Modeling Individual and Collaborative Problem Solving in Medical Problem-Based Learning

Siriwan Suebnukarn and Peter Haddawy

Computer Science and Information Management Program,
Asian Institute of Technology, Pathumthani,
12120, Thailand
`{Siriwan.Subnukarn, haddawy}@ait.ac.th`

Abstract. Since problem solving in group problem-based learning is a collaborative process, modeling individuals and the group is necessary if we wish to develop an intelligent tutoring system that can do things like focus the group discussion, promote collaboration, or suggest peer helpers. We have used Bayesian networks to model individual student knowledge and activity, as well as that of the group. The validity of the approach has been tested with student models in the areas of head injury, stroke and heart attack. Receiver operating characteristic (ROC) curve analysis shows that, the models are highly accurate in predicting individual student actions. Comparison with human tutors shows that group activity determined by the model agrees with that suggested by the majority of the human tutors with a high degree of statistical agreement (McNemar test, $p = 0.774$, Kappa = 0.823).

1 Background

Over the past few decades, problem-based learning (PBL) has been introduced as an alternative to traditional didactic medical education. PBL is designed to challenge learners to build up their knowledge and develop effective clinical reasoning skills around practical patient problems. PBL instructional models vary but the general approach is student-centered, small group, collaborative problem solving activities [2]. 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. We are investigating the potential use of concepts from Intelligent Tutoring Systems (ITSs) and Computer-Supported Collaborative Learning (CSCL) to develop an intelligent medical training system for PBL. In this paper we focus on the student modeling aspects of the problem.

Similar to one-to-one ITSs, e.g. ANDES [4], SQL-Tutor [15], our system requires an accurate model of clinical problem solving and a model of the student's state of knowledge so that the system can guide the students effectively. But in a PBL group clinical problem solving ability can vary from student to student since students differ in their background knowledge and skill. Thus modeling individuals and the group is

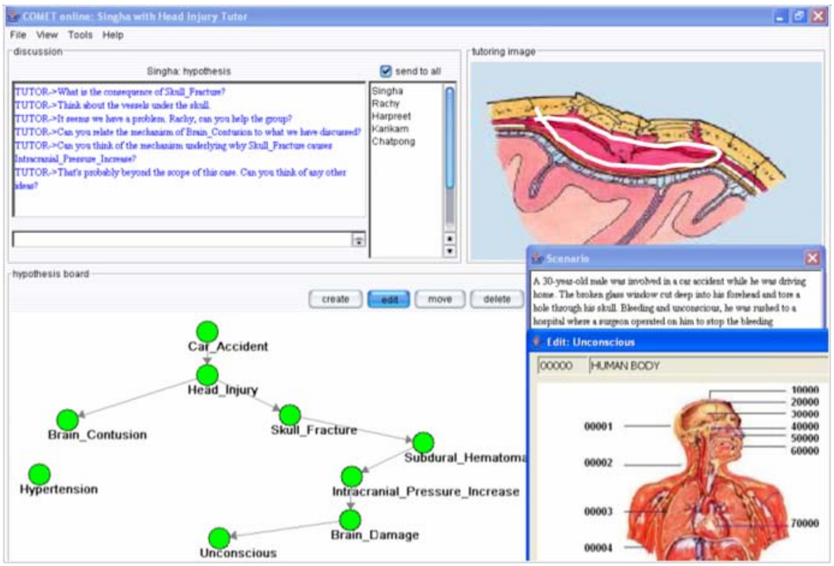
L. Ardissono, P. Brna, and A. Mitrovic (Eds.): UM 2005, LNAI 3538, pp. 377 – 386, 2005.

necessary if we wish to develop tutoring algorithms that can do things like focus the group discussion, promote collaboration, and suggest peer helpers.

Developing successful collaborative environments that satisfy each member's needs and contribute to the effectiveness of the group as a whole is an area that researchers have recently begun to address. The Docs 'n Drugs project [13] 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 [9] 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 [12] 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. Building upon results from Social Choice Theory, Masthoff [14] 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. 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. The modeling technique that we present in this paper has been implemented in COMET, a collaborative intelligent tutoring system for medical PBL [17]. Medical PBL is challenging due to the complexity of the knowledge involved, and the lack of standard, commonly accepted student problem-solving techniques. Thus, one objective of the work presented in this paper has been to identify prototypical patterns of student clinical reasoning to create student models that can be used by the tutoring module to generate the various tutoring hints.

2 COMET – COllaborative MEdical Tutor

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. COMET 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 (Fig. 1). COMET can currently support PBL in the domains of Head injury, Stroke and Heart attack. Generating appropriate tutorial actions in COMET requires a model of the students' clinical reasoning for 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 (BNs) as our modeling technique.



TUTOR->What is the consequence of Skull_Fracture?
TUTOR->Think about the vessels under the skull.
TUTOR->It seems we have a problem. Rachy, can you help the group?
TUTOR->Can you relate the mechanism of Brain_Contusion to what we have discussed?
TUTOR->Can you think of the mechanism underlying why Skull_Fracture causes Intracranial_Pressure_Increase?
TUTOR->That's probably beyond the scope of this case. Can you think of any other ideas?

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

Fig. 1. COMET Student interface. The hypothesis board provides the central shared group workspace. The discussion pane is the place for displaying tutoring hints

3 Clinical Reasoning Model

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

3.1 Domain Clinical Reasoning Model

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 [6], which is incorporated in the PBL process. It is characterized by the generation of multiple competing hypotheses from initial patient cues, followed by the collection of data to confirm or refute each hypothesis. Figure 2 shows a portion of the hypothesis structure created by one PBL group for the problem scenario on the bottom right of Figure 1. 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, BNs 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 3 steps [2]. (1) *Problem identification* is done by selecting problems from studying the case. This process is similar to "subgoaling" in means-ends problem solving [1]. We represent these problems with the "goal" node in the BN model (Fig. 3). (2) *Problem analyses* are developed for each problem. Students are encouraged to use their previous knowledge to solve the problem. We represent this knowledge with the “concept” node in the BN model. (3) The *hypotheses* are derived by applying medical concepts from the problem analyses. An “apply” node represents the student’s action of applying a “concept” to a “goal” to derive a “hypothesis”.

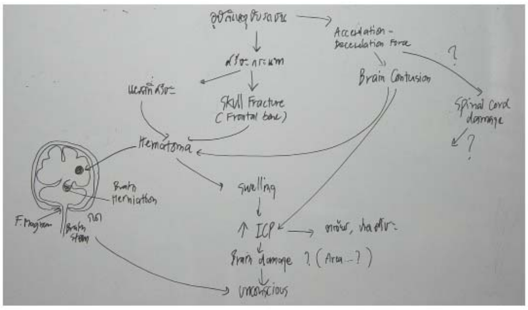


Fig. 2. A photograph of the white board after a PBL session at Thammasat University Medical School. The graph shows hypotheses with arrows indicating cause-effect relations among them. (Note: Some hypotheses are written in Thai)

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. 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 [7] 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 a set of consequences. Enabling conditions are illness features associated with the acquisition of illness (e.g., compromised host factors, hereditary factors). Faults are the major real malfunctions in illness (e.g., direct trauma, invasion of tissue by pathogenic organisms). Consequences are the secondary consequences of faults within the organism (e.g., unconsciousness, brain damage).

Figure 3 shows a portion of the BN domain model built for the head injury scenario of Figure 1. The model contains two types of information: (1) the hypothesis structure based on the differential diagnosis of the case (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. The figure shows the classification of the hypotheses into the three categories: 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 scenario took about one person-month to build. In Figure 3 (right half), we have seven possible faults associated with the single enabling condition car accident: *Head_Injury*, *Brain_Moving*,

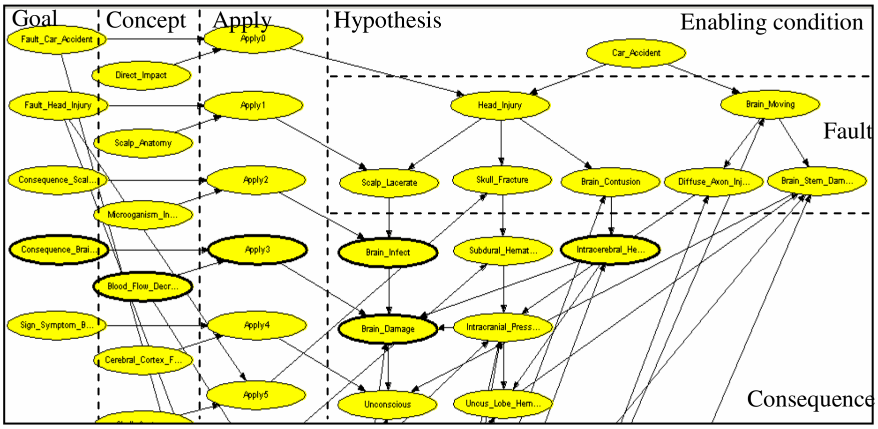


Fig. 3. Part of the Bayesian network student model. The complete network contains 66 nodes

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 causal 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 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 BN student model is similar to the student model used by Conati, et al [4]. Their model includes five types of nodes: Context-Rule, Rule-Application, Fact, Goal, and Strategy. The correspondence between their node types and ours is: Context-Rule = Concept, Rule-Application = Apply, Fact = Hypothesis, and Goal = Goal. Strategy nodes, which represent different correct solutions to a problem, are implicitly encoded in our model by the fact that students can enumerate the causal hypothesis structure in any order. Our model contains causal links among hypotheses, which are not present in their model. The reason for this is that in our medical domains a problem solution is represented by the hypotheses and causal links among them, while in their physics domains a problem solution is represented by a sequence of rule applications and the derived facts.

In the PBL sessions, the students create the hypotheses as well as the causal links between them (Fig. 2). The initial BN domain model described above represents the probabilities of the hypotheses but not the probabilities of the causal links between them. To capture the causal links as well, we modified the model by incorporating a new type of node representing the probability of a causal link between two hypotheses. In alternative model 1, for every hypothesis A that is a direct cause of a hypothesis B, we have a node representing the causal link between them. The two hypothesis nodes (A, B) are the parents of the link node (A→B), as shown in Figure 4b. The intuition is that the link cannot be created unless both hypotheses are created first. Alternative 2 is a combination of the initial student model and alternative 1 (Fig. 4c). Rather than eliminating the links in the Bayes net between hypothesis nodes, we retain them and simply add the link nodes. This model explicitly captures the probability that a student will create a hypothesis (B) if he creates the parent hypothesis (A), and the probability that the student will create the causal link between them. The probability tables for the link nodes in alternative 2 are the same as in alternative 1.

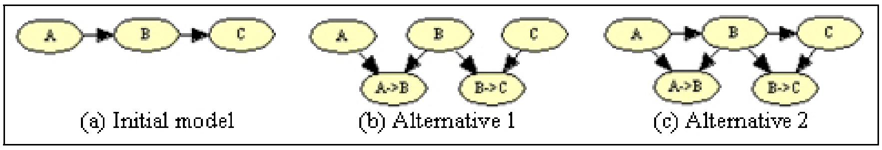


Fig. 4. A simple BN illustrating the hypothesis structure of the initial and alternative domain clinical reasoning models

The conditional probability tables for each network were obtained by learning from data obtained from the transcripts of PBL sessions. A total of 15 groups of third year medical students from Thammasat University Medical School 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 tape 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 provided by the HUGIN Researcher software to learn the conditional probabilities of each node [10].

3.2 Individual and Collaborative Student Clinical Reasoning Modeling

The domain clinical reasoning model is instantiated for each student prior to group discussion by entering that student's background knowledge as evidence. For example, if a student has a background in anatomy, we would instantiate the *Skull_Anatomy* and *Scalp_Anatomy* nodes. Since all students have *basic* knowledge in anatomy, physiology and pathology before they encounter the PBL tutorial sessions, we make the assumption that once a hypothesis in the domain model is created by one student in the group during discussion, every student knows that hypothesis. So as hypotheses are created, they are instantiated in each student model.

Conflict is an important aspect of group life. Researchers have suggested ways to ameliorate cognitive conflict and increase member productivity in group problem solving, which include combining individual novel ideas (divergent thinking) and narrowing this set to one alternative (convergent thinking) [11]. Following commonly accepted practice in medical PBL [7], 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, which is a function built into the Hugin software. Suppose we have the following hypotheses entered into the student model: *Car_Accident, Head_Injury, Intracranial_Pressure_Increase, and Unconscious*. The evidence is entered and propagated, and new beliefs are retrieved. Here we have six candidate paths, two of which are:

Path 2: ***Unconscious*** ← *Brain_Damage* ← ***Intracranial_Pressure_Increase*** ← *Subdural_Hematoma* ← *Skull_Fracture* ← ***Head_Injury*** ← ***Car_Accident***
Path 4: ***Unconscious*** ← *Brain_Damage* ← *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. Although groups can resolve disagreements in several ways, majorities are important [16], particularly in judgmental tasks that lack demonstrably correct answers (e.g. medical diagnosis). Thus, 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 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.

From our study of PBL sessions [17], we identified and implemented seven tutoring strategies commonly used by experienced human tutors: 1) focus group discussion, 2) promote open discussion, 3) deflect uneducated guessing, 4) avoid jumping critical steps, 5) address incomplete information, 6) refer to experts in the group, and 7) promote collaborative discussion. All strategies except strategy 7 use both the structure and the probabilities of the BN models. Strategies 1, 2, 5 make use of the group reasoning path.

4 Evaluation – Accuracy of the Student Models

In order to determine the accuracy of the model, we compared the probabilities of hypotheses and causal links from the student model with actual student actions considered as a "gold standard", and compared the group path generated by COMET and the path suggested by human tutors.

4.1 Experimental Design

We recruited 15 second-year medical students from Thammasat University Medical School. That is, they had not yet had PBL experience in Head injury, Stroke, or Heart attack. Stratified random sampling was applied to divide the students into 3 groups based on their background knowledge. Ten tutors with at least five years experience in conducting the brain and cardiovascular course were involved in the evaluation of the group path. Students were asked to answer pretest questions to determine their background knowledge. This information was used to instantiate the general student model for each individual student.

Students participated in the problem solving session on head injury, stroke and heart attack scenarios with COMET. Each student was asked to enumerate hypotheses and links using an offline client application. The student actions of creating hypotheses and their links served as a gold standard for comparing with the predicted probabilities from the BN student model. Then groups of 5 students worked collaboratively using an online client application. 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 the students' background knowledge. This data was used to compare with the group path generated by COMET.

4.2 Results

To determine whether our student models are accurate in predicting individual student actions, we evaluated them by means of receiver operating characteristic (ROC) curve analysis [3]. ROC curves plot sensitivity (true positive ratio) versus 1-specificity (true negative ratio) for a series of thresholds of the posterior probabilities of the nodes in the BN model. The area under the curve (AUC) represents an overall measurement of performance of the student model, with 1.0 a perfect test and 0.5 representing a model with no discriminating capacity. To measure the statistical significance of the difference between two AUCs, we used the between-area correlation and the standard error of the difference in areas [8].

Table 1. ROC curve analysis showing AUC for three student models

Model/Prediction	Head injury	Stoke	Heart attack	All scenarios
Initial/Hypotheses	0.731	0.809	0.843	0.814
Alternative 1/Hypotheses	0.859	0.793	0.917	0.848
Alternative 2/Hypotheses	0.909	0.765	0.868	0.832
Alternative 1/Causal links	0.895	0.814	0.843	0.848
Alternative 2/Causal links	0.897	0.838	0.905	0.899

Table 1 shows the ROC curve analysis of the three alternative models for the Head injury, Stroke, and Heart attack scenarios. For the Head injury scenario, there were no statistically significant differences between the AUCs for alternative 1 and alternative 2, while each of them was more accurate than the initial model in predicting which hypotheses students created. For the Stroke and Heart attack scenarios, there were no statistically significant differences between the AUCs of all three models. Averaging over all scenarios, alternative 1 and alternative 2 were more accurate in predicting which hypotheses students created than the initial model. However, there was no statistically significant difference between the AUCs for the alternative 1 and alternative 2. Alternative 2 was more accurate in predicting which causal links student created than alternative 1.

In order to evaluate the accuracy of our BN student model in predicting the group reasoning path, we compared the group reasoning path generated by COMET to the paths suggested by 10 human tutors for 3 scenarios and 3 groups. This gave us 90 data points for comparison. Total number of reasoning paths containing at least one node that was created for the Head injury, Stroke and Heart attack scenarios was 6, 65, and 125 respectively.

Table 2. 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%)

The results show that COMET's group paths are in line with the majority consensus of those suggested by the human tutors. For example, in the situation where COMET generated path no. 12, 85% of the tutor also suggested the same path, and 15% suggested path no. 14. 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 [5]. There were no statistical differences between the human tutors and COMET (McNemar test, $p = 0.774$). The results show a high degree of agreement between the group path generated by COMET and by the human tutors (Kappa index = 0.823).

5 Conclusions and Future Work

We have described a general domain-independent BN clinical reasoning model for medical PBL that integrates hypothesis structure based on differential diagnoses of the patient case and the application of the corresponding medical concepts in the problem solving process. Student background knowledge as well as individual and group reasoning behavior play an important role in modeling individual and collaborative student clinical reasoning. The positive result from the model's evaluation in three different scenarios provides encouraging support for our framework.

Future work will include more extensive evaluation. We are planning a full scale evaluation of COMET's effectiveness in imparting clinical reasoning skills and medical knowledge to students. Specifically, our empirical study will focus on student clinical reasoning gains obtained using COMET versus those obtained from human tutored PBL sessions.

Acknowledgements

We thank Hugin Expert for providing us the use of the Hugin Reseacher software. 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.

References

1. Anderson, J. R.: Problem solving and learning. American Psychologist. 48 (1993) 35-44
2. 2. Barrows, H. S.: A taxonomy of problem-based learning methods. Medical Education. 20 (1986) 481-486
3. Bradley, A. P.: The Use of Area under ROC Curve in the Evaluation of Learning Algorithms. Pattern Recognition. 30 (1995) 1145-1159
4. Conati, C., Gertner, A., VanLehn, K.: Using Bayesian Networks to Manage Uncertainty in Student Modeling, Journal of User Modeling and User-Adapted Interaction. 12 (2002) 371-417
5. Dawson, B., Trapp, R. G.: Basic & Clinical Biostatistics (McGraw-Hill, 2001)
6. Elstein, A. L., Shulman, L. S., Sprafka, S. A.: Medical Problem Solving - An Analysis of Clinical Reasoning (Cambridge: Harvard University Press, 1978)
7. Feltovich, P. J., 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) 128-142
8. Hanley, J. A., McNeil, B. J.: A Method of Comparing the Areas Under Receiver Operating Characteristic Curves Derived from the Same Case. Radiology. 148 (1983) 839-843
9. 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. (2003) 1-14
10. Lauritzen, S. L.: The EM-algorithm for graphical association models with missing data, Computational Statistics and Data Analysis. 1 (1995) 191-201
11. Levine J. M., Moreland, R. L.: Collaboration: The social context of theory development. Personality and Social Psychology Review. 8 (2004) 164-172
12. Lock, Z.,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. (2003) 25-34
13. 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. Proceedings of the AMIA Fall Symposium. (2001) 433-437
14. Masthoff, J.: Modeling a group of television viewers. Proceedings of the Workshop Future tv, in Intelligent Tutoring Systems Conference. (2002) 34-42
15. Mitrovic, A.: Experiences in Implementing Constraint-Based Modeling in SQL-Tutor, Proceedings of the 4th International Conference in Intelligent Tutoring Systems. (1998) 414-423
16. Stasser G., Kerr N. L., Davis J. H. : Influence processes and consensus models in decision-making groups, in P. B. Paulus, ed., Psychology of group influence (Hillsdale, Erlbaum, 1989) 279-326
17. Suebnukarn, S., Haddawy, P.: A collaborative intelligent tutoring system for medical problem-based learning. Proceedings of the 9th International Conference on Intelligent User Interfaces. (2004) 14-21