Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computers & Education 43 (2004) 345–359
COMPUTERS & EDUCATION
www.elsevier.com/locate/compedu

Prediction and assessment of student behaviour in open and distance education in computers using Bayesian networks

Michalis Xenos *

Hellenic Open University, 16 Saxtouri Str, GR 26221, Patras, Greece
R.A. Computer Technology Institute, Patras, Greece

Received 16 May 2003; accepted 18 September 2003

Abstract

This paper presents a methodological approach based on Bayesian Networks for modelling the behaviour of the students of a bachelor course in computers in an Open University that deploys distance educational methods. It describes the structure of the model, its application for modelling the behaviour of student groups in the Informatics Course of the Hellenic Open University, as well as the advantages of the presented method under conditions of uncertainty. The application of this model resulted in promising results as regards both prediction of student behaviour, based on modelled past experience, and assessment (i.e., identification of the reasons that led students to a given 'current' state). The method presented in this paper offers an effective way to model past experience, which can significantly aid in decision-making regarding the educational procedure. It can also be used for assessment purposes regarding a current state enabling tutors to identify mistakes or bad practices so as to avoid them in the future as well as identify successful practices that are worth repeating. The paper concludes that modelling is feasible and that the presented method is useful especially in cases of large amounts of data that are hard to draw conclusions from without any modelling. It is emphasised that the presented method does not make any predictions and assessments by itself; it is a valuable tool for modelling the educational experience of its user and exploiting the past data or data resulting from its use.

Keywords: Distance education and telelearning; Computer Science; Learning Strategies; Student Behaviour; Bayesian Networks

* Corresponding author. Present address: Hellenic Open University, School of Science and Technology, 23 Saxtouri Str, GR 26221, Patras, Greece. Tel.: +30-2610-367405; fax: +30-2610-362349.
E-mail address: xenos@eap.gr (M. Xenos).

doi:10.1016/j.compedu.2003.09.005

1. Introduction

In open and distance education the ability to predict and assess student behaviour is an inseparable part of the educational procedure. Especially in computer studies where the educational material continuously evolves adapting to new developments, prediction and assessment is of utmost importance. Normally, when dealing with a limited number of students, prediction and assessment is feasible without modelling, i.e., a motivated tutor responsible for 30 students in an open university can successfully handle the data concerning his/her students. On the other hand, in the case of the coordinator of 28 tutors, each one being responsible for about 30 students, thus resulting to a total of over 800 students – which is the presented case used as an example in this paper – modelling is essential for decision-making. When numbers increase, it becomes almost impossible to keep track of all the data required to make predictions and evaluate reached states. All the data may of course be kept in a computer database, but how can one handle such massive data and make decisions? How can the tutor track down behavioural patterns and locate student diversities that will enable the application of the appropriate educational approaches? The situation gets even more demanding in the case of courses facing high dropout rates, such as the Informatics course presented in this paper. In such case, for every module of the course a method is needed enabling one to make predictions of dropout rates, success rates, students likely to repeat the module, etc. Even more, one needs a way to assess a reached state, to evaluate student performance taking into account differences and to locate the reasons that are causing problems in the educational process, so as to be able to make corrective administrative decisions. In a few words, one needs a method to overcome conditions of uncertainty. In such cases, *modelling* is the only way to keep track of the increasing volume of student data. As presented in this paper, *causal modelling* (i.e., the use of Bayesian Networks) can prove to be very helpful in supporting education administrators and tutors in making decisions under conditions of uncertainty.

In an open and distance education course, the tutor plays the role of the 'interface' between the students and the university. Namely, in the open and distance educational model, the main role of the tutor consists in tracking down student needs and abilities and guiding them towards the right choices and decisions. Especially in a computer-related course where, normally, a variety of educational material is made available to students via the course web site, the above mentioned role of the tutor is of utmost importance. This role involves less traditional 'teaching' and more support in selecting the appropriate material from a variety of offered educational material, encouragement in difficulties, and – if possible – prediction of possible hurdles based on modelling and taking preventive measures. A tutor needs to keep a well organized and regularly updated student file enabling him/her to answer questions such as the following: (1) Based on the given student profile what is the student's success probability? The answer to this question will allow the tutor to better focus on the individual student needs, to locate students that require further encouragement or further educational assistance and to take measures towards success. (2) Based on student needs, what is the most appropriate educational material for specific students? A tutor should be able to answer this question before providing assistance to any student in selecting educational material. (3) What are the factors contributing to the success or failure of a student? The answer to this question will allow the tutor to repeat successful practices, determine the causes of failure and take corrective measures. Naturally, skilled tutors have their way of organising similar tasks, however, the increasing volume and complication of data and parameters affecting student performance can only be monitored with the utilisation of technology. In the presented case, the use of Bayesian Networks helped in modelling the behaviour of over 800 students each year and estimating hundreds of factors and parameters.

A literature review has showed that although causal modelling has been used in many different cases for decision-making under conditions of uncertainty in areas varying from engineering to aeronautics, it has not yet been applied into the educational procedure. Of great assistance in the creation of the presented model were previous surveys such as (Parker, 1999) regarding factors affecting student dropouts, (Kerka, 1996) discussing students' demographic characteristics and (Palmer, 2000) regarding computer usage. The results of these surveys were combined to the data derived from the particular course discussed in the presented example (Xenos, Pierrakeas, & Pintelas, 2002), as well as similar courses (Lupo & Erlich, 2001), in order to aid in constructing the probability tables required for the causal modelling.

This paper presents a methodological approach towards modelling student behaviour in conditions of uncertainty in order to enable prediction, status assessment and decision-making. The automated model, the development tools and the model's practical application are discussed. Specifically, section 2 presents the particular case of application in the Hellenic Open University, while section 3 briefly discusses the Bayesian Networks used for the described model. Section 4 presents the development of the model enhanced with examples and results from its application, while section 5 discusses conclusions and future goals.

2. Case description

The presented model was applied in the Informatics course of the School of Sciences and Technology of the Hellenic Open University (HOU); in particular, the modelling of student data from the 1st-year module, INF10, is presented to demonstrate the effectiveness of the discussed method. The Course of 'Informatics' is a 4-year course that comprises 12 modules and leads to a Bachelor Diploma in Informatics. Each student may register in one up to three modules per year. It should also be mentioned that each module is equivalent to 3 or 4 conventional university-level lessons, depending on the module's difficulty. The particularities and the nature of studies in the field of information technology, combined with the use of web-based material in the particular course, provided good conditions for the application of the model. Currently about 800 students are attending the INF10 module. These students are distributed in 28 groups, each comprising around 30 students. Each student group communicates with a single tutor and uses a variety of educational material either sent to the students or made available on the Web. A faculty member of the HOU serves as the coordinator of all tutors and all students. It must also be noted that the educational procedure in the Informatics course heavily depends on the use of computers. The tutor–student communication is based on e-mail and the use of electronic fora, while each module's Web site is an important part of the academic procedure. A variety of educational material is offered through the Web site including reading material, exercises, solutions to exercises, compilers, brief lectures, examples from past examinations, etc.

It is obvious that the large number of students makes it almost impossible for the coordinator to evaluate all student data without the use of a model that will enable the automation of certain measurements and estimates and will aid in decision-making. At this point, it is important to mention that the Informatics course in the HOU has to deal with a relatively high dropout rate. This rate ranges between 28% and 35% (Xenos et al., 2002) and mostly involves students that dropped out because of their failure in the INF10 module. This rate is by no means very high, especially for such a demanding course as the Informatics course, also considering that studies report dropout rates that range from 20% to 30% in Europe (Rumble, 1992), while in Asiatic countries the corresponding rates are as high as 50% (Narasimharao, 2000; Shin & Kim, 1999). Students drop out of their studies due to a number of reasons. Such reasons include difficulty in meeting the requirements of the module either because the students realise that they need to spend more study time than initially estimated, or because the module turned out to be more difficult than expected, or because the conditions under which these students started their studies have changed (Xenos et al., 2002). However, a certain percentage of dropout students would have continued their studies, if their needs had been traced and handled in advance. Even in the case of students who finally drop out, it is important for administrators and tutors to be able to identify the dropout causes based on an automated model. In this way, they can learn to handle particular student cases appropriately and progressively eliminate drop out causes that are not due to intrinsic, i.e., related to the student, factors. This is exactly what the presented method and the example of its application in the INF10 module aim at: predict and assess student behaviour in education in computers.

Each student attending the INF10 module has to reach six important milestones with regard to his/her academic progress. These include the results of each of the four written assignments that will determine if the student is allowed to participate in the final face-to-face examinations and the two examinations themselves. Naturally, a number of individual milestones affecting student performance also have to be taken into account such as important changes at work, changes in the student's family status, health problems, family problems, financial problems, etc. These kinds of differences influence not only each student's progress, but also their perception of the course's difficulty and in many cases determine the successful or unsuccessful completion of the course.

3. A brief introduction to Bayesian networks

The *Bayesian Networks* (BN) are also called Belief Networks, Causal Probabilistic Networks, and Probabilistic Influence Diagrams. The BN are types of graphical models, namely they are graphs the nodes of which illustrate random values and the acnes the cross-correlations between independent assumptions (Lauritzen & Spiegelhalter, 1990). Specifically, the BN is a special category of graphic model, in that the nodes represent variables and the acnes the relations between them (Jensen, 1996). Therefore a BN is a graphic network that describes the relations of probabilities between the variables. The use of BN brings *reason* in conditions of uncertainty and combines the advantages of intuition and experience with those of a mathematic model. The use of BN makes it possible to define the relation between the various nodes (variables) and to consistently estimate the way in which the initial probabilities influence uncertain conclusions such as the future behaviour of students. In this case, BN are used for *future estimation*. Furthermore, the BN can be used to speculate about the states of the initial nodes, based on a given final and some intermediate variables, for example when it is known that a student has failed in a module (final variable) given a certain professional and family status (intermediate nodes). In this case BN are used for the *assessment of an existing state*.

In order to define the relations between the variables, firstly the dependent probabilities that describe the relations between a 'child' node and its 'parent' nodes must be determined for each node. If the values of each variable are distinct, then the probabilities for each node can be described in a Node Probability Table (NPT). This table presents the probability that a 'child' node is assigned a certain value for each combination of possible values of the 'parent' nodes (Kschischang, Frey, & Loeliger, 2001). For example, Fig. 1 presents two 'parent' nodes (nodes B and C) and one 'child' node (node A). The probability table of node A reflects the probability $P(A|B,C)$ for all possible combinations of A, B, C. Thus, since there are two possible states ($b1, b2$) for node B of Fig. 1, and three possible states ($c1, c2, c3$) for node C and node A ($a1, a2, a3$), then the NPT of node A will include $3 * 2 * 3 = 18$ elements

$$P(A|B) \cdot P(B) = P(A,B), \tag{1}$$

$$P(B|A) \cdot P(A) = P(A,B), \tag{2}$$

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}. \tag{3}$$

The mathematic model on which Bayesian Networks are based is the theorem developed by the mathematician and theologian Thomas Bayes in the 18th century. The BN and the relevant theorem are based on the principle that conditional probabilities are more essential than joint probabilities. It is easier to define probability $P(A|B)$ without reference to the joint probability $P(A,B)$. The theorem of Bayes (3) is derived from Eqs. (1) and (2). In Eqs. (1)–(3), A is a given probability about an event, and B is additional information about it. $P(A|B)$ is known as *posterior probability*, i.e., probability of A after taking into consideration how the additional information B influences a certain event. $P(A)$ is called *prior probability* of A. $P(B|A)$ represents the probability of an event given a hypothesis A and considering that the hypothesis A is correct. $P(B)$ is independent from event B and can be considered as a normalisation factor. It is important to note that these probabilities are *hypothetical* and determine the confidence level of estimation under certain conditions. Thus the theorem is based on the previous analyses of the probabilities of these

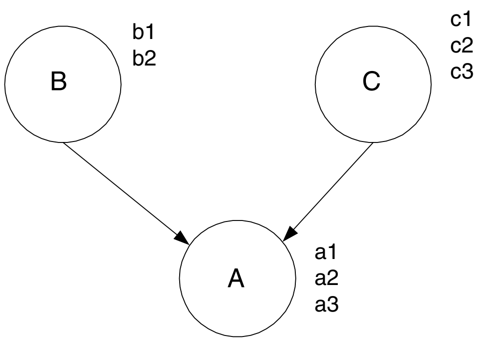


Fig. 1. A simple Bayesian Network.

conditions. Based on all the above, the BN can be used to monitor the changes of a case A in relation to a new event B. Namely, the future probability (posterior) $P(A|B)$ is calculated by multiplying the previous probability (prior) $P(A)$ by the probability (likelihood) $P(B|A)$, that is B will occur if A is true.

Building a Bayesian network to model a real-world situation is a two-step process including: (a) construction of a graph representing the qualitative influences of the situation modelled and (b) assignment of probability tables to each node in the graph. The conditional independence statements implied by the graph significantly simplify the elicitation of these probabilities. For instance, assuming that a student's performance in the final exams depends only on his/her preparation and his/her concentration during the exams. The graph of this (over simplified) example should comprise the parent nodes 'preparation' and 'concentration' and the child node 'exams'. For simplicity purposes, two states 'good' and 'poor' are assigned to both parent nodes and two states 'succeeded' and 'failed' to the child node. The graph of this example is shown in Fig. 2. The NPT table must be built on the basis of previous data or of the experience of the designer of the model, which is the case here, if no data exist. For example in the case that student preparation is 'good' and student concentration during the exams is also 'good', the designer assigns 85% probability of 'succeeded' and 15% of 'failed' in the exams. These probabilities must of course be tested during the application of the model and may change to reflect the real-world situation. After assigning all the values in the NPT table (in this case $2 * 2 * 2 = 8$ values) the model is ready for application. After assigning all values, the NPT table of the sample graph will look as Table 1.

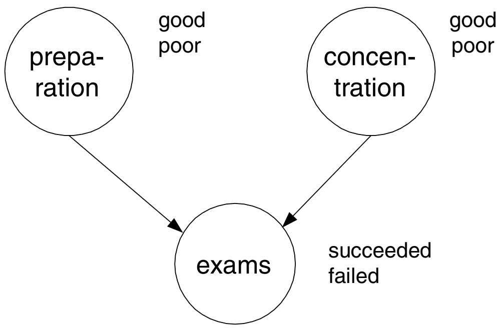


Fig. 2. Sample graph.

Table 1
Values of the NPT table

Parent nodes		Child node (exams)	
Preparation	Concentration	Good (%)	Poor (%)
Good	Good	85	15
Good	Poor	60	40
Poor	Good	30	70
Poor	Poor	5	95

Summarising, the Bayesian Networks enable the modelling of a sequence of probable events in combination to a set of conditions in order to test a set of hypotheses relating to an event. In cases where a number of alternative scenarios exist that can influence a certain event, the BN provide the opportunity to test each scenario and define then extent to which it will influence the particular event, as well as to analyse the consequences of each alternative on the final event. For this reason, the BN can be utilised for modelling individual differences of student behaviour in cases of distance education such as the case of INF10 in the HOU that is presented below.

4. The proposed method and model

4.1. The concept of the method and the model

The concept underlying the presented method is the creation of a continuously growing network that will accumulate the experience of all previous years, derived from data collection and analysis, as well as the estimations provided by the model itself. A model based on the Bayesian Networks is expected to change dynamically based on the input of its own results, i.e., its results can be used as feedback to improve the NPT and consequently the accuracy of the model. In the presented case, the model has been applied in the INF10 module of the Informatics course incorporating the experience of the past 2 years (out of the 3 years that the Informatics course has been running) and is expected to be applied in the entire Informatics course in the following years.

The model has been implemented using the *Microsoft © MSBNx Authoring and Evaluation Tool version 1.4.2*. The initial data of the probability tables was based on analyses of previous data collected from the course. This data provided a good starting point, since the empirical building of each table of probabilities can prove to be a hard task. However, even in the case of limited data, it is still possible to build the probability tables, but with a high level of uncertainty. In such a case, as already mentioned, the input of the model's own results can be used to improve the tables, although this is a rather lengthy process requiring a significant amount of time before reliable results can be obtained. On the other hand, faster fine-tuning of the model and accurate and reliable results can be obtained based on the collection and analysis of data from previous academic years.

Part of the BN of the presented model is shown in Fig. 3. The schema has been significantly simplified for presentation purposes, since the proper presentation of the entire BN of the model would require two A3 pages. The current version of the model (an *.xbn file that runs on MSBNx) can be made available on e-mail request. Specifically, Fig. 3 illustrates the final 'child' node representing the two possible states of a given student: the student either continues the studies or decides to drop out, and some 'parent' nodes – such as the use of computer, the hours of work, the use of the available web-based additional material, the use of the internet, etc – influencing this final node including the cross-relations between them. For the convenience of the reader, reference numbers have been assigned arbitrarily to all nodes of the schema, while all possible states of each node are presented. In order to give to the reader an idea of the number of parameters that can be taken into account thanks to the presented model it should be mentioned that, even though a simplified version of the model is illustrated in Fig. 3, for certain nodes (e.g., nodes No. 9 and 10) as many as $35 = 243$ values must be filled in the NPT; in other cases this number is much lower

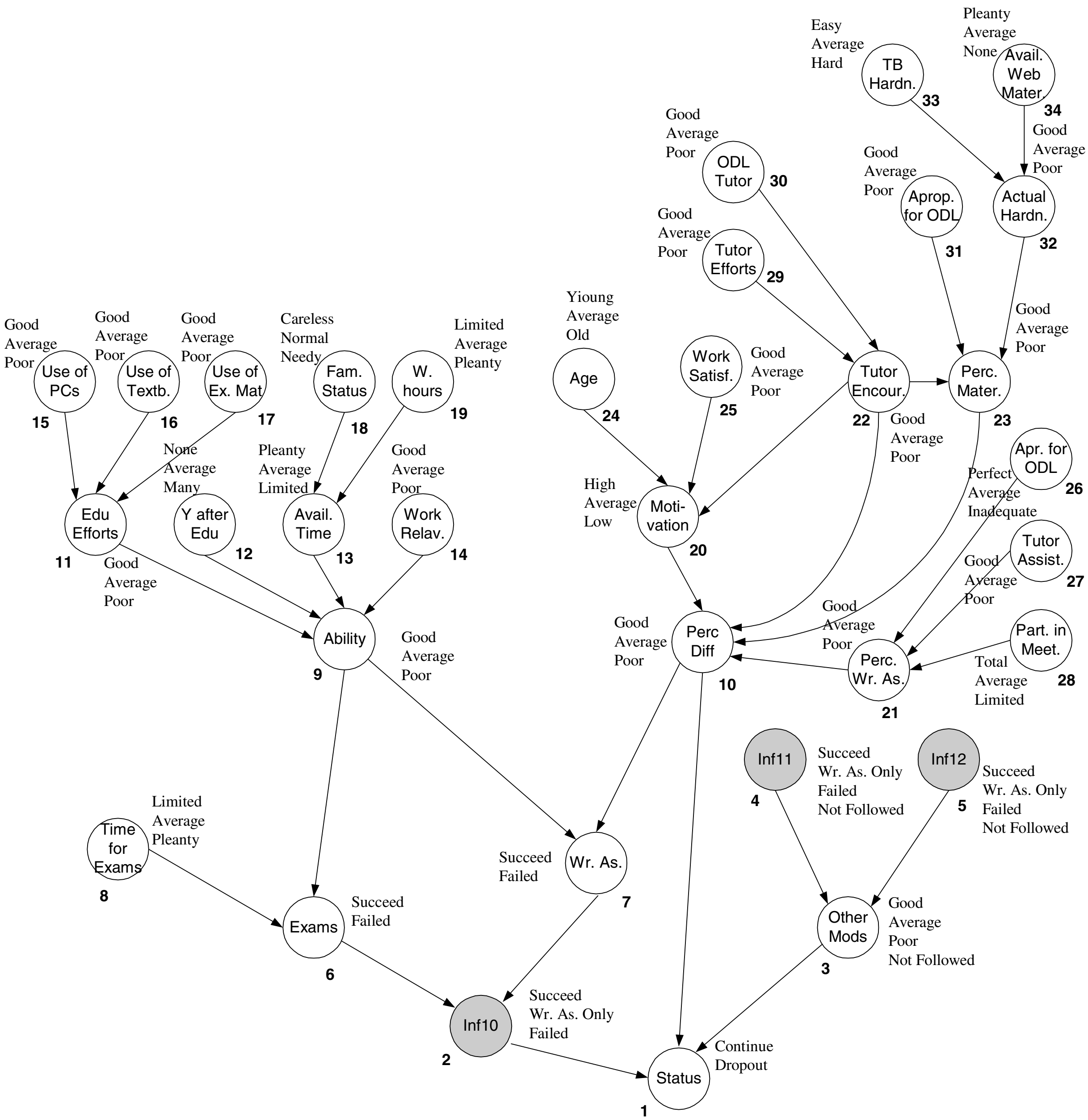
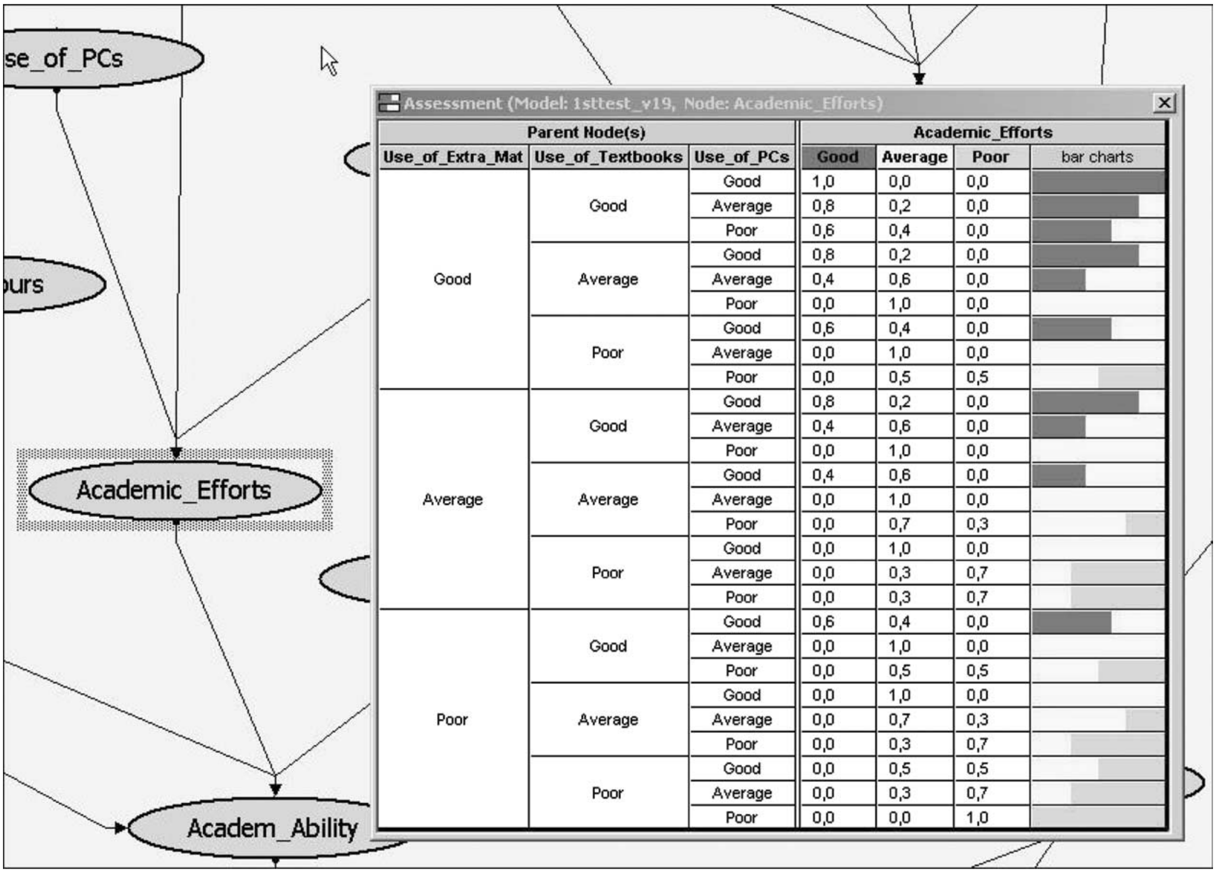


Fig. 3. Outline of the proposed model.

(e.g., 64 values for the NPT of node No. 3, 18 values for the NPT of node No. 6, etc.). Fig. 4 illustrates part of the interface of the MSBNx tool while used for the completion of the NPT of a medium to small size node, such as node No. 11 in Fig. 3.

As already mentioned, the presented model has the ability to 'learn' from past estimations and achieve more and more accurate results. This is exactly what makes it appropriate for constantly



Parent Node(s)			Academic_Efforts			
Use_of_Extra_Mat	Use_of_Textbooks	Use_of_PCs	Good	Average	Poor	bar charts
Good	Good	Good	1,0	0,0	0,0	
		Average	0,8	0,2	0,0	
		Poor	0,6	0,4	0,0	
	Average	Good	0,8	0,2	0,0	
		Average	0,4	0,6	0,0	
		Poor	0,0	1,0	0,0	
	Poor	Good	0,6	0,4	0,0	
		Average	0,0	1,0	0,0	
		Poor	0,0	0,5	0,5	
Average	Good	Good	0,8	0,2	0,0	
		Average	0,4	0,6	0,0	
		Poor	0,0	1,0	0,0	
	Average	Good	0,4	0,6	0,0	
		Average	0,0	1,0	0,0	
		Poor	0,0	0,7	0,3	
	Poor	Good	0,0	1,0	0,0	
		Average	0,0	0,3	0,7	
		Poor	0,0	0,3	0,7	
Poor	Good	Good	0,6	0,4	0,0	
		Average	0,0	1,0	0,0	
		Poor	0,0	0,5	0,5	
	Average	Good	0,0	1,0	0,0	
		Average	0,0	0,7	0,3	
		Poor	0,0	0,3	0,7	
	Poor	Good	0,0	0,5	0,5	
		Average	0,0	0,3	0,7	
		Poor	0,0	0,0	1,0	

Fig. 4. The Node Probability Table of node 11.

evolving educational environments, like the HOU. The currently used Bayesian Network is the 34th version of the model, utilising experiences gained over the past 3 years and results of all the previous versions.

In Fig. 3, the highlighted nodes Nos. 2, 4 and 5 correspond to student progress in the three 1st-year modules. Of these nodes, only the 'parent' nodes of node No. 2 are presented. In the extensive design of the model, nodes Nos. 4 and 5 are equally expanded. This is a characteristic example of the expandability of model since the probability tables for these nodes are based on the 'parent' nodes of node 2 (INF10 module); i.e., using the 'parent' nodes of node 2 it was very easy to design the 'parent' nodes of nodes 4 and 5 (not shown here). It should also be noted that a number of 'parent' nodes that further analyse nodes Nos. 15, 16, 17, 25, 27, 28, 29, 30, 31, 33, and 34 are not shown in Fig. 3, due to practical space limitations. An expansion of node 15 into 3 other nodes taken from the current version of the model is shown in Fig. 5. Most of these nodes were not included in the first versions of the model and were added in later versions to further expand certain nodes. This is typical for a BN that could start using few nodes and expand later on as more data and more accurate estimates became available.

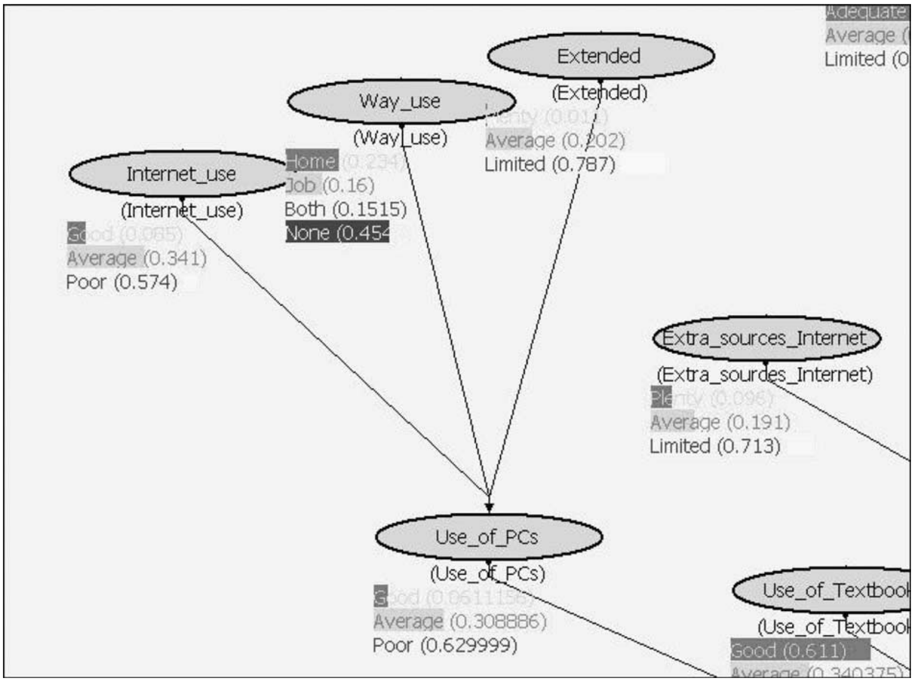


Fig. 5. Expanding node 15.

4.2. Application of the model

The application of the model is based on entering data in the form of defined states called *evidence*, for certain nodes. If no evidence is entered in the model, then the model provides estimation based on the previous experience of the three years that the Informatics course has been operating (it has been available since the academic year 2000–2001). Based on this experience, each student has approximately 72% probability to successfully continue his/her studies and 28% to drop out, which is shown in the final 'child' node. The import of evidence in the model activates the probabilities in the NPT of each relevant node and as result some parent estimations and correspondingly the final estimation change. In a rather simple case, if all we know about a student is that he registered in 2 modules (INF10 and INF11) and failed in INF10, while he successfully completed the written assignments of INF11 but failed in the respective final exams, then the probability of drop out changes from 28% to 32%. Of course, normally, there is more evidence to be imported in the model, thus affecting the final estimation accordingly, such as the use of computers at home or in the office, the use of the Internet, the use of the additional educational material available on the web sites, etc.

This type of model usage is called *forward*. In similar cases, the model's application offers prediction and the results can be used to prevent events from occurring. For example, if the dropout probability of a student is 45%, given that enough evidence has been imported in the model, then such a high percentage is a good indication that the tutor needs to adopt a different approach towards the specific student aiming at helping him overcome the difficulties he might be facing. Furthermore, the tutor may try to simulate (using the Bayesian Network) several approaches that he considers appropriate for this individual and select the combination that offers optimum results. To achieve this, the tutor needs to import his planned actions as evidence and monitor the model's predictions. On the other hand, if the model's estimations result in unusually high dropout rates for an entire group, then most likely a drastic change of the tutor's approach is required, or even the coordinator's intervention. One of the strong points of the model is that the probabilities in each NPT can be updated dynamically, even during the application of the model. This facility is called *propagation* and can be used to continuously improve the quality of estimations based on experience from previous application of the model.

The results derived from the *forward* application of the model (prediction) have been very satisfactory. Of course, like most automated tools, this model, too, should be regarded as an assisting tool and not as an oracle; one should always keep in mind that predictions are simply based on the data imported as evidence and on past experience in the form of NPTs. A limited number of cases in which the model failed for various reasons to predict changes in the student behaviour exist. For example, in one case a student dropped out although the probability estimated by the model was as low as 16.5%. The reason for this was her family's decision to move in the USA. Even in this case, however, 16.5% indicates that unanticipated events can always occur and cause dropouts. Fig. 6 presents an example of the model's application using the MSBNx tool. Evidence has been imported in certain nodes and, accordingly, a certain state in these nodes is considered as 'fact'. The probability of such state is equal to 1. In Fig. 6 this is illustrated by only

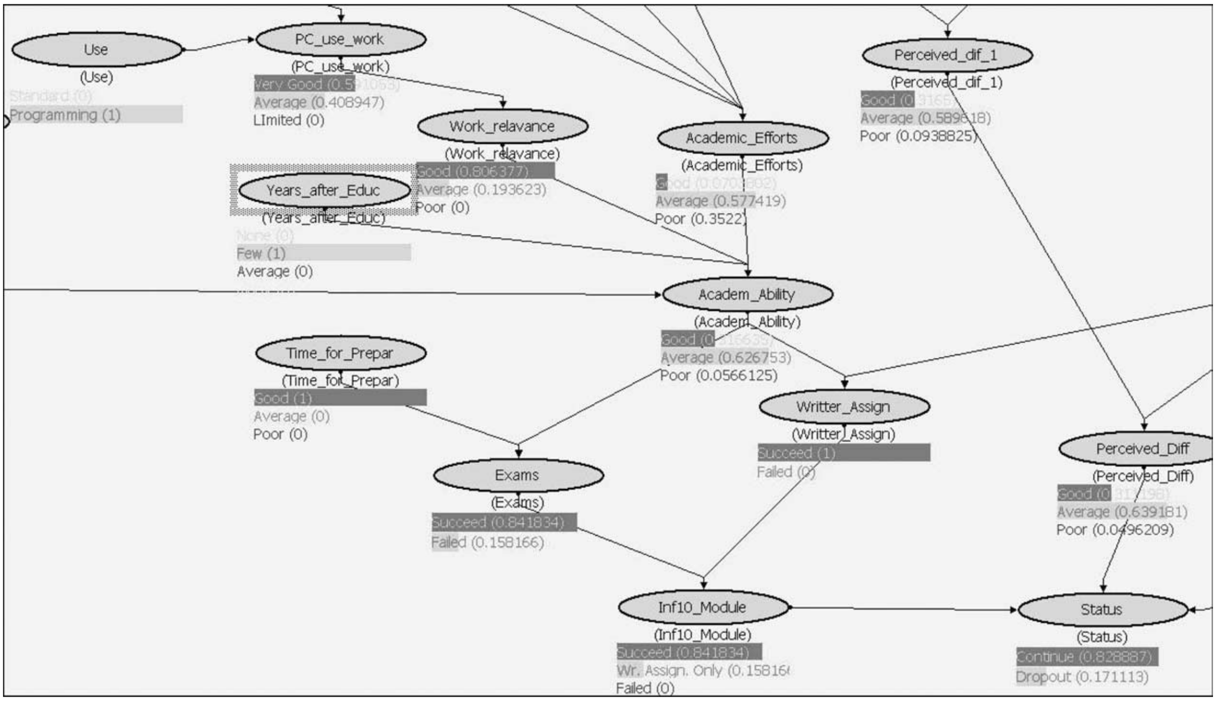


Fig. 6. Forward application of the model.

one highlighted state bar, as in the case of nodes Nos. 7 and 8 also present in Fig. 5 (the model illustrated in Fig. 6 is the full model, not the simplified version of Fig. 3, so more nodes appear as some nodes of Fig. 3 are further expanded). For the rest of the nodes shown in Fig. 6, estimates based on imported evidence are offered. Based on the inserted evidence, the probability of the child node for the state 'continue' has changed from 72% to 83%.

It must be noted that during the forward application of the model, predictions can be made not only for the final child node, but also for any non-parent node of the model, i.e., any of the intermediate nodes of the model. In this way prediction for a student's behaviour can be made based of the indications of intermediate nodes and contribute to changing the way in which the tutor approaches this student. For example in Fig. 7, evidences have been inserted into all parent nodes of the intermediate node named 'Perceived Diff'. This intermediate node represents the probability that the student is confident that he/she will successfully complete the particular module. As shown in Fig. 7, the student thinks that this module is extremely difficult and thus doubts that he can successfully complete it. The probability of the state 'good' is 0, of 'average' 60.6% and of poor

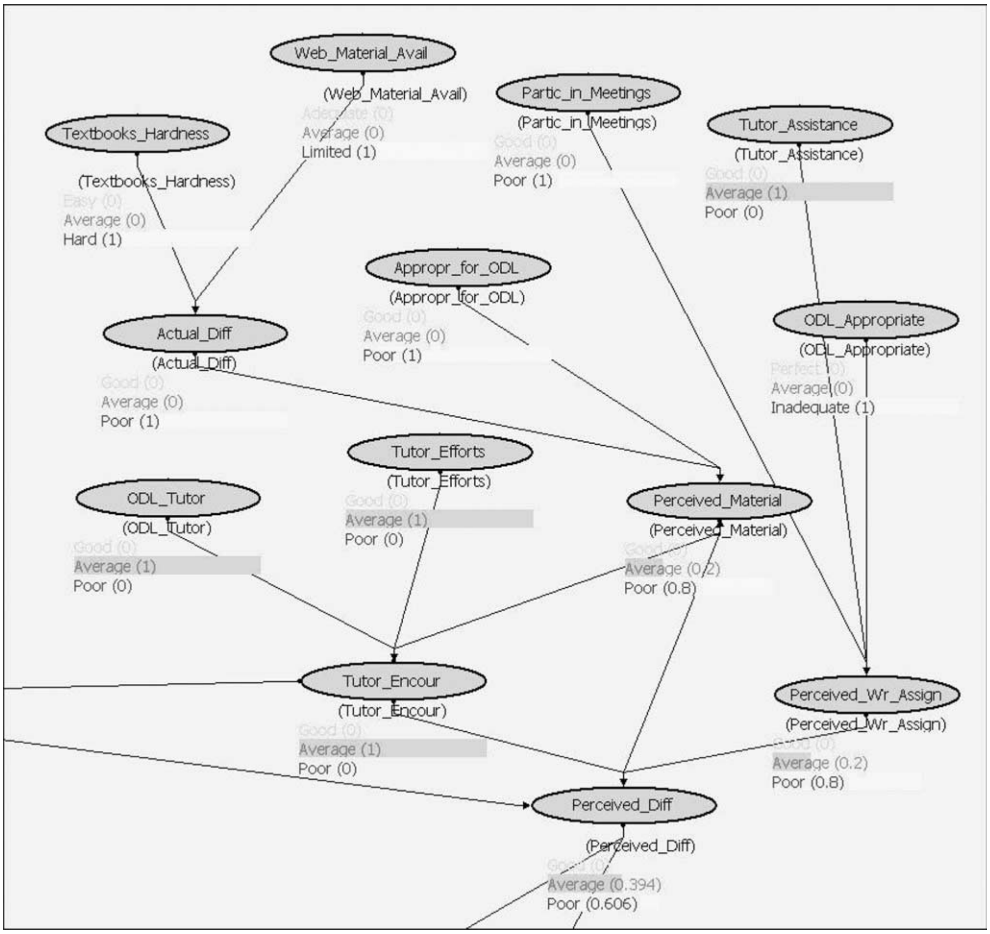


Fig. 7. Forward application of the model for an intermediate node.

39.4%. This indication is certainly worrying and should lead the tutor to take action so as to encourage the student and straighten his/her self-esteem. In this case the forward application of the model leads to a prediction related to an intermediate node (the one representing the student's perceived difficulty of the module) and not the final child node.

In the case that the model is used for the *assessment* of a given state that has already occurred, therefore there is evidence about this state, the model's use is called *backward*. For example, when it is known for a fact that an individual student has dropped out, the probability of dropout is equal to1. Assuming that evidence about certain initial nodes is known – for example whether this student uses a computer at work, number of working hours, if this student had registered in other modules, etc. – then one use these facts to make assessments about certain 'parent' nodes for which there is no evidence at all. For example, assessments can be made about the student's motivation for studying, the perceived difficulty concerning the written assignments, the encouragement received by the tutor, etc.

A particularity of the use of the model for *backward* assessment should be mentioned at this point: in some cases there is no objective way of confirming an assessment. For example, how can one objectively confirm the sate of the node 'motivation'? Nevertheless, in all cases where clues were available for confirming the results, the model exceeded even the most optimistic expectations. Fig. 8 presents the case of a student as an example of backward usage of the model. This student had successfully completed the written assignments, but failed in the examinations. These facts, along with other personal data (not all nodes are shown in Fig. 8) were imported as *evidence* in the corresponding nodes. Based on the results provided by the model for this particular case, it was highly probable that the specific student would not be sufficiently prepared for the final examinations. As illustrated in Fig. 8, the probability of 'good preparation' was only 0.109 (that is approximately 11%), when the corresponding percentage for other students that also failed in the examinations was not lower than 20%. This assessment based on the model's results was confirmed by means of a telephonic interview with the student. The interview revealed that during the month before the final examinations a member of the student's family faced a serious health problem and for this reason the student failed to prepare for the examinations.

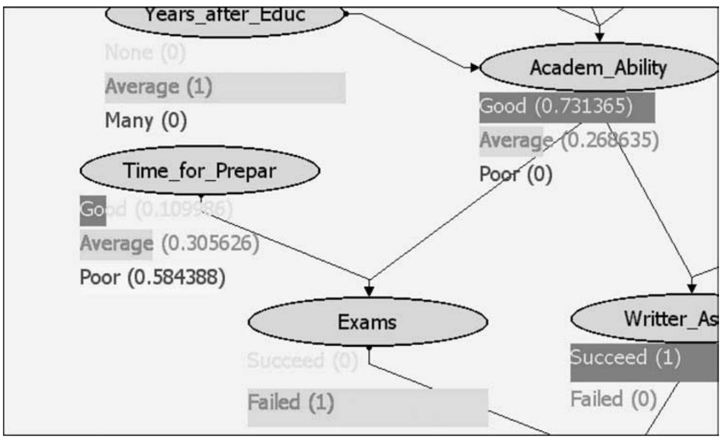


Fig. 8. An example of backward assessment.

5. Conclusions and future work

Before concluding, it should be pointed out again that the effectiveness of the model is closely related to the quality of the BN design and accuracy of values entered in the probability tables. Before accurate and useful results can be obtained, a collection of analytical data should be performed. Such data should then be used for the adaptation of the model to the particularities of each case, and the calculation of the NPT values. This implies paying significant efforts for implementing the model, but the real profit lays in the automation of estimates and assessments gained. In cases where the assessment or prediction should take into consideration very large amounts of data from a variety of sources, such as in the presented case, the use of modelling and automation is no longer simply 'profit'; it becomes a necessity. The presented model aims at helping towards this direction.

It was never in the author's intentions to develop a model that makes predictions and assessments by itself! The goal was to develop a method for modelling the educational experience of its designer and exploiting the past data or data resulting from its use. The model derived from using the presented method can aid in drawing conclusions based on the data that were incorporated into the model. The application of the model in one of the modules of the Informatics course offered very satisfactory results. Furthermore, the model can prove to be a valuable tool assisting in decision making in conditions of uncertainty, especially when a large number of parameters is involved. A future goal is to develop a model based on the presented method for the evaluation of the educational material and methods in the HOU. Finally, it must be noted that the main advantage of the proposed model is its modularity. This means that a limited network can be built and tested for a small part of a problem and then the same network can be expanded to include more parts of the same problem, while exploiting results from its use to improve the values in the tables of probabilities.

Acknowledgements

The author would like to thank the students, tutors and personnel of Hellenic Open University for their assistance in the research.

References

Jensen, F. (1996). *An introduction to Bayesian networks*. UCL Press.

Kerka, S. (1996). Distance learning, the internet and the world wide web, ERIC digests: ERIC document reproduction service no. ED395214.

Kschischang, F., Frey, B., & Loeliger, H. (2001). Factor graphs and the sum product algorithm. IEEE Transactions on Information Theory, February issue.

Lauritzen, S., & Spiegelhalter, J. (1990). Local computations with probabilities on graphical structures and their applications to expert systems. In G Shafer & J Pearl Morgan Kaufmann (Eds.), (pp. 415–448). California: San Mateo.

Lupo, D., & Erlich, Z. (2001). Computer literacy and applications via distance e-learning. *Computers & Education, 36*(1), 333–345.

Narasimharao, B. (2000). Issues in preparing Open University learners for Open University system. Available: http://www.cemca.org/ignou-icde/paper23.html.

Palmer, S. (2000). On- and off-campus computer usage in engineering education. *Computers & Education, 34*(1), 141–154.

Parker, A. (1999). A study of variables that predict dropout from distance education. *International Journal of Educational Technology, 1*(2), 1–10.

Rumble, G. (1992). *The management of distance learning systems*. Paris: UNESCO/IIEP.

Shin, N., & Kim, J. (1999). An exploration of learner progress and drop-out in Korea National Open University. *Distance Education – An International Journal, 20*(1).

Xenos, M., Pierrakeas, C., & Pintelas, P. (2002). A survey on student dropout rates and dropout causes concerning the students in the Course of Informatics of the Hellenic Open University. *Computers & Education International Journal, 39*, 361–377.