Bayesian Networks for Learner Modeling

Anouar Tadlaoui Mouenis, Khaldi Mohamed, Aammou Souhaib

Abstract— Bayesian networks tend to be increasingly used for the management of uncertainty in modelling of the learner. They have been successfully used in many systems, with different objectives. However their use as part of the cognitive process modelling raises a number of problems. On the one hand the underlying semantics of arcs is often not clearly explained, and on the other hand the evolution of the knowledge structure is not taken into account. Our work focuses on the question of the orientation of the arcs, and more generally on the structure of Bayesian network modeling of the learner. We try to show in this work how this question is crucial. In addition, the issue of structural adjustment in the network behavior of the learner sometimes had been suggested, and while different results from cognitive psychology attests to the existence of structural differences by level of expertise. The central hypothesis of our work is that has been a link between the structure of the learner model and level of expertise. We present our probabilistic graphical models of multi-networks to take into account several networks within the same model. The experiments presented in this work are arguments in favor of our hypothesis on the link between the level of expertise of the learner and the structure of Bayesian network.

Index Terms— Bayesians Networks, cognitive diagnosis, Learner modeling;

I. INTRODUCTION

The main hypothesis of our work is that there is a link between the structure of the learner model and the level of expertise. First , and after talking briefly about the definition of Bayesian networks , we'll discuss in a detailed way the difficulties encountered in the use of Bayesian networks learner modeling (by focusing on the most common case that is to say, the construction by elicitating the expert knowledge). Then we present the multi- networks wich are a probabilistic graphical models that take into account several networks within the same model and thereby provide a framework in which we will test our hypotheses. Finnaly, we are going to end our paper by showing the implementation and the experimental results of a learner model using multi-network in a specific case. The problem of this paper can be summarized as follows: why and how to take into account several competing Bayesian networks into the same model of the learner? This consideration is it experimentally justified?

Manuscript Received on August 2014.

Mr. Anouar Tadlaoui Mouenis, Abdelmalek Essaâdi University, Faculty of Sciences, Laboratory of Informatics, Research Operational and Statistic Applied (Lirosa), Tétouan, Morocco.

Prof. Khaldi Mohamed, Abdelmalek Essaâdi University, Faculty of Sciences, Laboratory of Informatics, Research Operational and Statistic Applied (Lirosa), Tétouan, Morocco.

Prof. Aammou Souhaib, Abdelmalek Essaâdi University, Faculty of Sciences, Laboratory of Informatics, Research Operational and Statistic Applied (Lirosa), Tétouan, Morocco.

II. BAYESIAN NETWORKS

A. Definition

A Bayesian network [1] is a directed acyclic graph and in which the nodes correspond to the variables (user properties) and links represent probabilistic relationships influence. These variables can belong both to the field of knowledge, the knowledge base and / or cognitive model. Each node represents the belief system about possible values (levels, states) of the variable. Thus, the conditional probability distribution must be specified for each node. If the variables are discrete, they can be represented as a table. The graph is also called the "structure" of the model, and the probability tables its "parameters". The structure and parameters can be provided by experts, or calculated from data, although in general, the structure is defined by experts and the calculated parameters from experimental data.

B. Construction of a Bayesian Network

As we have seen in the definition, the complete specification of a Bayesian network requires specifying a share structure (directed acyclic graph that underlies) and other parameters (probability tables). To do this, two approaches are possible and can be combined: the collection of expertise and machine learning, which is one of the attractions of Bayesian networks. In the case of collection of expertise, the definition of the network structure begins with the identification of possible nodes and the distinction between (unobservable) informational variables (inputs) or hypothetical. The existence of an arc can be analyzed in terms of influence of one variable on another, but its orientation is more difficult. Traditionally, an arc is directed from A to B if A is a cause of B, but we will see that this interpretation is not as simple in the case of the learner modeling. The parameters are in turn attached in an approximate manner by using frequentists or qualitative information. Since Bayesian network is a probability distribution, we can use maximum likelihood as statistical learning parameters criterion. The result is as a Bayesian network whose structure is fixed and E which is a comprehensive basis of example, the maximum likelihood is achieved if the parameters of the Bayesian network are equal to the frequencies of the same features observed in E. statistical learning structure requires for its development test to determine whether or not the random variables are conditionally independent [2].

C. Examples of Learner Modeling by Bayesian Networks

Andes [3] is a tutor for help in learning problem solving in Newtonian mechanics that assesses the state of knowledge of the learner to recognize the plan he follows and predict its possible actions future. Capit [4] to select the most effective remediation for a given learner, a teacher in the English punctuation system aims. To do this, the decision theory is used: network modeling learner consists only of informational variables representing either the actions of the learner or remediation decisions and their uses. It is initialized from a test group using machine learning methods

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.*

and is then used to make decisions about instructional strategies to follow.

III. ARCS ORIENTATION

The definition of the structure of a student model based on Bayesian network from expert knowledge, is often done from a representation of domain knowledge, we do not discuss here the validity of the skeleton (graph undirected) resulting in the type of links that are taken into account (mainly partitive or generic). Therefore, obtaining the model structure from the skeleton of the orientation requires arcs. These guidelines determine the diagnosis obtained, since as they play a key role in the relationship dependency between variables.

A. The Influence of the Orientation of the Arcs on the Diagnosis

We point out an example of the influence on the arcs orientation of the diagnosis obtained. “Fig. 2” shows a simplified version of the model Hydrive, where the performances of the electronics learner can be observed. The set of variables representing the different skills of the learner is affected by this observation. Thus, if the student is found to be competent in electronics, for example because of their initial training, the diagnostic model is the same in all other diciplines. Consider “Fig. 2” a graph having the same skeleton but with other orientations of the arcs. The spread of the information here is much more limited. The diagnosis we got depends not only on arcs orientation’s choices, but also it seems that the orientation depends on the expertise level of the learner. Indeed, it seems reasonable to consider the network of “Fig. 1” for a subject having followed all the training (for which we can assume a homogeneity especially in some skill levels), the network 3 seems more appropriate for a about the beginning of training, which can be very proficient in a particular area because of its course without the need to master all the skills.

B. What Choices for the Orientation of Arcs?

In literature, the choice is massively in favor of a focus node of the domain layer to those of the task layer. The question remains open regarding the arcs linking the nodes in the domain layer. It is common to present this choice as an alternative choice between a general orientation to the individual or its opposite. We find in literature examples of these two choices ([5], [6]), even if the justification given for this choice is not always totally convincing, even if sometimes it doesn’t exist. This dichotomy itself is to our knowledge never has been partly explained the concept of equivalent Bayesian networks under Markov [7]. The orientation of the arcs from the general to the particular is suitable when the skills of the learner have certain homogeneity because of dependencies of this orientation. The orientation on the other side is most appropriate for learners with diverse skills. Moreover, if we take the findings on the relationship between network structure and level of expertise of the learner, we conclude that the model must allow this structural change.

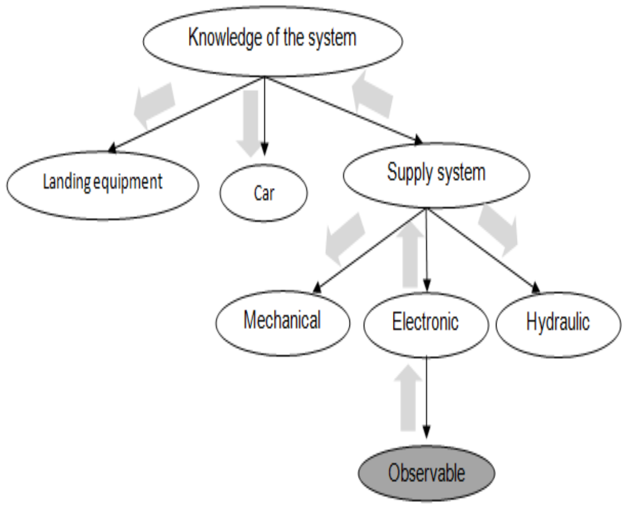

Figure 1.Flow of information in the learner model of Hydrive

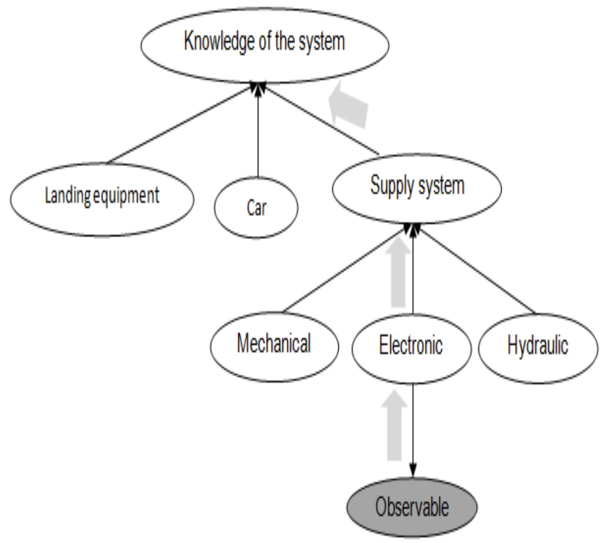

Figure 2. Flow of information in the modified learner model of Hydrive

C. Asymmetry Dependence

In probability, we talk about asymmetric dependence when two dependent variables are being given certain values of one of them being independent data and other values [8]. In [3], they presented a typical case of asymmetric dependence in learner modeling (which is not marked as such)). It’s the description of a model consisting of a Bayesian network representing the knowledge of the student on the whole decomposition into prime factors. The construction of the learner model is made from the following observation of teachers, if a student knows the decomposition of a number, then (usually) he knows those of its divisors, and the opposite is true. In other words knowledge of the decomposition of a number and those of its divisors that are dependent if the student knows the breakdown of the number in question: it is indeed a case of dependence asymmetric (since it depends on the values of the variables). This situation cannot be modeled by a Bayesian network. Therefore, the constructed model (which is a Bayesian network) cannot really take into account observations of teachers.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

IV. LEARNER'S MODEL UPDATING

The updating requires acquiring information about the user's behavior and adjusting the user model. The acquisition is the process of gathering user input corresponding to user interactions in a hypermedia application, such as pages visited, the steps of solving problems of a learner, etc... The problem is the interpretation of data (mouse clicks, keyboard input, etc...) In actions or proposals which are not insignificant. The acquisition process consists of two phases: data collection and diagnosis. In the diagnostic process two steps can be distinguished: transformation and evaluation.

A. Data Collection

The main problems related to data collection are: the reliability of the data, the amount of data available and the detail's level of the data. How much data is needed depends on the granularity of the model. In hypermedia systems, there is an additional problem concerning the registration of user interactions, primarily due to HTTP protocol, which offers little support in the process of data collection. Thus, data are obtained by additional ways.

B. Diagnosis

The diagnosis consists of two stages: a transformation of the data collected and an assessment of user behavior. Processing consists in extracting important information from the data collected in order to judge the qualifications of the user. This can be done in two ways according Ragnemalm [9] and Vassileva [10]. The inputs from the user's behavior can be converted into a closer representation of the model. The techniques used for this conversion may be domain dependent. The user input must be described in a set of proposals [11] The problem is to find a function:

$$\text{interpret } (\{d_1, d_2, d_3, \dots, d_n\}) = (\{p_1, p_2, p_3, \dots, p_m\})$$
$$\text{Such as } B_S B_{Upj} \text{ for } j = 1 \text{ à } m$$

But the interpretation is generally more complexed because it should be based on what entries stated in terms of user beliefs. The properties of the user in the model can be converted in the screen closer to the user input, that is to say to the criteria or patterns for identifying the user input. It can be defined by analogy to the deduction of patterns recognition (recognition patterns):

$$\text{Deduct } (\{p_1, p_2, p_3, \dots, p_m\}) = (\{r_1, r_2, r_3, \dots, r_n\})$$
$$\text{Such as } B_S B_{Upj} \text{ for } j = 1 \text{ à } m$$

And the set of patterns recognition $\{ r_1, r_2, r_3, \dots, r_n \}$ compared to user input during the assessment $\{d_1, d_2, d_3, \dots, d_n\}$. The evaluation refers to the process of comparing the user's behavior in a certain design with "best" behavior, which is explicitly or implicitly represented in the expert model. The diagnostic process is to match the data (d) with a model (UM) or inversely match the model to the data. Thus, the process of diagnosis ranged between two extremes:

- Approach purely directed by the data (data-driven): the diagnosis is made on the user behavior without reference to a predefined template,
- Approach purely directed by the models (model-driven): this method produces the expected patterns and matches the behavior of the user.

Data-driven approaches are appropriate for simple domains. Model-driven approaches are appropriate for complex domains. Most diagnostic methods are situated between these two extremes, such as reconstruction, model tracing, or induction. A brief explanation is given of some of them. The following work can serve as a reference for a detailed description of diagnostic methods: Self [11], Jameson [12] and Ragnemalm [9].

V. A MODEL CONSTITUTED OF A SET OF NETWORKS: THE MULTI-NETWORKS

In this section, we present the multi-networks that are probabilistic representations of knowledge to be taken into account asymmetries dependencies.

A. Presentation

Multi-networks generalize the Bayesian networks in the same direction as mixtures of Gaussian generalized Gaussian distributions: a multi-network consists of several Bayesian networks, each of which has an associated probability. Different Bayesian networks are formed constituting the same nodes, but have different topologies. Because of its generalizing character (a Bayesian network is a multi-network with a single component), this formalism allows to take full advantage of the expertise in learner modeling using Bayesian networks. In the following, we assume that the multi-network consists of n Bayesian networks $n_1, \dots, n_n$, whose respective probabilities are rated $p_1, \dots, p_n$. The probability of an event E is given by: $P(E) = \sum_{i=1}^{n} P(E|n_i)$, ou $P(E|n_i)$ is the probability of E in the networkn_i.

B. Construction of the Various Networks

We limit ourselves to the analysis of learner responses to exercises in areas where knowledge of the learner may decline in knowledge and know-how. We consider three types of nodes: the knowledge and know-how, which are the domain layer, and items that model the responses of the learner and provide for their task layer. We further assume that we have a representation of domain knowledge in the form of non-oriented graph in which the items are related only to know-how. The causal traditional interpretation of arcs in a Bayesian network (the arc is oriented from A to B when A is a cause of B) does not seem the most relevant to us in the case of the learner modeling. The question of whether a student is good in databases management because he solves a lot of exercises or because, he solves a lot of exercises because he is good in databases management, seems particularly difficult to resolve. It is the same regardley the deletion of any causal link between the knowledge that a learner on databases and those of its objects. Therefore, to determine the direction of the arcs between the different variables, we use a systematic analysis of the dependencies between variables. In summary, this analysis allows us to confirm the validity of the choice usually done in literature of arcs orientation of the domain layer nodes to those of the task layer, while confirming the impossibility of determining in first place the orientation between knowledge nodes.

C. Learning Parameters

When using multi-networks that we consider has a significant simplification compared to the general case: we assume that the structures of different networks considered and conditional probability tables are known. Only to be determined in first place the probabilities of the different networks and the law of the hidden variable determining the

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

probability of each network. As long as we consider that each learner is modeled by a network, we propose to use the EM algorithm [13] after assigned to each data corresponding to it. specifically (see "Fig. 3", the data is recorded $d_1, \dots, d_n$) after the random initialization of probability of each network, each data is assigned to the most likely network, then it shows us the missing parameters of each of them using the EM algorithm. Which then calculate the probability of each data network knowing then that the loop is repeated until convergence of p_i.

D. Diagnosis

A multi-network provides two types of diagnostics: local diagnosis and comprehensive diagnosis. The local diagnosis is obtained by calculating, for each variable in the domain layer K the probability table knowing the values of K observable. Conceptually this diagnosis is the same type as that carried out in a student model based on a single Bayesian network: it is a calculation of conditional probabilities. However, this calculation takes into account the different networks corresponding to different levels of expertise. The overall diagnosis is a such quite nature different: it doesn't express in terms of belief on the acquisition of particular knowledge, but in terms of network that best matches the actions of the learner. This network is one that maximizes $P(n_i|d_j)$ among the networks considered in the model. Since different model networks change in terms of structure, the overall diagnostic provides guidance on structuring the knowledge of the learner. It is important to note that this is the structure of multi-network that gives meaning to the phrase "network that models the best the learner."

VI. EXPERIMENTS

In this section we will present the experiments that we have established [14], the results that we have obtained and the algorithm used to achieve the results that back up our hypotheses.

A. Model Description

The exercises proposed require knowledge of relational modalization (the relationship, attribute, domain, the operators ...) and database design (the entity model association and standardization relationship). The layers skeleton of different Bayesian networks constituting the model multi-network of the learner consists of a node for each conceptual knowledge or modeling needed to solve exercises to which are added nodes Datasets , Design of database and relational modalization.

1) 5 Selected Structures

Following our analysis, from the skeleton we've obtained 8 different networks (two possible orientations at each of three nodes: DataBase, design databases and relational modalization). Each of these networks can be interpreted in terms of homogeneity or heterogeneity of knowledge of the learner. For example, the network 6 in "Fig. 4" models a learner with homogeneous knowledge in database design and heterogeneous knowledge in relational modalization. Eliminating networks which are not fairly interpreted (e.g. modeling a network with heterogeneous learning design database knowledge but homogeneous in Databases), we reduce the number of components of the model of the learner to 5 "Fig. 4". The layer task of the various networks consists of 34 Items modeling answers to the exercises, each of them being connected with the know-how to implement them.

2) Parameters

Relationships between nodes in the domain layer are not noisy, the root nodes (ie those with no parents) following an equiprobable law. At the task layer, the parameters represent the probability of being wrong in its area of competence (uniformly set at 0.15), or guessing the correct answer (based on the number of possible answers). The probability of occurrence of each of the 5 networks is set at 0.2.

B. Results

We worked on a set of 400 files of answers to the exercises. These learners are 400 students in the first and second year of a technical diploma called a "Diplôme universitaire de technologie" . For each of them, we use the class as an indicator of the level of expertise, given that some students of the second year may be more competent than some students of the first year. Since the overall diagnosis allows us to obtain the network (and therefore the structure) which models best learner, and we want to test a hypothesis on the link between this structure and the level of expertise of the learner, we so try to see if there are significant differences between these two populations from the perspective of the overall diagnosis. In other words we want to know if there is a correlation between the result of the overall diagnosis and the grade level. Distributional differences obtained between the students in the first year and those of the second year were measured using the Chi's test [15]. We have made experiments with or without learning, is being made on all or part of data. Note that when we talk about learning, it's just learning the parameters of the multi-network, and the indication of the class of the study is not in the files used in this learning.

Determine randomly p_i	
Repeat	
	For i varying from 1 to m and j = 1 to n For i=1 to m For j=1 to n Calculate : $P(n_j\|d_i)=\frac{p_j P(d_i\|n_j)}{\sum_{k=}^{n} p_k P(d_i\|n_k)}$
	For i=1 to m For j=1 to n Calculate P_j : $P_j=\frac{\sum_{k=}^{m} P(n_j\|d_k)}{m}$
	Assign d_i to n_j such as $j = argmax_{k \in \|1,\dots,n\|} P(n_k\|d_i)$
	Learn the parameters of each n_j using the EM algorithm from d_i.
Until convergence of p_i	

Figure 3. The Algorithm used in the Experiment

We provide in Table 1 the results obtained after learning all the data, all the results we have obtained are going in the same direction. First, the assumption of independence is rejected (test significant at 0.01). Then, if we look at the distribution of different networks in different classes, we find that first year there are overrepresentation networks modeling of learners with homogeneous knowledge in database design and underrepresentation of those modeling learners heterogeneous in databases design, and that is the exact

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

opposite in the second year knowledge.

Table 1. Correlation between Educational Level and the Overall Diagnostic Result

Structure	First year				Second year			
	Observed	Theoretical	Gap		Observed	Theoretical	Gap	-
BN 1	71	58,56	5,45	+	22	35,52	6,13	-
BN 2	12	9,22	2,84	+	4	6,78	1,09	-
BN 6	26	29,12	0,01	+	29	29,69	0,04	-
BN 7	4	5,65	0,53	-	8	4,65	0,64	+
BN 8	108	128,2	1,99	-	111	98,75	2,61	+

These results confirm our hypotheses. On the one hand, we have clearly shown a correlation between the level of expertise and the network structure modeling the learner, which is an argument in favor of the inclusion of several networks within the same model. On the other hand, we find a good over-representation of homogeneous networks (those whose edges are oriented from general to particular) among the more experienced learners and an over-representation of heterogeneous networks among those who are the least.

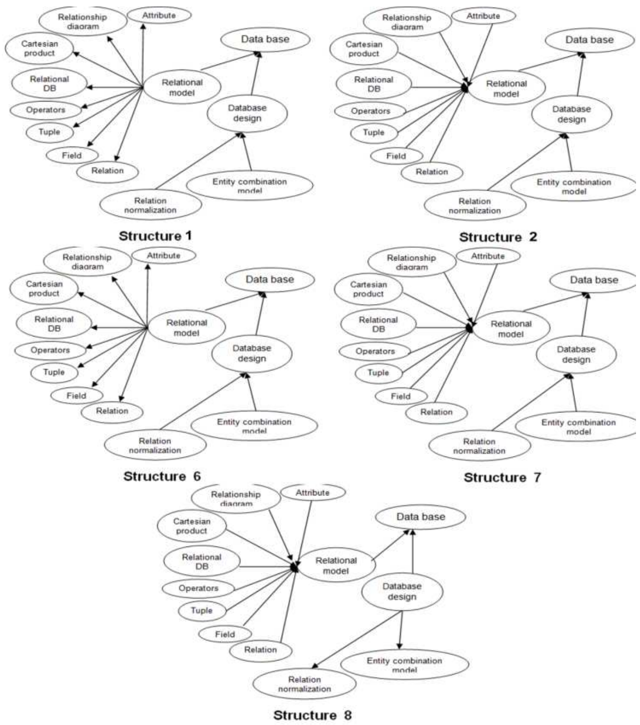

Figure 4. 5 Bayesian Networks Composing the Model

VII. Conclusion

We have shown how in a theoretical point of view and also taking into account the analysis of the literature, it seems justified to consider several Bayesian networks in a model of the learner, whether to take account the asymmetries of dependence or structural evolution of knowledge of the learner. We suggest using for this multi-networks that permit such a consideration in a mathematically well-founded framework. The experiments presented in this paper are arguments in favor of our hypothesis on the link between the level of expertise of the learner and the structure of Bayesian network that models. We see a main direction in which to pursue this work. Which is applying our ideas to more advanced conceptual models than those used so far, using PR-OWL, a Bayesian extension to the OWL Ontology Language that provides a framework for authoring probabilistic ontologies.

References

[1] [PEARL 88] PEARL, J., Probabilistic Reasoning in Intelligent System, Morgan Kaufmann. 1988.

[2] [BECKER & NAIM 99] BECKER, A., NAÏM, P., Les réseaux Bayésiens, modèles graphiques de connaissances, Eyrolles, 1999.

[3] [CONATI et al., 02] CONATI, C., GERTNER, A., VANLEHN, K., Using Bayesian networks to manage uncertainty in student modeling, Journal of User Modeling and User-Adapted Interaction, volume 12 (4), pages 371-417, 2002.

[4] [MAYO & MITROVIC 01]] Mayo M., Mitrovic. Optimising, A., ITS behaviour with Bayesian networks and decision theory, in International Journal of Artificial Intelligence in Education, n°12, pages 124-153, 2001.

[5] [MISLEVY & GITOMER 1996] Mislevy R., Gitomer D. (1996). The role of probability-based inference in an intelligent tutoring system, User-Modeling and User- Adapted Interaction, n°5, pp 253-282.

[6] [ZAPATA-RIVERA&GREER 2002] Zapata-Rivera J.-D., Greer J.(2000) « Inspecting and Visualizing Distributed Bayesian Student Models », Intelligent Tutoring Systems 00, pp. 544-553.

[7] [VERMA & PEARL 1991] Geiger D., Heckermann D. (1996). Knowledge representation and inference in similarity networks and Bayesian multinets, Artificial Intelligence, volume 82 (1–2), pp. 45–74.

[8] [GEIGER & HECKERMAN 96] GEIGER, D., HECKERMAN, D., Knowledge representation and inference in similarity networks and Bayesian multinets, Artificial Intelligence, volume 82 (1–2), pages 45–74, 1996.

[9] [RAGNEMALM, 1995] RAGNEMALM E. (1995). Student Diagnosis in Practice; Bridging the Gap. User Modeling and User-Adapted Interaction (UMUAI) International Journal, 5, Kluwer Academic Publishers, pp. 93-116.

[10] [VASSILEVA, 1996] VASSILEVA J. (1996). A Task-centered Approach for User Modelling in a Hypermedia Office Documentation System, User Modeling and User-Adapted Interaction Journal, 6, pp. 185-223

[11] [SELF ,1991] SELF J. (1991). Formal Approaches to Student Modeling. Technical Report AI-59. Lancaster University, England.

[12] [JAMESON 96] JAMESON, A., Numerical uncertainty management in user and student modeling: an overview of systems and issues, in User- Adapted Interaction, volume 5 (3- 4), n°5, pages 193-251, 1996.

[13] [LAUITZEN, 1996] Lauritzen S.L. (1995). The EM-algorithm for graphical association models with missing data, Computational Statistics and Data Analysis, vol 1, pp 191- 201.

[14] [M. ANOUAR TADLAOUI et al. 2014] M. Anouar Tadlaoui, M. Khaldi, S. Aammou (2014) TOWARDS A LEARNING MODEL BASED ON BAYESIAN NETWORKS,*EDULEARN14 Proceedings*, pp. 3185-3193.

[15] [CHI et al., 1981] Chi M.T.H., Feltovitch P.J., Glaser R. (1981). Categorization and representation of physics problems by experts and novices, Cognitive Science, vol 5, pp.121-152.

Mouenis Anouar Tadlaoui, is with Laboratory of Informatics, Research Operational and Statistic Applied (LIROSA) at Faculty of Sciences, Abdelmalek Essaadi University, Tetwan, Morocco.

Mohamed Khaldi, is with Laboratory of Informatics, Research Operational and Statistic Applied (LIROSA) at Faculty of Sciences, Abdelmalek Essaadi University, Tetwan, Morocco.

Souhaib Aammou, is with Laboratory of Informatics, Research Operational and Statistic Applied (LIROSA) at Faculty of Sciences, Abdelmalek Essaadi University, Tetwan, Morocco.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.