Student evaluation model using bayesian network in an intelligent E-learning system

Article *in* IIOAB Journal · January 2016

Meghamala Sinha
Oregon State University

ISSN: 0976-3104
SPECIAL ISSUE
Chakraborty and Sinha

THE IIOAB JOURNAL

ARTICLE OPEN ACCESS

STUDENT EVALUATION MODEL USING BAYESIAN NETWORK IN AN INTELLIGENT E-LEARNING SYSTEM

Baisakhi Chakraborty and Meghamala Sinha*

Department of Information Technology, National Institute of Technology, Durgapur, INDIA

ABSTRACT

An Intelligent Tutoring System (ITS) is a type of knowledge based system whose main agenda is to efficiently substitute a human tutor by a machine. Unlike the traditional classroom teaching ways, an ITS has the ability to fit according to the necessity of each student. More and more emphasis is laid on different types of e-learning systems now days. In this paper a probability based ITS system is proposed consisting of four models namely student model, tutor model, domain model and a student evaluation model. The emphasis has been given on the student evaluation model where an element of uncertainty has been introduced and handled by Bayesian network. The purpose of the student evaluation model is to correctly detect the knowledge level of each student based on their response to questions. The uncertainty factor has been defined by terms guess and slip parameters. The two parameters are defined as follows: (a) Guess is the probability that a student of low intelligence gives a correct response to a difficult question whereas (b) Slip is the probability that a student of high intelligence gives an incorrect response to an easy question. During evaluation of the knowledge level of a student, we have incorporated the uncertainty factors of guess and slip with the help of Bayes' rule and have found desirable results that take into account the possibility of slippage or guess.

Received on: 1st -Dec-2015
Revised on: 6th-Dec-2015
Accepted on: 20th-Dec-2015
Published on: 10th –Jan- 2016

KEY WORDS

E-learning; Bayesian Network; intelligent tutoring system (ITS); personalized learning; student modeling

***Corresponding author:** **Email:** meghamala.sinha@gmail.com

INTRODUCTION

An Intelligent Tutoring System (ITS) is a special type of knowledge based system that replaces a human tutor by a machine which provides personalized tailored instructions and feedbacks to the user [1]. The major difference between the ITS and the traditional classroom is that ITS can fit according to the necessity of each students. It is impossible for a human tutor to cater to the needs for every student in a classroom. Other advantage of the intelligent tutoring systems is removal of time and space complexity of the real world unlike a regular classroom. It aims to provide a reformed education system.

Framework of an ITS system

A typical intelligent tutoring system consists of four components:

1. The Domain Model
2. The Student Model
3. The Tutoring Model
4. The User Interface Model

The Domain Model or Expert Model contains the detailed description of what the expert user's knowledge consists of. This model contains a superset of all concepts, strategies, rules etc. That is, given a particular problem, the domain model contains all possible steps for its solution [2]. It acts as a reference to evaluate a student's knowledge level as he/she solves a problem.

The Student Model contains the description of the knowledge level of the student along with their misconception and knowledge gaps. It can be represented as an overlay on the domain model. This means that as a student solves

ISSN: 0976-3104

a problem, his/her activity is traced according to the domain model in order to correctly identify the presence of the required knowledge in the student. Student Modeling is the most crucial task of the ITS.

The Tutor Model acts as a support to the students to help with their learning process. It takes input from both student model and domain model in order to provide recommendations and instructions to the student [3].
The User Interface Model acts as an interface between the ITS and the student logged in.

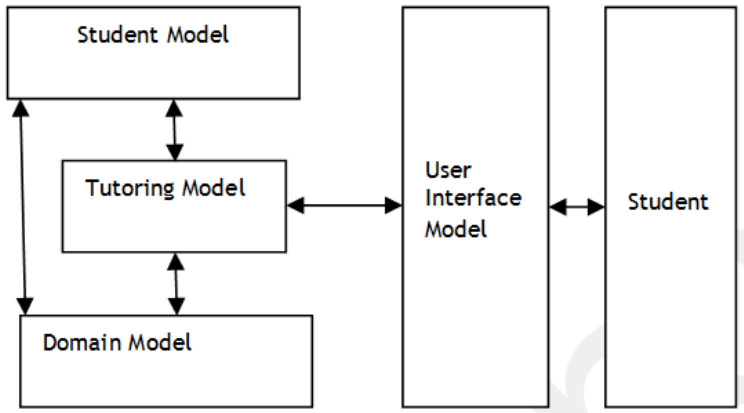

Fig: 1. Typical ITS framework

Challenges in Existing ITS

The major challenge faced by e-learning systems, whose goal is to determine a students' knowledge about a particular domain, is that how much precisely and correctly it evaluates a student based on his/her responses. Even though a student provides a correct (or incorrect) response, it cannot be correctly concluded by the system what the true state of knowledge of the particular student really is. If a student whose past academic history is excellent and continuous assessment is good, but there are few mistakes is his response not expected from him, there is a high possibility that the student did a silly mistake in his/her answer. On the contrary, if a student with low continuous performance suddenly receives excellent grades, there is a possibility of guess which is a combination of random answering and adopting unfair means. We have taken into account such possibilities and have added a student evaluation model component that inputs the student current knowledge level and their response into a Bayesian Network to evaluate probabilities of guess or slippage. Then these probabilities are used to calculate the updated knowledge level of the student.

Our Approach

Our ITS model allows a student logged in the system gain knowledge on concept/domains and particular subjects with the help of an e-tutor. The level of knowledge of a student at an instant is measured with the help of the student evaluation model. This model makes it possible for students with a pre-requisite knowledge level to access questions and answers them. Based on the responses to the question, this module provides a grade to the student. If the student achieves a grade above a particular threshold, then he/she is entitled to questions of higher knowledge level. Likewise his/her knowledge level is enhanced as he/she goes on solving the question in this ITS model. Based on his/her knowledge level, the student evaluation model also suggests study material which are accessible via the tutoring module. The underlying technology used is Bayesian Networks [4].

Bayesian Network is a compact and theoretically sound probabilistic graphical structures which are used as a tool for building a model to represent probability distribution over a given problem domain. It is a mathematically solid and efficient mechanism to provide insights on imprecise and uncertain information. In a Bayesian Network, any observable state or feature is represented by a node and any interdependencies among the features are represented by directed arcs. Each node is associated with a conditional probability table (CPT) representing probability of their occurrence conditions by their parent nodes. We have designed a Bayesian Network shown in **Figure– 4** to evaluate the probability of slip or guess for each student answering questions.

COMPUTER SCIENCE

ISSN: 0976-3104

RELATED WORK

Array As mentioned earlier e-learning systems have lots of advantages over a traditional tutor in a classroom [5]. Firstly, they are easily available over the internet hence overcoming time complexity as well as geographical complexity. Secondly E-learning systems are lot more attractive due to incorporation of rich multimedia and interactive features, such as, video, audio, animation etc. But there is a major scope of improvement in these static learning systems. It is the necessity of introduction of personalization by making the study materials more adaptive and interactive according to the requirement of the students [6].

Although there are several practical applications of Artificial Intelligence in the development of such systems, the best solution to this problem is a "one-size-fits-all" approach proposed by Brusilovsky and Maybury [7]. Such systems come with the advantage to adapt according to the users current knowledge level and skills. Using this idea Intelligent tutoring systems were build in order to assist students while solving problem [8]. It was also necessary to address the problem of handling uncertain or incomplete information while evaluation of users knowledge. Bayesian Network is one of the strongest probabilistic graphical models which can handle uncertain or imprecise data [9, 10, 11]. It has been applied in various applications like health, e-commerce, tutoring systems etc.
We will briefly review some of the approaches used in the past to develop student models using probabilistic methods [12].

OLAE (Martin and VanLehn.1995) which stands for "On-Line Assessment of Expertise or Off-Line Assessment of Expertise" is an assessment system that collects information about 'what a student knows' during problem-solving in introductory college physics [13]. This information is used by the teacher to make decisions. For each given problem, OLAE builds a Bayesian network that relates knowledge to actions. For example, what rules should be known in order to solve a particular equation. By using this network, the system knows whether a student knows and appropriately uses each of the rules.

POLA (Conati and VanLehn, 1996) stands for Probabilistic On-Line Assessment [14]. It provides a framework for modeling and assessment of student knowledge while they solve problem in introductory physics class. It differs from OLAE by applying probability reasoning to execute both knowledge and model tracing. In order to provide a compact representation of all available solution to a problem, an AND/OR graph is created. Then the Bayesian network is build incrementally. Finally the network provides the student's knowledge assessment regarding the physics problem.

HYDRIVE (Mislevy and Gitomer, 1996) has been developed to help simulate the important cognitive and relative features of troubleshooting on the flight line [15].The problem is given in a video where the pilot is seen describing some aircraft malfunction to the technician. The interface then offers the student several options for performing troubleshooting procedures with the help of online technical support material. The student's performance is thus evaluated based on how he utilizes the given information provided in the material. The students understanding are characterized in terms of dimensional variables which is represented and updated by a Bayesian network. Rule based inference also is an important part in this system.

ANDES (Conati et al.,1997) is an Intelligent Tutoring System for solving Newtonian physics problems via coached problem solving(VanLehn,1996) [16].This model takes help of Bayesian network for assessment of student knowledge, students' action prediction and plan recognition. An approximate algorithm is used to trace and update the network. Here the tutor proceeds along with the student, as he gives the correct response to a problem. Whenever there is the student come across a doubt, the tutor provides tailored hints that help the student to overcome his doubt. Each problem is associated with a solution graph. The Bayesian network is constructed automatically from the respectable solution graph and the corresponding conditional probably are updated. These networks are too large to be solved and the worst case of propagation in a Bayesian network is NP-hard.

The last decade there has been a growing interest among researchers to utilize the theoretically sound approach of Bayesian Network in user modeling systems, especially in the field of education [17]. The old Microsoft Office Assistant used the concept of Bayesian Network to provide users with suggestions and predictions. A survey of systems using Bayesian network paradigm to use or student modeling is given in [18] We have extended the ITS framework to introduce an evaluation model which uses Bayesian Network for mixing prior statistics and user inputs in order to for user knowledge evaluation.

PROPOSED SYSTEM

The necessity of any ITS model to be modular is its ease to update the system when necessary. A simple change in the system does not require the need to reconstruct the whole model. This is the reason why a typical ITS system is modeled into different components.

THE IIOAB JOURNAL
ISSN: 0976-3104

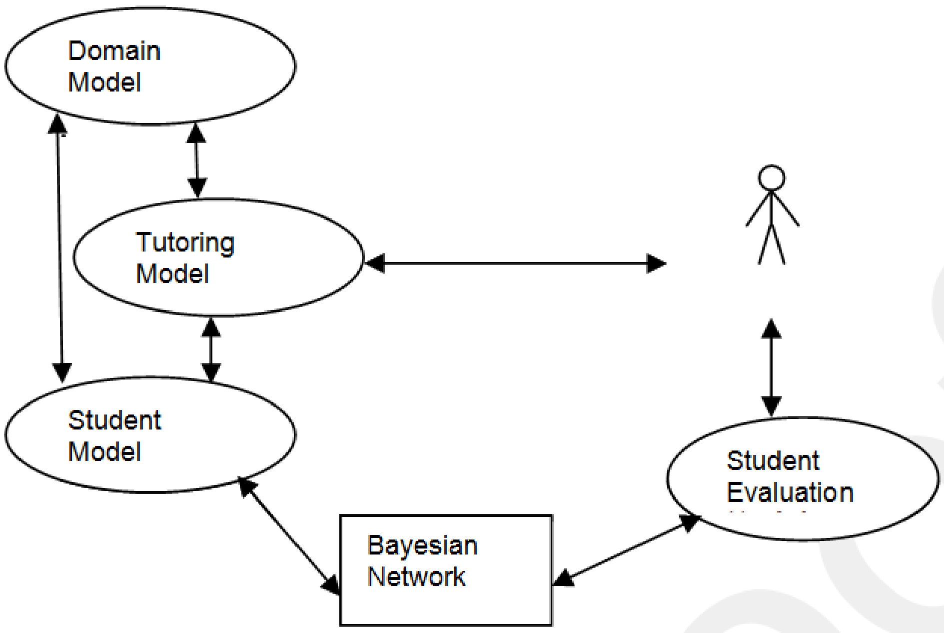

Fig: 2. Proposed ITS framework

……………………………………………………………………………………………………….

Our proposed domain module contains the detailed structure and information for each course .It consists of the learning material of all the sub-topics of the course. Our domain model is designed in the form of a tree like structure. The particular course 'Database Management System' is used as for the domain content case study in the proposed system. It is divided into all the fundamental sub-topics, which are again sub-divided. Each of the leaf nodes represents a particular concept. Corresponding to each concept there is a set of question that will be used to evaluate student knowledge about that concept.

Student Model

In our proposed system we need to initialize the knowledge level for each student before he/she starts the learning process [19]. We are able to determine the initial knowledge level of each student for each concept from the following flowchart [**Figure-3**].

Firstly a student logs in the system. Then he/she chooses from the available courses. A pre-evaluation test is performed which is used to determine the knowledge levels for each concepts. This is used to initialize the prior student database. The pre-evaluation test consists of a set of 20 questions. Each question is made up of a combination of several concepts. Here each of the concepts in the question has certain weight associated with it. Hence the response to each question is used to calculate the concept weights. This procedure gives us a certain idea about the initial knowledge level for each student.

Each student will be having a definite knowledge level for each concept. Once we finish the initial evaluation, all the calculated concepts are categorized into three levels poor, good and excellent. This is helpful to keep a well defined understanding of the current knowledge level for each concept so as to provide the right guidance to the student. The required study material corresponding to each concept level is provided to the students by an e-tutor so that he/she can improve themselves by learning.

Study-Material Adaptation

After the pre-test a adaptation process combines the student model with the domain model to deliver appropriate course content to the student. Different kinds of adaptive technology are available. The proposed system uses the link hiding technique.

Link hiding is implemented by showing only relevant links that are suitable to the current knowledge level of the student and hiding irrelevant links from her/him. Link hiding protects the student from the complexity of all section links in a course and reduces her/his cognitive overload in hyperspace. The proposed system hides section links that have section knowledge levels lower than the student's overall knowledge level.

THE IIOAB JOURNAL
ISSN: 0976-3104

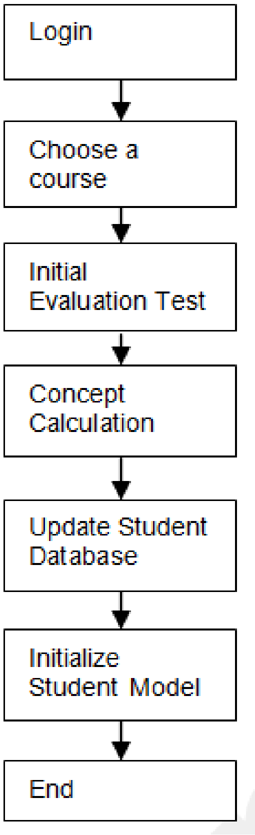

Fig:3. Flowchart for initialization of student model

Student Evaluation Model

After the students have completed all the learning exercises, he/she has to undergo the final evaluation test which will recalculate the concept levels efficiently. Questions from different concepts are randomly picked and given to the students and their responses along with the previous knowledge levels are calculated using the concept of Bayesian network based on the Corbett and Anderson’s Bayesian Knowledge Tracing model [20]. This is the key functionality of our Student Evaluation Model.

To explain this, we start by modeling a simple case, where a single question of a certain difficulty level is given by the system. In order to obtain information about the student's current knowledge about a domain, we use the student's answers or responses to given questions. But it is not necessary that a student gives a correct answer only if he/she has the knowledge related to the domain. Student can make random guess. On other hand given a student has knowledge about a domain he/she make mistakenly make a slip. This is known as unreliable information. So the system should consider all the evidences it has, including the student's response, to decide what is the current student knowledge. This requires reasoning under uncertainty which is handled by a belief network where the Bayesian estimate has been applied.

We compute the probability of a student’s current knowledge about a particular domain using the fundamentals of Corbett and Anderson’s knowledge tracing model.

i. Here we differentiate each question given to a student into particular levels of difficulty (QL) (from 0-1).
ii. All the students assumed to be already graded to a certain intelligent level (IL) (ranging from 0-1).
iii. In our proposed model a certain student appearing for the question either response correctly (C) or incorrectly.
iv. A student with lower knowledge level is more likely to give an incorrect response, but there is a slight chance he/she might answer correctly by making either a guess or cheating (called G).
v. On other hand a student with a higher knowledge level can also answer incorrectly, by making a silly mistake or slip (called S)

vi. Let Ki-1 be the initial knowledge probability of a student before answering a question, either correctly or incorrectly. This parameter will be dynamically updated for each response.

vii. The student also has a probability of learning a skill (called L), while answering a question.

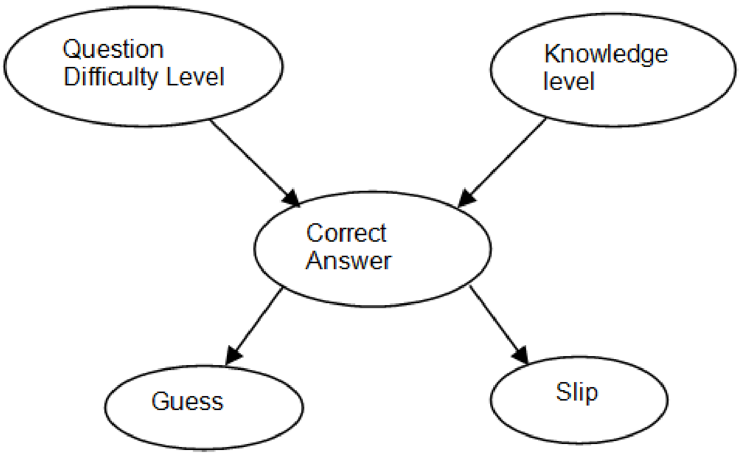

Fig:4. Bayesian Network Construction

……...

Hence we compute equations to compute student's knowledge from Bayesian Knowledge Tracing [21]

$$P(K_{i-1}|C_N) = \frac{P(K_{i-1})(1 - P(S))}{P(K_{i-1})(1 - P(S)) + (1 - P(K_{i-1}))\, P(G)}$$

Equation 1.1

$$P(K_{i-1}|\sim C_N) = \frac{P(K_{i-1})P(S)}{P(K_{i-1})(P(S)) + (1 - P(K_{i-1}))\,(1 - P(G))}$$

Equation 1.2

As we already mentioned the all the students are already has an initial knowledge probability which is to be re-evaluated and updated, taking into account the response given to a question by help of Bayesian rule. The updating is done as follows:

After we receive the output post the nth interaction the appropriate changes in the student knowledge at nth time can be incorporated for this we have to determine whether a student is in the learned state or not. We take into account two possible cases for this:

$P\,(K_i \mid K_{i-1}, answer)$

This probability denotes being in learned state given the student was already in the learned state.

$P\,(K_i \mid \sim K_{i-1}, answer)$

This probability denotes being in learned state given the student was not in the learned state.

Adding the two above mentioned cases we get the probability of learning of each student:

$$P\,(K_i \mid answer) = P\,(K_i \mid K_{i-1}, answer) * P\,(K_{i-1} \mid answer) + P\,(K_i \mid \sim K_{i-1}, answer) * P\,(\sim K_{i-1} \mid answer)$$

Equation 1.3

THE IIOAB JOURNAL
ISSN: 0976-3104

EXPERIMENTATION AND RESULTS

We evaluated 15 students, taking 5 students each from 3 levels of a certain concept from the system. Next they are tested on question of different difficulty levels and each time their results are recorded.

Table: 1. Response record from experiment

Student id	Knowledge level of a concept	Question level	Response
IT/01	0.10	High	Correct
IT/02	0.37	Average	Incorrect
IT/03	0.23	Low	Correct
IT/04	0.19	Low	Correct
IT/05	0.31	High	Incorrect
IT/06	0.43	High	Incorrect
IT/07	0.56	Average	Incorrect
IT/08	0.67	Low	Correct
IT/09	0.49	Low	Correct
IT/10	0.51	High	Incorrect
IT/11	0.78	High	Correct
IT/12	0.88	Average	Correct
IT/13	0.90	Low	Incorrect
IT/14	0.70	Average	Correct
IT/15	0.95	High	Incorrect

Anomalies and Correction

We can consider some anomalies from above data. For example by considering an extreme case of student with id IT/01.Here the student is of low intelligent level and is given a high difficulty level question which he answers correctly. So there is chance that he might have either guessed the answer or have taken some unfair means to do so. In case of a traditional case we have to wrongly assign a high marks to this student. But in our system we have incorporated some probability of guessing with the help of Bayesian network to impute some degree of correction.

Again, by considering another extreme case of student with id IT/13.Here the student is of high intelligent level and is given a low difficulty level question which he answers wrongly. So there is chance that he might have either made a silly error or have wrongly imputed the answer. In case of a traditional case we have to wrongly assign a low marks to this student. But in our system we have incorporated some probability of slip with the help of Bayesian network to impute some degree of correction.

The calculation for the above two cases have been shown in the following section.

Calculation of results based on Bayesian Network

Table: 2.Conditional Probability Table for Guess parameter.

Question Difficulty Level	Intelligence Level	Correct Response	Probability (P(G=T|IL,C,QL)
High	1	True	0.88
High	1	False	-
High	2	True	0.76
High	2	False	-
High	3	True	0.35
High	3	False	-
Low	1	True	0.43
Low	1	False	-
Low	2	True	0.23
Low	2	False	-
Low	3	True	0.12
Low	3	False	-
Average	1	True	0.73
Average	1	False	-
Average	2	True	0.67
Average	2	False	-
Average	3	True	0.29
Average	3	False	-

Table: 3.Conditional Probability Table for Slip parameter.

Question Difficulty Level	Intelligence Level	Correct Response	Probability (P(S=T|IL,C, QL)
High	1	True	-
High	1	False	0.008
High	2	True	-
High	2	False	0.05
High	3	True	-
High	3	False	0.22
Low	1	True	-
Low	1	False	0.08
Low	2	True	-
Low	2	False	0.76
Low	3	True	-
Low	3	False	0.89
Average	1	True	-
Average	1	False	0.23
Average	2	True	-
Average	2	False	0.42
Average	3	True	-
Average	3	False	0.6

Showing calculation of student knowledge for students with id (IT/01) with intelligence level 0.01(Low Intelligence Level) who is given a question with high difficulty level given that he/she answers it correctly:

THE IIOAB JOURNAL ISSN: 0976-3104

$$P(K_{i-1}|C_n) = \frac{P(K_{i-1})(1-P(S))}{P(K_{i-1})(1- P(SLIP)) + (1-P(K_{i-1}))\,P(G)}$$

$$= \frac{0.1}{0.1+0.9*0.73}$$

$$= \mathbf{0.13} \quad \textbf{(Low Intelligence Level)}$$

In a traditional system this student will be graded a high score without taking into any account of whether he/she has performed any guess or unfair mean.

Showing calculation of student knowledge for students with id (IT/13) with intelligence level 0.09 (High Intelligence Level) who is given a question with low difficulty level given that he/she answers it incorrectly:

$$P(K_{i-1}|\sim C_n) = \frac{P(K_{i-1})P(S)}{P(K_{i-1})(P(SLIP)) + (1-P(K_{i-1}))\,(1-P(G))}$$

$$= \frac{0.9*0.89}{0.9*0.89+0.1}$$

$$= \mathbf{0.889} \quad \textbf{(High Intelligence Level)}$$

In a traditional system this student will be graded a low score without taking into any account of whether he/she has performed any slippage or silly mistake.

CONCLUSION AND FUTURE WORK

In this paper we have proposed an Intelligent Tutoring System which has the potential to enhance the traditional e-Learning systems by incorporating reasoning based probabilistic approach namely Bayesian Network to evaluate a student's knowledge. The aim of this system is to provide a student the correct feedback at the correct time.

We have seen from our experiment results that evaluation of current knowledge level of each student is corrected by incorporating both the Guess and Slip parameter in the Bayesian network. Hence in case of a highly intelligent person performing a slippage, their knowledge level is reduced by a small percent. Also in case of a low intelligent person performing a guess, their knowledge level is increased by a small percent. Since experiment was conducted on 15 students, our aim is to increase the number of students from various backgrounds in order to obtain a wide range of probabilities. A large set of data will give us a more convincing result and will also enhance the validity of our proposed system. But such an attempt will also give rise to the challenge of managing a large number of data. We need to device a mechanism to handle the computation complexity of Bayesian Network for large data input.

The other major directions for future work regarding our system include the following:

- Enhancing the study material recommendation feature of the proposed system with advanced adaptive hypermedia technology (AHS).
- Evaluate individual students learning style from their performance and usage of study materials.
- Creating a more adaptive tutor model which will maintain an estimate of all the probabilities and the learning styles and present more individualized exercises and feedback for improvement of the student.
- Extending the proposed Bayesian network to a Dynamic Bayesian Network (DBN) which can update student's knowledge over long time spans.

ACKNOWLEDGEMENT

None

CONFLICT OF INTEREST

Authors declare no conflict of interest.

FINANCIAL DISCLOSURE

No financial support was received to carry out this project

REFERENCES

[1] Joseph Psotka, Sharon A. Mutter. [1988] Intelligent Tutoring Systems: Lessons Learned. Lawrence Erlbaum Associates. ISBN 0-8058-0192–8.

[2] Nkambou R, Mizoguchi R, Bourdeau J. [2010] Advances in intelligent tutoring systems. Heidelberg: *Springer*.

[3] Anderson H, Koedinger M. [1997] Intelligent tutoring goes to school in the Big City. *International Journal of Artificial Intelligence in Education*, 8: 30–43.

[4] 4 Millán, Eva, and José Luis Pérez-De-La-Cruz.[2002] A Bayesian diagnostic algorithm for student modeling and its evaluation. User Modeling and User-Adapted Interaction 12.2-3: 281–330.

[5] Butz, Cory J, Shan Hua, and R Brien Maguire. [2006]A web-based bayesian intelligent tutoring system for computer programming. Web Intelligence and Agent Systems 4.1: 77–97.

[6] P Brusilovsky.[1999] Adaptive and intelligent technologies for Web-based education, *Special Issue on Intelligent Systems and Teleteaching*, 4: 19–25.

[7] P Brusilovsky and MT Maybury. [2002] From adaptive hypermedia to adaptive Web, *Communications of the ACM, Special Issue on the Adaptive Web*, 45(5): 31–33.

[8] C Liu, L Zheng, J Ji, C Yang and W Yang. [2001]Electronic homework on the WWW, in: Proceedings of First Asia-Pacific Conference on Web Intelligence, Maebashi City, Japan, , pp. 540–547.

[9] J Pearl. [1988] Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann.

[10] SKM Wong and CJ Butz.[2001] Constructing the dependency structure of a multi-agent probabilistic network, *IEEE Transactions on Knowledge and Data Engineering*, 13(3) :395–415.

[11] SKM Wong, CJ Butz and D Wu. [2000]On the implication problem for probabilistic conditional independency, *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 30(6) :785–805.

[12] JM Agosta E Millan and JL Perez de la Cruz.[2001] Bayesian student modelling and the problem of parameter specification. British *Journal of Educational Technology*, 32 (2)

[13] Kurt VanLehn and pages=179-221 year=1998-22 Joel Martin title=Evaluation of an assessment system based on Bayesian student modeling, journal=*International Journal of Artificial Intelligence in Education*

[14] Cristina Conati and Kurt VanLehn Pola.[1996] A student modeling framework for probabilistic on-line assessment of problem solving performance. Proceedings of the Fifth International Conference on User Modeling.

[15] Linda S Steinberg and Drew H Gitomer.[1996] Intelligent tutoring and assessment built on an understanding of a technical problem-solving task. *Instructional Science*, 24.3:223{258 }.

[16] Cristina Conati.[1997] On-line student modeling for coached problem solving using bayesian networks. *User Modeling. Springer Vienna*.

[17] Chrysafiadi K and Virvou M.[2013]] Student modelling approaches: A literature review for the last decade. *Expert Systems with Applications*, 40(11): 4715–4722

[18] Jameson Anthony. [1995]Numerical uncertainty management in user and student modeling: An overview of systems and issues.*User Modeling and User-Adapted Interaction* 5.3-4: 193–251.

[19] Esichaikul, Vatcharaporn, Supaporn Lamnoi, and Clemens Bechter. [2011]Student modelling in adaptive e-learning systems. Knowledge Management & E-Learning: *An International Journal (KM&EL)* 3.3: 342–355.

[20] ATCorbett and JR Anderson.[1995] Knowledge tracing: modelling the acquisition of procedural knowledge. *User modelling and user-adapted interaction* 4:253–278.

[21] D Baker, Ryan SJ, Albert T Corbett, and Vincent Aleven.[2008] More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. *Intelligent Tutoring Systems*. Springer Berlin Heidelberg.