Tan, S. C. et al. (Eds.) (2013). Workshop Proceedings of the 21st International Conference on Computers in Education. Indonesia: Asia-Pacific Society for Computers in Education

# Adaptive Question Generation for Student Modeling in Probabilistic Domains

**Nabila KHODEIR*** **& Nayer WANAS**
*Informatics Dep., Electronics Research Institute, Egypt*
*nabila_khodeir@yahoo.com

**Abstract:** Problem solving behavior remains to be the most trustable source for modeling student knowledge in intelligent tutoring systems. In this work we focus on diagnostic problem solving, as an essential question type associated with probabilistic domains. Student answer for such questions indicates the knowledge discrepancies between the student and his/her stored model. In this paper we introduce an algorithm that adaptively generates different appropriate follow-up questions to accurately determine the knowledge discrepancies in the student model. Answers to these follow-up questions are used to update the student model. Verification is conducted on the updated model based on the matching between student and generated model answers to the presented questions. Results suggest that tracking the student knowledge discrepancies using the generated follow-up questions improves the prediction accuracy of the student answers by 20% compared to relying only on the diagnostic questions alone. In addition, approximation of the student model enhanced by 40% relative to that obtained using the diagnostic questions alone.

**Keywords:** Intelligent Tutoring System, Student Modeling, Abduction, question generation

## 1. Introduction

Intelligent Tutoring Systems (ITSs) (Brusilovsky, 2003), are computer-assisted tutoring systems that build a model for the objectives, preferences and knowledge of an individual student in order to adapt the system to his/her learning needs. Based on the student model ITSs are able to support the student with a lot of interactions and personalization. Modeling of the student knowledge drew heightened attention in the literature (Carmona et al., 2005), (Pahl and Kenny, 2009). Several models have been suggested for the student knowledge modeling namely the i) Overlay model, ii) Perturbation model and iii) Differential model. The overlay model represents the user model as a subset of the expert model (Carmona et al., 2005), (Melis and Siekmann, 2004). Therefore, its modeling process is basically based on the representation of the expert or the domain model. Several representations have been adopted for the overlay model including production rules (Corbett et al., 1993), and Bayesian Networks (BN) (Conati et al., 2002). The perturbation model extends the overlay model by adding the representation of the student incorrect knowledge (misconceptions or bugs) (Yacef, 2005). The differential model, on the other hand, represents both the student knowledge and the differences between student and expert knowledge (the knowledge the user lacks) (Burton and Brown, 1976). We invoked the idea of a differential model to utilize the differences between the student model answer and the student answer to update the student model. The student model is initialized based on some assumptions or prior information, and in some cases could be initialized as the expert model, In turn the answer generated from the model prior to matching the student perfectly can be different from the student answer to the presented question. Such differences express the discrepancies between the student knowledge and the student model. The contribution of this paper is an algorithm that utilizes such discrepancies to automatically generate a series of follow-up questions, and in turn the student answers to these questions are used to update the student model.

Most adaptive tutoring systems that model the student knowledge deal with domains represented by deterministic models that define the domain by a set of variables and describe the relations between them by fixed rules. However, in the real world, especially in applications such as forecasting, troubleshooting and medical diagnosing, a degree of uncertainty is inherent which requires the use of probabilistic models to represent such domains. Domains of this nature are probability based inference in complex networks of interdependent variables. Bayesian Networks are widely used approach to represent such domains to handle the uncertainty of their relations. (Suebnukam and Haddawy, 2005) presented an example of modeling the student knowledge in probabilistic domains. They suggested a modeling algorithm that focuses on the skill of reasoning through domain variables relations around practical patient problems in medical domains. This work suggests a modeling algorithm that utilizes diagnostic skill to infer the student knowledge in probabilistic domains through different questions that are automatically generated.

Automatic generation of questions supports the functionality of ITSs, in addition to dialogue systems (Piwek, Stoyanchev, 2010), and Question Answering (QA) systems (Kalady et al., 2010). Most question generation techniques revolve around linguistic study including syntactic and semantic analysis for the given document to generate questions (Heilman and Smith, 2009), (Becker et al., 2010). In turn, factual and definitional questions are the common types of generated questions in these approaches (Heilman and Smith, 2010). However, queries associated with some domains cannot be generated or answered based on linguistic analysis. For example, Probabilistic domain represents a difficult problem in this regard. In this paper we proposed an approach to generate different questions types and their answers automatically by utilizing the Bayesian Network (BN) knowledge representation ( Korb and Nicholson, 2011) for probabilistic domains.

The rest of this paper is organized as follows. Section two presents the proposed different questions types and their generation process. The proposed updating technique is illustrated in section three. Thereafter, we explore the experimental results that illustrate the performance evaluation of the algorithm implementation in section four, and discussion of the result and conclusion is given in section five.

## 2. Proposed Questions Types and Their Generation Techniques in Probabilistic Domains

We identify three questions types that vary in the level of thinking required to be answered according to Bloom's Taxonomy (Bloom, 1956). Diagnostic and comparison questions, which belong to the higher thinking level, and feature specifications question, that belongs to the lower thinking level. Next section illustrates the different questions types in more details.

### *2.1 Proposed Questions Types*

Probabilistic domains are usually associated with diagnostic questions which require identifying the most probable explanation given a set of evidences. We consider such questions as the essential questions especially in relation to ambiguous cases, where more than one hypothesis that explains the question evidences exists. In such cases the student is asked to provide a ranked list of possible hypotheses for the question evidences. Diagnostic questions for ambiguous cases are chosen since answers for such questions reveal more information about the student knowledge. Answer for such question need recalling of information, information analysis, and judgment skill to select and arrange the most possible hypotheses.

Comparison questions are used as follow-up questions to track the student beliefs about the relations strengths between specific evidence and different hypotheses. On the other hand, feature specifications questions are used to pursue what the student think about the relations existence between specific hypothesis and different evidences.

### *2.2 Questions Generation Contexts and Techniques*

The modeling process relies on discrepancies between the student answer and answer generated from his/her model to diagnostic question. Each diagnostic question deals with a specific sub-graph in the domain that is represented by a BN. The diagnostic question generation process is preceded by choosing some evidences that are have common relations with a set of hypotheses. These evidences constitute the generated question which is validated to have more than one hypothesis to achieve the condition of ambiguity case. Thereafter, the question is presented to the student to get his/her answer in addition to generating the question answer by applying abduction inference mechanism on the BN that represent the student model (Nilsson, 1998). An example of diagnostic question is presented in Figure 1.

***Diagnostic Question***
If you have a case with maculopapular rash, abdominal pain, and malaise. What are the most probable diseases? Choose and Rank from the following diseases beginning by 1 to the highest likely diagnosis?

| | |
|---|---|
| ☐ Rubella | ☐ Roseola infantum |
| ☐ Infectious mononucleosis | ☐ Scarlet fever |
| ☐ Measles | ☐ Chicken pox |

Figure 1. Example of diagnostic question

The answer provided by the student to the diagnostic question is compared to that generated by the student model using the abduction algorithm. If the answers match, we declare that the student model doesn't require regulation. On the other hand, if there is a discrepancy between the model and student answers the student model needs to be regulated. Then, generation of follow-up questions begins by analyzing the difference(s) between the two answers. Since the answer is a ranked list of hypotheses, the difference between the two answers can be one of the following:

1. One or more missing hypothesis.
2. One or more extra hypothesis.
3. Answers are presented in a different order.
4. Combined error of missing and extra hypotheses in a correct or incorrect order.

The follow-up questions generation process proceeds in two phases. First, check the missing and extra hypothesis and generates feature specifications question that ask about the existence and absent of relations between evidences and the missed or/and extra hypotheses. According to the student answer the student model is regulated. Fig. 2 gives example of the generated scaffolding questions according to the differences between the two answers.

***Main Question***
If you have a case with maculopapular rash, and high fever, what are the most probable diseases?

***Student Answer:***
1. Roseola infantum
2. Measles
3. Infectious mononucleosis

***Student Model Answer:***
1. Roseola infantum
2. Scarlet fever
3. Measles

***First follow up question***
Infectious monoucleosis is associated with
1. Maculopapular rash
2. High fever
3. None of the above

***Second follow up question***
Scarlet fever is unassociated with
1. Maculopapular rash
2. High fever
3. None of the above

Figure 2. Answers Differences and corresponding Generated Scaffolding Questions

After the student model is regulated the matching between the student answer and generated student model answer is checked. The second phase of the follow-up questions is initiated in the case of a mismatch between the two answers with regard to the order of the correct hypotheses. The generated follow-up questions in this phase are comparative questions which needs higher thinking skills compared to the first phase, where the student is asked to compare between two hypotheses in regard of question evidences. Based on the student answer the student model is updated. Fig. 3 gives an example of the generated comparative questions in the second phase.

## 3. Updating Techniques

The student model is initialized by the expert model that represents some pediatric diseases from the medical domain in the form of a BN. According to the student answer, the student model is updated to give the same response as the student. Three different approaches for updating process was suggested i) coarse, ii) refined, and iii) blended (Khodeir et al., 2012). The coarse updating technique is based on sharp actions in updating of the student model BN. On the other hand, the refined updating is based on repetitive regulation of the student model BN gradually until matching between student answer and model answer occurs. Regulation will be terminated if the difference between the answers increases or the weights reach the threshold values which are zero for decreasing weight regulation and one for increasing weight regulation. Blended updating technique is adaptable way for student model regulation. According to the difference between the student answer and the model answer the updating model is selected. Missing or adding of hypotheses triggers coarse updating technique while wrong order difference initiates the refined updating technique.

***Main Question***
If you have a case with maculopapular rash, and high fever What are the most probable diseases?

***Student Answer:***
1. Roseola infantum
2. Measles
3. Infectious monoucleosis

***Student Model Answer:***
1. Roseola infantum
2. Infectious monoucleosis
3. Measles

***First follow up question***
Measles is less associated than Roseola infantum with
1. Maculopapular rash
2. High fever
3. Non of the above

***Second follow up question***
Measles is more associated than Infectious monoucleosis with
1. Maculopapular rash
2. High fever
3. Non of the above

***Third follow up question***
Infectious monoucleosis is less associated than Measles or Roseola infantum with
1. Maculopapular rash
2. High fever
3. Non of the above

Figure 3. Answers Differences and corresponding Generated Scaffolding Questions

### *3.1 Coarse Updating Technique*

The coarse update is conducted by adding or removing of relations for the differences in the hypotheses in the following manner

#### *3.1.1 Missing hypothesis or hypotheses*

The student model is regulated by establishing missing relations or modifying the weight of the existing relations $P(H_i|E_j)$ between evidence $E_j$ and the missed hypothesis $H_i$ for each question. Different selected weights are allocated to each established relation. The weights of the newly added links are equal to the weight of the matching or nearest matching between the rank of the missing hypothesis $P(H_k|E_j)$ . In case of absence of other relations, the weight is assigned to be equal 0.5. This is expressed by Equation 1.

$$P(H_i|E_j) = \begin{cases} P(H_i|E_j) & if \quad P(H_k|E_j) \neq 0 \;\; and \;\; (H_i)_{rank} \geq (H_k)_{rank} \\ 0.5 & if \quad P(H_k|E_j) = 0 \end{cases} \quad (1)$$

### 3.1.2 *Extra hypothesis or hypotheses*

The student model is regulated by removing the existing relations $P(H_i|E_j)$ between the hypotheses added and the question evidence. This is expressed by Equation 2.

$$P(H_i|E_j) = 0 \qquad if \quad P(H_k|E_j) \neq 0 \quad (2)$$

## *3.2 Refined Updating Technique*

The refined update is performed using successive increase or decrease in the weights and prevalence using a fixed step. The step value indicates the speed of settling of the algorithm. The linear refined modification is shown in Equation 3.

$$W_{i+1} = W_i \pm \Delta W_i \quad (3)$$

Where $W_{i+1}$ is the new weight, $W_i$ is the old weight, and $\Delta W_i$ is a constant value which is increased gradually by step equal to 0.1 with considering of the weight value is within the range [0,1].

In turn, the refined update increases the relations' weights between the hypothesis with the lowest rank and the evidence. In addition, the weights of the relations between the hypothesis with the highest rank and the evidence are decreased. If incremental updates fail to achieve a match through available iterations (until the updating weights reach a threshold value) the student model is updated by increasing the prevalence of the hypothesis with the lowest rank and decreasing the prevalence of the hypothesis with the highest rank.

## *3.3 Blended Updating Technique*

The blended updating technique is a combination of coarse updating and refined updating techniques. Coarse updating is used when the student answer is heavily skewed from the student model answer in presence or absence of the hypotheses themselves. This is due to the fact that the student answer in this case is highly diverse from the model answer and needs significant modification. On the other hand, refined updating is used when the student answer differs from the student model answer in the order of hypotheses. This stems from the fact that the student answer in this case is in close proximity to the student model answer and needs limited modification. Figure 4 illustrates an example of the student model blended updating where the updating step in refined technique is equal to 0.2.

It is worth mentioning that updating techniques are applied on two cases. The first case relies on the student answers to diagnostic questions alone. In this case, blended updating technique is applied. Coarse updating is applied on all relations between the missing and extra hypotheses in the student model answer and the mentioned evidences in the question in addition to applying of the refined technique on erroneously ordered hypotheses. Refined updating techniques are used to update all relations between the wrong order hypothesis and the question evidences. On the other hand, the second case is based on the student answers to follow-up questions. This case is characterized by selectively updating relations between hypotheses and the question evidences according to the student answers to the presented questions. Coarse updating is used to update the student model according to feature specifications questions answers while refined updating is utilized for compare questions answers. Measuring the impact of the two cases on the modeling process is the target of the evaluation section.

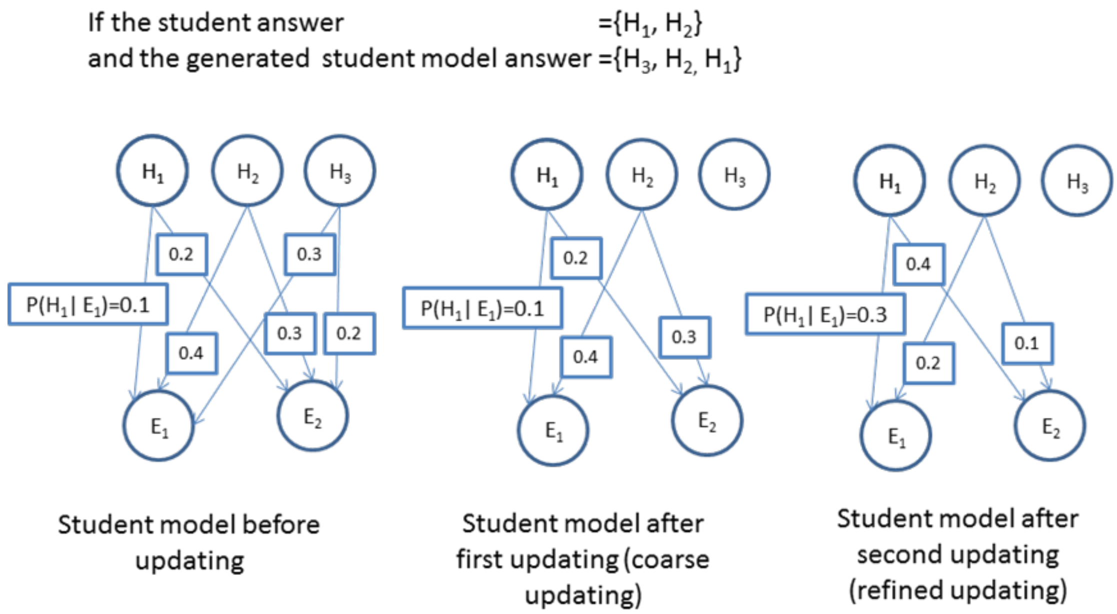


Figure 4. Illustration of Modification in case of Combined Error in the Generated Student Model Answer Using Blended Updating Technique

## 4. Evaluation

The accuracy of the student model using diagnostic questions and follow-up questions were evaluated from two perspectives: its ability to predict the outcome of an individual student answer implying its misconceptions and the approximation of the student model compared to the actual model. Measuring the approximation requires to run the experiment on students whose prior knowledge could be accurately assessed. This allows comparison between the resulted student knowledge model and the actual student knowledge. Therefore we use a simulated student approach in the evaluation process. (Van Lehn and Jones, 1998), and (Millan, et al., 2002) have suggested simulated student approaches for evaluating student models. Simulated students enable the measurement of the difference between the simulated student and the updated student knowledge model obtained quantitatively.

The proposed mechanism to generate the simulated students is based on an existing domain BN. Simulated students' models are randomly modified BNs that represent the students' knowledge. The simulated student response is assessed by processing the generated BN to generate the target student answer on the posted question. Simulated students BNs that represent the students' knowledge are automatically generated by perturbation of the knowledge model. The perturbation proceeds on two levels; 1) the links level, where some links are removed and some are added, and in 2) the weight level where some links weights values are changed. The perturbation process is constrained by specific ratios to prevent generation of an extremely perturbed BN that might be unrealistic. In this work, the medical domain is used as an example of probabilistic domains. The domain knowledge selected for evaluation is based on information from pediatric experts. Six diseases are selected based on their overlapping symptoms to allow generation of diagnostic questions. The relations between diseases and symptoms have causal relations with probabilities in addition to the prevalence of each disease are represented as a Bayesian Network. This representation is exploited to update the student model. Moreover it is utilized to verify the updating of the student model.

We aim to measure the impact of using generated follow-up questions on the efficiency of the student knowledge modeling process. We test the performance of the modeling algorithm using diagnostic questions, in addition to the effect of using follow-up questions. The blended updating technique is used in the both cases. The evaluation begins by generating three groups of twenty random different diagnostic questions. The diagnostic questions are used for the updating algorithm. Then, a different set of twenty questions are tested against the new updated student model for measuring the student answer prediction accuracy using the diagnostic questions $SPA_{DQ}$.

The same diagnostic questions are used to evaluate the updating algorithm by using follow-up questions. Mismatching between the student answer and generated student model answer initiate the generation of follow-up questions according the differences between the two answers. The follow-up questions are used for the updating algorithm. Another set of twenty questions are tested against the new adjusted model for measuring the student answer prediction accuracy using follow-up questions $SPA_{FQ}$. Prediction accuracy expresses the comparison between the student answer and the generated answer using the final updated student model. It is worth mentioning that, applying the refined technique proceeds by steps. Different updating steps (0.1,0.2 to 1) were used in measuring the prediction accuracy.

Figure 5 indicates enhancement of the performance of the algorithm using follow-up questions over using diagnostic questions by up to 20%.

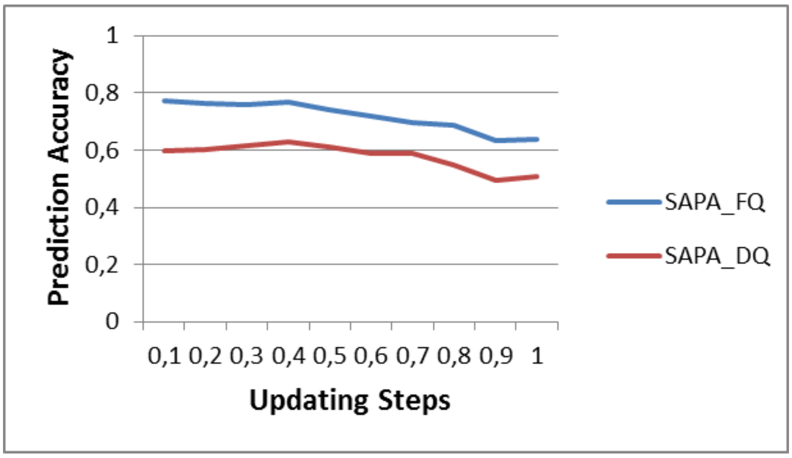


Figure 5. Prediction accuracy using follow-up questions $SPA_{FQ}$ and prediction accuracy using the diagnostic questions $SPA_{DQ}$

It is worth mentioning that, the performance degrades for all updating techniques with the increasing of step size. This stems from the fact that, for larger steps the algorithm follows coarse actions instead of refined actions which lead to degradation of the performance.

To evaluate the approximation of the student model in the two cases (using diagnostic questions only and using diagnostic and follow-up questions) relative to the actual student model, we used the Root Mean Squared Error (RMSE) between the different models. RMSE is applied on the data that represents the differences in weights between the BN that represent the student model and the BN that represents the simulated student. The results show how using follow-up questions enhances the accuracy of the student model up to 40%. As shown in the following table the RMSE is significantly decreased in case of follow-up questions.

Table 1. RMSE for the differences between the student model and the actual model through different updating steps

| Updating Steps | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|
| Using diagnostic questions | 9.38 | 9.34 | 9.24 | 9.15 | 9.30 | 9.11 | 9.46 | 9.31 | 9.93 | 9.98 |
| Using follow-up questions | 5.78 | 6.12 | 6.00 | 6.04 | 6.50 | 6.55 | 6.98 | 7.26 | 8.12 | 8.20 |

## 5. Discussion and Conclusion

In this paper, we presented an algorithm to approximate the student model in a probabilistic domain. The algorithm utilizes generated follow-up questions that track the discrepancies between the student knowledge and his/her knowledge model. The algorithm is invoked when the student answer mismatches the expected answer evaluated using the student model. The blended updating approach is applied in the modeling process. In addition, different granularity levels are evaluated by changing the value of the updating step and the output of this parametric study is indicated. An experimental evaluation of the approaches has been conducted. Random models for the student are generated. A series of twenty questions are automatically generated and presented to the system based on the domain structure. The performance of the algorithm is evaluated using both the prediction accuracy of the student answers to the questions and the number of required trials to this estimation. The results suggested that using follow-up questions gives better performance with respect to accuracy compared to using diagnostic questions alone especially in small updating step by 20%. In addition, approximation of the student model enhanced by 40% relative to that is obtained using the diagnostic questions.

The algorithm aims to obtain a more accurate student knowledge model that contains the correct and incorrect knowledge represented in BN form. Then, the model can be used to control tutoring system, such as Intelligent Tutoring System, interactions with the student to rectify his/her errors. The proposed follow-up questions can be also used in the context of learning by utilizing the discrepancies between the student knowledge model and domain model.

## References


Becker, L. Nielsen, R., Okoye, I., Sumner, T., & Ward, W. (2010). What’s next? Target Concept Identification and equencing. *Third Workshop on Question Generation*, 35–44. *The Tenth International Conference on Intelligent Tutoring Systems (ITS2010)*, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.

Bloom, B. (1956) Taxonomy of Educational Objectives. David McKay Co. Inc.

Brusilovskiy, P. (2003). Adaptive and intelligent web-based educational systems. *Artificial Intelligence in Education*, 13, 159—169.

Burton, R., and Brown, J. (1976). A tutoring and student modeling paradigm for gaming environments. In: *SIGCSE-SIGCUE joint symposium on Computer science education*, (pp. 236-246). ACM.

Carmona, C., Milán, E., and Pérez de-la Cruz, J. (2005). Introducing prerequisite relations in a multi-layered Bayesian student model. In: L. Ardissono, P. Brna, & A. Mitrovic (Eds), *Proceedings of UM 2005* (pp. 347–-356). LNAI, Springer-Verlag Publishers.

Conati, C., Gertner, A., & Vanlehn. K. (2002). Using Bayesian networks to manage uncertainty in student modeling. *User Modeling and User-Adapted Interaction*, 12, 371-417.

Corbett A., Anderson, J., & OBrien, A. (1993). The predictive validity of student modeling in the act programming tutor. In: Artificially AI-ED 93,( pp. 457-464).

E. Millan, J. Perez-De-La-Cruz, & Luis. (2002). A bayesian diagnostic algorithm for student modeling and its evaluation. *User Modeling and User-Adapted Interaction*, 12, 281-330

Heilman, M., Smith, N. (2009) Question generation via overgenerating transformations and ranking. *Technical Report CMU-LTI-09-013*, Carnegie Mellon University.

Heilman, M. & Smith, N. (2010) Extracting Simplified Statements for Factual Question Generation. *Third Workshop on Question Generation*, 35–44. *Tenth International Conference on Intelligent Tutoring Systems ITS2010*, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.

Kalady, S., Elikkottil1, A., Das, R. (2010). Natural Language Question Generation Using Syntax and Keywords. *Third Workshop on Question Generation*, 1-10. *Tenth International Conference on Intelligent Tutoring Systems (ITS2010)*, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.

Khodeir, N., Wanas, N., Darwish, N., & Hegazy, N.: Bayesian Based Student Knowledge Modeling in Intelligent Tutoring Systems. (2012). *38th Annual Conference of the IEEE Industrial Electronics Society*, 12-17, Montreal, Canada.

Melis, E., & Siekmann, J (2004). Activemath: An intelligent tutoring system for mathematics. In: Rutkowski, L., Siekmann, J., Tadeusiewicz, R., & Zadeh, L. (Eds), *Seventh International Conference Artificial Intelligence and Soft Computing (ICAISC)*. LNAI, vol. 3070, (pp. 91-101). Springer-Verlag.

Nilsson, D. (1998). An efficient algorithm for finding the m most probable configurations in probabilistic expert systems. *Statistics and Computing*, 8, 159-173.

Pahl, C., & Kenny, C. (2009). Interactive correction and recommendation for computer language learning and training. *IEEE Transaction On Knowledge And Data Engineering*, 21, 854-866.

Piwek, P., Stoyanchev, S. (2010). Question Generation in the CODA project. *Third Workshop on Question Generation*, 29–34. *The Tenth International Conference on Intelligent Tutoring Systems (ITS2010)*, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.

Suebnukarn, S., & Haddawy, P. (2005). Modelling individual and collaborative problem solving in medical problem-based learning. *User Modeling 2005*, 377-386. Phonix, Arizona, USA.

VanLehn, K., & Jones, R. (1998) Student modeling from conventional test data: A bayesian approach without priors. *Proceedings of 4th International Conference ITS98*, (pp. 434-443). Berlin Heidelberg: Springer Verlag,