Fracture)
Students	Brain_Damage	Brain_Damage is a node along the group reasoning path but not the node that COMET wants the group to focus on
Tutor	Think about the vessels under the skull	<i>Strategy 2:</i> COMET gives the next hint using the medical concept parent node of Subdural_Hematoma
Students	Unconscious	Unconscious is a node along the group reasoning path but not the node that COMET wants the group to focus on
Tutor	It seems we have a problem. Val, can you help the group?	<i>Strategy 7:</i> Val is called on since she has the highest probability of knowing the Subdural_Hematoma node among the students
Val	Subdural_Hematoma	Subdural_Hematoma is correct. COMET gives no hint
Students	Skull_Fracture → Subdural_Hematoma → Intracranial_Pressure_Increase	Students draw the correct links between the hypotheses. COMET gives no hint
Student	Head_Injury	Head_Injury is a node along the group reasoning path. COMET gives no hint
Student	Brain_Contusion	Brain_Contusion is mildly divergent from the group reasoning path
Tutor	Can you relate the mechanism of Brain_Contusion to what we have discussed?	<i>Strategy 3:</i> COMET gives the first hint asking the group to connect the mildly divergent hypothesis to the group reasoning path
Students	Head_Injury → Brain_Contusion	Students draw the correct link between the hypotheses. COMET gives no hint
Students	Intracranial_Pressure_Increase → unconscious	Students draw a link jumping from Intracranial_Pressure_Increase to unconscious
Tutor	Can you think of the mechanism underlying why Intracranial_Pressure_Increase causes unconscious?	<i>Strategy 5:</i> COMET gives the first hint asking the group to think about the nodes between Intracranial_Pressure_Increase and unconscious
Students	Intracranial_Pressure_Increase → Brain_Damage → unconscious	Students draw the correct link between the hypotheses. COMET gives no hint
Student	Hypertension	Hypertension is not in the domain
Tutor	That is probably beyond the scope of this case. Can you think of any other ideas?	<i>Strategy 3:</i> COMET responds to Hypertension which is severely divergent from the group reasoning path

guided the students to new hypotheses. Table 1 shows the interactions with the system after the students read the problem scenario. The system selects hint strategies based on the student input on the hypothesis board.

6. 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. Since the group path forms the basis for several of the hint strategies, we also compared the group path generated by COMET to the paths suggested by human tutors.

6.1. Evaluation of the hints

6.1.1. Experimental design

We recruited ten tutors from the faculty at Thammasat University Medical School: three physiology majors, two anatomy majors, one anatomy major, one pathology major, one pharmacology major, and two neurosurgeons. Each tutor had at least 5 years experience in conducting the brain course at Thammasat. All of them were well-grounded and skilled medical PBL tutors.

For each of the eight hinting strategies discussed in the previous section, we created three situations in which COMET would generate that hint, giving a total of 24 test scenarios. Video clips of each scenario were taken from COMET's client application window. Each video clip showed only the hypothesis board and scenario description, so that the information available to the human tutors was the same as that available to COMET. The ten tutors were each asked to analyze the video clip of each scenario 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.

6.1.2. Results

Table 2 shows the percentage of human tutors using various strategies for each strategy used by COMET. For example, the first row indicates that in the situations in which COMET used strategy 1, 60% of the human tutors also used strategy 1 with the same tutoring content, while 7% used strategy 2, 20% used strategy 4, and the remaining 13% used strategy 1 but referred to a different tutoring goal to focus the group. The degree of agreement between COMET and the human tutors differed

Table 2 Results comparing COMET and human tutor hints

COMET's strategy	Human tutors' strategy (% of tutors suggesting the strategy)			
1	1 (60)	1 ^a (13)	2 (7)	4 (20)
2	2 (60)	1 (33)	3 (7)	
3	3 (87)	4 (13)		
4	4 (87)	5 (7)	1 (6)	
5	5 (80)	5 ^a (13)	6 (7)	
6	6 (67)	5 (7)	No hint (26)	
7	7 (60)	4 (13)	5 (20)	
8	8 (93)	8 ^a (7)		

The first column under human tutors' strategy indicates that percentage of human tutors whose hint strategy and content matched that of COMET.

^a Same strategy but different content.

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 in the situation that the group jumped critical steps, 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 repeat the avoiding jumping a critical step strategy, using a yet more specific hint. On average, 74.17% of the human tutors used the same hint strategy and content as COMET. The most similar human tutor agreed with COMET 83% of the time and the least similar tutor agreed 62% of the time.

To test the statistical significance of the agreement between the system and the human tutors, we used the McNemar test and Kappa statistic, which are commonly used in medicine to determine the degree of agreement between two alternative testing procedures. Kendall's W coefficient of concordance was used to summarize the agreement among human tutors [32]. There were no statistical differences between the human tutors and COMET (McNemar test, $p = 0.652$). Additionally, there was a high degree of agreement between the hints generated by COMET and by the human tutors (κ -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. There was a high degree of agreement among the responses of the various human tutors with Kendall's W coefficient of concordance among human tutor responses = 0.780. Our results indicate that the responses generated by COMET are in line with the majority consensus responses of the human tutors tested.

Table 3 Results comparing COMET and human tutor group paths

Scenario	COMET's path	Human tutors' path (% of tutors suggesting the path)
Head injury	12	12 (85) 14 (15)
	14	14 (70) 12 (10) Others (20)
Stroke	23	23 (85) 24 (10) Others (5)
	24	24 (60) 23 (20) Others (20)
Heart attack	31	31 (90) 32 (10)
	32	32 (70) 31 (20) Others (10)

6.2. Evaluation of the group path

6.2.1. Experimental design

In the group path study, we simulated three groups of five students from different backgrounds. Partial solutions were created by randomly giving 50% of the hypothesis nodes and their causal links in each scenario. Ten tutors were asked to identify the reasoning path that the group should follow for each scenario and each group, given the partial solutions and the information about students' background knowledge. We compared the group reasoning path generated by COMET to the path suggested by 10 human tutors for three scenarios and three groups. This gave us 90 data points for comparison. The total number of reasoning paths containing at least one node that was mentioned for the head injury, stroke and heart attack scenarios ranged from 6 to 125.

6.2.2. Results

Table 3 shows that COMET's group paths are in line with the majority consensus of the human tutors. For example, in the situation where COMET generated path no. 12, 85% of the tutors also suggested the same path, and 15% suggested path 14. There were no statistically significant differences between the human tutors and COMET (McNemar test, $p = 0.774$). The results showed a high degree of agreement between the group path generated by COMET and by the human tutors (κ -index = 0.823). There was a high degree of agreement among the group paths suggested by various human tutors (Kendall's W coefficient of concordance among human tutor responses = 0.979).

7. Discussion

Some recent work has attempted to integrate computer-supported collaborative environments with some form of user modeling. Jameson et al. [26] propose a generative model of individual group mem-

bers, 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 [27] 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 [33] addresses the issue of combining models of individuals' preferences in order to infer group preferences. The work is illustrated with the problem of selecting appropriate television programming for a group. In designing a complex instructional module of computer supported collaborative learning, Soller [34] used plan recognition to model student actions on a shared workspace and infer the presence of behaviors such as conflict and coordination. Inaba and Mizokuchi [35] proposed an ontology-based approach to clarifying behavior and roles for learners. They showed how this can be used to assign an appropriate role to each learner, to generate the conditions for assigning that role, and to reason about the educational benefits of assigning the roles. They built a system that supports group formation by identifying an appropriate learner for each role in collaborative learning settings. Our work departs from previous efforts to incorporate user modeling into computer supported collaborative learning environments by focusing on modeling individual and group problem solving behavior.

In developing our domain clinical reasoning model, we drew inspiration from the Andes ITS. The Andes BN student model represents an individual student's domain knowledge as well as his learning goals and plans [36]. The model is used not only to recognize the student's plans and goals but also to predict the student's inferences [8]. Even though our medical domain is different from very structured domains such as math and physics, where problem solving amounts to applying one of a set of formulas or techniques to work toward a solution, our results from the study of medical PBL sessions and expert problem representation indicate that the abstractions in medical reasoning are goal-oriented, using medical basic science concepts to derive hypotheses or solutions. So, we use the same basic idea as in Andes for structuring the general student clinical reasoning model. Our model differs from the one in Andes in two ways. First, Andes and COMET represent

the internal causal relationships of the hypothesis structure differently. Second, in Andes the conditional probability tables are created by hand.

Facilitating group PBL through a collaborative intelligent tutor requires a system to support the spectrum of activities that groups engage in while learning. Tuckman [37] identifies the “stages of group development” as forming, storming, norming and performing. He identifies the group facilitator strategies involved in these stages as orientation, conflict resolution, development of group cohesion, and insight, respectively. The PBL tutoring strategies that we have identified and implemented correspond closely to these well-known strategies. Orientation is the process of becoming familiar with the students’ backgrounds. COMET’s abilities to enter each student’s background into the individual student model and to refer to experts in the group are instances of orientation. Strategy 3 “create open environment for discussion” is a form of conflict resolution. The objective of our “focus group discussion” and “address incomplete information” strategies is to develop group cohesion. Tuckman’s insight strategy covers the stage in which group members generalize and apply learned issues to other situations. This stage is beyond the scope of COMET.

As described by Shulman [38], pedagogical competence is a tightly-linked combination of pedagogical knowledge (knowing how to teach) and pedagogical content knowledge (knowing what to teach). The COMET strategies discussed above are all aspects of knowing how to teach, or focusing on the learning process. COMET’s strategies “deflect uneducated guessing” and “avoid jumping critical steps”, as well as the use of specific hints referring to medical concepts are examples of knowing what to teach.

To develop the collaborative tutoring algorithms, we use question templates for hint generation. This idea is similar to the hints from CIRCSIM-tutor [10], which uses short template questions in order to conduct a dialogue about the student’s qualitative analysis of a cardiophysiological feedback system. The main difference is that our analysis of students’ responses is based primarily on the information about the structure and probabilities of the BN individual student models and the group reasoning path, while the assessment model of CIRCSIM is rule-based. Moreover, our tutoring strategies focus on collaborative group problem solving.

8. Conclusions and future work

In this paper we have described representational techniques and algorithms for generating tutoring

hints in an intelligent tutor for medical PBL. We have described a BN clinical reasoning model that integrates hypothesis structure based on the illness script with the application of the relevant medical concepts in the problem solving process. We model each student with an instance of our general domain clinical reasoning model. The group is reasoned about by combining information from the models of the individual students. The generality of the representation is demonstrated by our ability to implement three qualitatively different scenarios using the same general network structure. Examples from classroom discourse were analyzed and eight specific types of tutoring hint strategies identified. We developed generic algorithms that can replicate the eight identified tutoring hint strategies, taking as input the BN student model and the student actions. The developed techniques have been implemented and integrated into the COMET group tutoring system. COMET is able to provide tutorial hints that guide individual students and the group as a whole. The results of the evaluations of the accuracy of the student models in determining the group-reasoning path in three different scenarios provide encouraging support for the framework of COMET’s clinical reasoning model. Empirical evaluation shows a high degree of agreement between the hints generated by COMET and those of experienced human tutors.

We plan to extend and improve the system in several directions and to perform more extensive evaluation. COMET can currently support PBL problem analysis in the domains of head injury, stroke and heart attack. We plan to develop 47 more domains to cover all essential knowledge that medical students must learn in the pre-clinical course as taught at Thammasat University Medical School. Since creating the domain model is not a trivial task and requires significant expert knowledge, we plan to develop an authoring system by employing medical resources like SNOMED and the Unified Medical Language System (UMLS) Semantic Network to assist in creation of new cases.

One limitation of this work is that students’ interactions in the chat tool are beyond the capability of the student-modeling module to interpret. It would be useful to add some language processing capabilities so that COMET can track and comment on the discussion. We would also like to add the ability to communicate with the system through anatomical sketches. As can be seen in the bottom left-hand corner of Fig. 3, it is common for students to make sketches in the course of analyzing a PBL scenario. COMET should support this form of communication by not only providing a white board but by also being able to parse the sketches so that the

tutor can follow and comment on the students' problem solving process. 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.

Acknowledgements

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 Medical 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. The constructive comments of the two anonymous reviewers greatly helped to improve this paper.

References

- [1] Ridderikoff J. Medical problem solving: an exploration of strategies. *Med Educ* 1991;25:196–207.
- [2] Barrows HS. A taxonomy of problem-based learning methods. *Med Educ* 1986;20:481–6.
- [3] Patel VL, Kaufman DR, Arocha JF. Conceptual change in the biomedical and health sciences domain. In: Glaser R, editor. *Advances in instructional psychology: educational design and cognitive science*. Hillsdale, NJ: Lawrence Erlbaum Associates; 2000. p. 329–92.
- [4] Schmidt HG, Machiels-Bongaerts M, Hermans H, Cate TJ, Venekamp R, Boshuizen HPA. The development of diagnostic competence: comparison of problem-based, and integrated, and a conventional curriculum. *Acad Med* 1996;71:658–64.
- [5] Dolmans DH, Wolfhagen IH, Scherpbier AJ, Vleuten CP. Relationship of tutors' group-dynamics skills to their performance rating in problem-based learning. *Acad Med* 2001;76:473–6.
- [6] Conati C, Gertner A, VanLehn K. Using Bayesian networks to manage uncertainty in student modeling. *J User Model User-adapted Interact* 2002;12:371–417.
- [7] Shute VJ. SMART: student modeling approach for responsive tutoring. *J User Model User-adapted Interact* 1995;5:1–44.
- [8] Crowley RS, Medvedeva OP. A general architecture for intelligent tutoring of diagnostic classification problem solving. In: Musen M, editor. *Proceedings of AMIA Symposium*. 2003. p. 185–9.
- [9] Mitrovic A. Experiences in implementing constraint-based modeling in SQL-tutor. In: Goettl BP, Halff HM, Redfield CL, Shute VJ, editors. *Proceedings of the Fourth International Conference in Intelligent Tutoring Systems*. 1998. p. 414–23.
- [10] Martha W, Brandle S, Chang R, Freedman R, Glass M, Lee YH, et al. CIRCSIM-tutor: an intelligent tutoring system using natural language dialogue. In: Moore FW, editor. *Proceedings of Midwest AI and Cognitive Science Conference*. 2001. p. 16–23.
- [11] Das M, Mpofu DJS, Hasan MY, Stewart TS. Student perceptions of tutor skills in problem-based learning tutorials. *Med Educ* 2002;36:272–8.
- [12] Mahling DE, Sorrows BB, Skogseid I. A collaborative environment for semi-structured medical problem based learning. In: Schnase JL, Cunniss EL, editors. *Proceedings of Computer-supported Collaborative Learning*. 1995. p. 218–24.
- [13] Suthers D, Toth E, Weiner A. An integrated approach to implementing collaborative inquiry in the classroom. In: Hall R, Miyake N, Enyedy N, editors. *Proceedings of Computer-supported Collaborative Learning*. 1997. p. 272–9.
- [14] Martens A, Bernauer J, Illmann T, Seitz A. Docs 'n Drugs—the virtual polyclinic: an intelligent tutoring system for web-based and case-oriented training in medicine. In: Bakken S, editor. *Proceedings of the AMIA Symposium*. 2001. p. 433–7.
- [15] Suebnukarn S, Haddawy P. A collaborative intelligent tutoring system for medical problem-based learning. In: Nunes NJ, Rich C, editors. *Proceedings of the Ninth International Conference on Intelligent User Interface*. 2004. p. 14–21.
- [16] Barrows HS. Problem-based learning in medicine and beyond: a brief overview. In: Wilkerson L, Gislaers WH, editors. *Bringing problem-based learning to higher education: theory and practice*. San Francisco, CA: Jossey-Bass; 1996. p. 3–12.
- [17] Koschmann T, Kelson AC, Feltovich PJ, Barrows HS. Computer-supported problem-based learning: a principled approach to the use of computers in collaborative learning. In: Koschmann T, editor. *CSCL: theory and practice of an emerging paradigm*. Mahwah, NJ: Lawrence Erlbaum Associates; 1996. p. 83–124.
- [18] Wenger E. *Artificial intelligence and tutoring systems—computational and cognitive approaches to the communication of knowledge*. CA: Morgan Kaufmann, 1987.
- [19] Suthers D, Girardeau L, Hundhausen C. Deictic roles of external representations in face-to-face and online collaboration. In: Wasson B, Ludvigsen S, Hoppe U, Ulrich, editors. *Proceedings of Computer-supported Collaborative Learning*. 2003. p. 173–82.
- [20] Cote RA, Rothwel DJ, Beckett RS, Palotay JL. SNOMED international: the systematized nomenclature of human and veterinary medicine. Northfield, IL: College of American Pathologists, 1993.
- [21] Moore KL, Dalley AF. *Clinically oriented anatomy*. Philadelphia: Lippincott William and Wilkins, 1999.
- [22] Elstein AL, Shulman LS, Sprafka SA. *Medical problem solving—an analysis of clinical reasoning*. MA: Harvard University Press, 1978.
- [23] Anderson JR. Problem solving and learning. *Am Psychol* 1993;48:35–44.
- [24] Feltovich PJ, Barrows HS. Issues of generality in medical problem solving. In: Schmidt HG, De Volder ML, editors. *Tutorials in problem-based learning: a new direction in teaching the health professions*. Assen, Maastricht: Van Gorcum; 1984. p. 128–42.
- [25] Lauritzen SL. The EM-algorithm for graphical association models with missing data. *Comput Stat Data Anal* 1995;1:191–201.
- [26] Jameson A, Baldes S, Kleinbauer T. Generative models of group members as support for group collaboration. In:

- Gaudioso E, editor. Workshop on User and Group Models for Web-based Adaptive Collaborative Environments, Proceedings of the Ninth International Conference on User Modeling. 2003. p. 1–14.
- [27] Lock Z, Kudenko D. Multi-component user models of team members. In: Gaudioso E, editor. Workshop on User and Group Models for Web-based Adaptive Collaborative Environments, Proceedings of the Ninth International Conference on User Modeling. 2003. p. 25–34.
- [28] Chi MTH, Siler S, Jeong H, Yamauchi T, Hausmann R. Learning from human tutoring. *Cogn Sci* 2001;25:471–533.
- [29] Barrows HS. The tutorial process. Springfield, IL: Southern Illinois University School of Medicine, 1994.
- [30] Heffernan NT, Koedinger KR. An intelligent tutoring system incorporating a model of an experienced human tutor. In: Cerri SA, Gouardères G, Paraguaçu F, editors. Proceedings of the Sixth International Conference on Intelligent Tutoring Systems. 2002. p. 596–608.
- [31] Major N. Modeling teaching strategies. *J AI Educ* 1995;6: 117–52.
- [32] Dawson B, Trapp RG. Basic and clinical biostatistics. New York: McGraw-Hill, 2001.
- [33] Masthoff J. In: Cerri SA, Gouardères G, Paraguaçu F, editors. Modeling a group of television viewers. Proceedings of the Workshop Future TV, Proceedings of the Sixth International Conference on Intelligent Tutoring Systems. 2002. p. 34–42.
- [34] Soller A. Computational modeling and analysis of knowledge sharing in collaborative distance learning. *User Model User-adapted Interact* 2004;14:351–81.
- [35] Inaba A, Mizoguchi R. Learners' roles and predictable educational benefits in collaborative learning an ontological approach to support design and analysis of CSCL. In: Lester JC, Vicari RM, Paraguaçu F, editors. Proceedings of Intelligent Tutoring Systems. 2004. p. 285–94.
- [36] VanLehn K, Martin J. Evaluation of an assessment system based on Bayesian student modeling. *Int J Artif Intell Educ* 1998;8:179–221.
- [37] Tuckman BW. Developmental sequence in small groups. *Psychol Bull* 1965;63:384–99.
- [38] Shulman L. Paradigms and research programs in the study of teaching: a contemporary perspective. In: Wittrock MC, editor. Handbook of research on teaching. New York: Mac-Millan; 1986. p. 3–36.