A BAYESIAN APPROACH TO PREDICT PERFORMANCE OF A STUDENT (BAPPS): A Case with Ethiopian Students

Rahel Bekele
&
Wolfgang Menzel

Department of Informatik, University of Hamburg
Vogt-kölln-Str. 30, 22527 Hamburg
Germany
Bekele@informatik.uni-hamburg.de
Menzel@informatik.uni-hamburg.de

ABSTRACT

The importance of accurate estimation of student's future performance is essential in order to provide the student with adequate assistance in the learning process. To this end, this research aimed at investigating the use of Bayesian networks for predicting performance of a student, based on values of some identified attributes. We presented empirical experiments on the prediction of performance with a data set of high school students containing 8 attributes. The paper demonstrates an application of the Bayesian approach in the field of education and shows that the Bayesian network classifier has a potential to be used as a tool for prediction of student performance.

KEY WORDS
Bayesian approach, Modelling, machine learning, Student Performance, Stochastic Modelling.

1. Introduction

1.1. Statement of Purpose

Accurate prediction of student performance is helpful in order to provide a student with the necessary assistance in the learning process. However, in a manual set up, it is usually difficult to come up with rule sets that are needed to predict performance. Hence we need to look for readily capable methods for dealing with the task.

This research has looked specifically at personal, social and cultural features that may be used in automatic prediction of performance. In this connection, there have been a number of research studies published in the literature[1-5], which attempt to explain features and related factors for low performance of students. They identify elements that intervene with educational goals namely: teaching/learning strategy (system causal factor), parents (family causal factors), teachers (academic causal factor), and students (personal causal factor). In Ethiopia, for instance, some factors for low performance of students that are related to these causal factors are: difficulty to understand the English language which is the medium of instruction[6], individual personalities such as lack of motivation, introvert personalities and lack of self confidence which results in fear of students to ask questions in class, [6,7,8,9], and lack of assistance by instructors to students due mainly to heavy teaching loads [10].

It is obviously impossible to know everything about a student. Therefore, it becomes necessary to choose the most relevant and useful information about a student that may influence performance.

From the foregoing, the following research questions provided the specific focus of our study. (i) What are the major/strong attributes that may intervene with performance? (ii) What is the technique for acquiring values for those identified attributes? (iii) How can we automatically predict performance based on those values?.

Practically, it is often difficult to model the exact values of attributes of a student with respect to some attributes. Thus, the procedure may include too much uncertainty. If those attributes are uncertain, then this uncertainty will transfer to the prediction, which may also result in poorly adapted classification of performance. The Bayesian approach was employed to tackle this problem since it was found to be a clear and manageable language for expressing what we are certain and uncertain about.[11]

With the goal of predicting performance, the specific aims of the research work were (i) to identify major attributes which help to effectively predict performance; (ii) to develop an instrument for measuring those attributes; (iii) to investigate the application of belief network tools in classifying students into categories of performance.

This paper makes four main contributions. (i) it demonstrates that Belief networks can be applied to predict student performance. (ii) it demonstrates a new application of the Bayesian approach in the field of education; (iii) the output of the study may be used as a benchmark for educationalists and others interested in the area to predict student performance in various subjects (iv) the outcomes of the research are helpful to further extend the functionality of the student model component of an intelligent tutoring system. i.e. the student model may additionally contain aspects of the social and personal attributes as well as the predicted performance of the student.

Subsequent sections of this paper contain a brief overview of the Bayesian network classifiers, related works methodology adopted, our experimental work, conclusions and future work.

1.2. Bayesian Network Classifiers

1.2.1. General

One of the fundamental problems in machine learning, data analysis, and pattern recognition is classification of observed instances into predetermined categories of classes. For example, students could be classified into categories according to their performance. In this case the categories will be Above Satisfactory, Below Satisfactory and Satisfactory.

Classification problems have been widely studied in statistics and artificial intelligence(AI) and a variety of different classification approaches have been developed. Some of the most popular approaches used in AI include decision trees,[13], neural networks [14], genetic algorithms [15] and Bayesian network classifiers [16]. The theories and concepts of Bayesian Networks were invented by Judea Pearl in the 1980s [17].

1.2.2. Bayesian Networks

General Bayesian network classifiers are known as Bayesian networks, belief networks or causal probabilistic networks. [17,18]. They draw their roots from a branch of probability and statistics known as decision theory[19], which involves the theory of how to minimize risk and loss when making decisions based on uncertain information. Moreover, given that quite often data can not be classified with deterministic correct certainty, and associated with every classification problem is a risk/loss function that indicates the severity of an incorrect classification, Bayesian learning involves the process of calculating the most probable hypothesis that would correctly classify an object or piece of data, based on Baye's rule. Some attractive aspects of Bayesian learning include: each training vector can be used to update probability distributions which in turn affect the probability that a given hypothesis is true; provides more flexibility in that a hypothesis does not get completely ruled out from few examples; and prior knowledge can be easily implemented in the form of prior probability distributions[19].

The structure of a Bayesian network is a graphical illustration of the interactions among the set of variables that it models. It consists of a directed acyclic graph and conditional probability distributions associated with the vertices of the graph. The directed acyclic graph represents the structure of the application domain. Nodes which are usually drawn as circles or ovals, represent random variables and arcs represent direct probabilistic dependencies among them. [20,21]. With every vertex is associated a table of conditional probabilities of the vertex given each state of its parents. We denote the conditional probability table using the notation $P(x_i|par(x_i))$, where lower case x_i denotes values of the corresponding random variable X_i and $par(x_i)$ denotes a state of the parents of X_i. The graph together with the conditional probability tables define the joint probability distribution contained in the data.

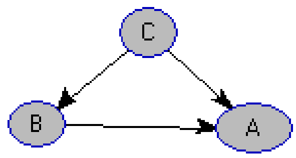

Figure 1: A Bayesian Network

Using the probabilistic chain rule, the joint distribution can be written in the product form:

$$P(x_1,x_2,x_3,\ldots) = \prod P(x_i|par(x_i))$$

Where the product goes from i=1 upto n and n is the number of vertices in the graph.

An example of a simple Bayesian network is given in figure 1. The corresponding joint probability distribution for the figure can be written in the form :

$$P(a,b,c) = P(a|b,c)\ P(b|c)P(c).$$

In a Bayesian network all variables are treated in the same way and any one can be regarded as the class variable Classification. A Bayesian network classifier involves performing probabilistic inference on the Bayesian network using one of the available probabilistic inference algorithms. [22,23,24].

1.3. Related Works

During the past decade, Bayesian networks have gained popularity in AI as a means of representing and reasoning with uncertain knowledge. To this end, there is a good deal of research work in the application of the Bayesian network[25-30].

More domain specific works have focused on probabilistic student models. Andes Intelligent Tutoring System for Physics [31] uses a belief network to represent alternate plans that may be used to solve physics problems. Student actions are analyzed to update the probabilities of the respective plans. Vanlehn[32,33] presented an On-Line assessment of Expertise(OLAE) that collects data from student solving problems in introductory college physics and analyzes the data with probabilistic methods that determine what knowledge the student is using and presents the results of the analysis. For each problem, the system automatically creates a Bayesian net that relates knowledge represented as first-order rules, to particular actions, such as written questions. Using the resulting Bayesian network, OLAE observes a student's behavior and computes the probabilities of the level of knowledge of the student and accurate use of rules.

The research presented by Murray[34] inferred a student model from performance data using a Bayesian belief network. The belief network modeled the relationship between knowledge and performance for either test items or task actions. The measure of how well a student knows a skill is represented as a probability distribution over skill levels. Questions or expected actions are classified according to the same categories by the expected difficulty of answering them correctly or selecting the correct action.

The research presented in this paper is different from what is already presented in the literature, in that it does not use the domain knowledge to determine the level of performance, rather it uses other social and personal attributes to predict performance by employing Bayesian network modeling technique.

2. Methodology

2.1. Identification and Measurement of Attributes

For the purpose of our experiment, mathematics was taken as the subject area. This was mainly because of familiarity of the researchers with the subject.

In order to select the relevant attributes which predict performance, a list of potential attributes were identified from discussion with colleagues, expert opinion and review of literature. These attributes were then distributed to about 40 individuals to collect their opinion on the relevancy of each. Based on the collected information, the eight selected attributes for the purpose of this experiment were gender, group work attitude, interest for mathematics, achievement motivation, self confidence , shyness, English performance and mathematics performance..

The test subjects were students of one senior high school in Addis Ababa who were in their last year of preparatory program to join higher learning institutions. Data on gender, English performance and mathematics performance were collected from record office of the school. Questionnaires for measuring the other attributes were developed in consultation with psychologists and from the available literature. Two consecutive pilot surveys containing 96 items were carried out before the final questionnaire was developed. Likert scale[35] was adopted to register the extent of agreement or disagreement with a particular item in the questionnaire. In the second pilot survey, the items for measuring each attribute were found to be reliable with coefficient of alpha > 0.77. In the final questionnaire, there were a total of 75 items included where each of the attributes was measured by 15 items Some lie detector statements were also included in order to check seriousness of the students in filling out the questionnaire.

The minimum sample size required was determined with respect to the probability of the rare event. For example, getting a high value for all the personality values is considered to be a rare event. From the pilot survey, the probability of getting a student with high achievement motivation was the least (4/64). Therefore, this probability was used to calculate the minimum sample size, which was 504. A total of 571 data records were collected in the final survey. Before the records were used for the experiment a careful examination of each questionnaire was done. Questionnaires which had more than half of the lie detector statements answered positively were disregarded. A total of 514 data records were finally used for the experiment.

2.2. Data Preparation

The bulk of the effort was invested in assembling and integrating the data and in preparing distinct files for training dataset and test dataset. Typographical errors in the data were avoided because each value of the attribute was generated using a statistical package.

For the purpose of testing the applicability of five scale options for making the possible categories of the identified variables, the mean within three standard deviation and, two standard deviation were calculated. In both cases the number of students at the extremes were found insignificant or almost null. Thus a three scale option (the mean within one standard deviation) has been considered.

The possible categories(outcomes) given for each variable were, therefore, gender(male, female); group work attitude(positive, indifferent, negative); interest for math(interested, indifferent, uninterested); achievement motivation (high, medium, low); self confidence (high, medium, low); shyness (extrovert, medium, introvert); English performance (above satisfactory, satisfactory, below satisfactory) and mathematics performance (above satisfactory, satisfactory, below satisfactory). Values of the attributes. above mean plus standard deviation ($>\bar{x}+s$) were put in high valued category, between mean minus standard deviation and mean plus standard deviation[$\bar{x}-s$, $\bar{x}+s$] were put in the average category and below mean minus standard deviation ($<\bar{x}-s$) were put in the low valued category.

Based on these category values, the attributes in each data record were changed to their respective qualitative information. The observed 514 data records were ordered in such a way that math performance is a function of gender, group work attitude, interest for mathematics, achievement motivation, self confidence, shyness and English performance.

An attempt was made to extract the confusions in the actual data records. This means the data records were extracted which have the same values for all the seven attributes but a different category for math performance. A total of 54 such errors were observed, which makes 10%. Thus the maximum degree of accuracy in predicting performance of a student was 90%.

3. Experimental work and results

3.1. Belief Network Modeling

The belief network modeling software employed for the purpose of the experiment was the Bayesian Network in Java software package [36].

A percentage split was used to partition the dataset into training and test data. Since experimental results may be influenced by the selection of the test and training sets, several experiments were carried out by splitting the data into 2, 3, 5 and then 10 partitions. For instance, if we take the data partition of 3, each partition in turn was used for testing while the remainder was used for training. This process repeats three times and at the end, every instance has been used exactly once for testing. Finally, the average result of the 3 fold cross validation was considered. The following illustrates one of the Bayesian networks learned during the experiment.

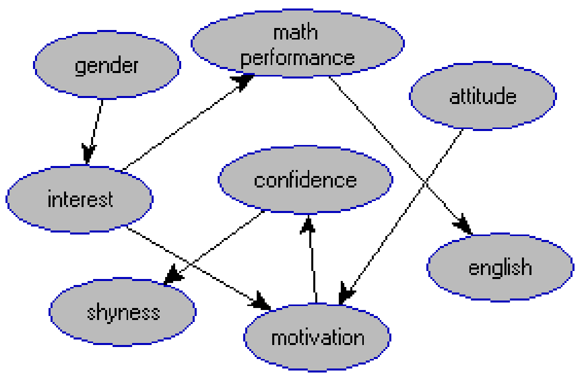

Figure 2 : Belief network modeling

Each node was described by a probability distribution conditional on its direct predecessors. Nodes with no predecessors are described by prior probability distributions. For example, node attitude in the network was described by the prior probability distribution over its three outcomes: Positive, Indifferent and Negative. The other nodes were described by a probability distribution over their outcomes (eg. Above Satisfactory, Satisfactory, Below Satisfactory for math performance) conditional on the outcomes of their predecessor. By observing values of attributes, captured from the model, we can compute the probability of performance .i.e

P(math performance | gender, group work attitude, interest for math, achievement motivation, self confidence, shyness, English performance).

During the experiment, in order to test the performance of each of the learned network, first the test data records were saved as text file without the value of the math performance. This file was then used to see the prediction performance of the model. A Java program was written to parse each of the test data records into the respective values of gender, group work attitude, interest for math, achievement motivation, self confidence, shyness and English performance. The classes and methods in the Bayesian network tool in Java particularly, the Logic Sampling algorithm, was used for the purpose of assessing the prediction performance. Once the program loaded the corresponding learned network, the parsed information of each record was fed into the graph as evidence values. The program after consulting the model based on the evidence values, evaluated the probability values assigned for the three categories of math performance. It then took the performance category having the maximum probability value. The program then wrote this value for the corresponding record.

Upon completion of the prediction, the program compared the predicted performance category with that of the observed categories in the original file. For instance, an example of the confusion matrix that resulted from the 3 fold cross validation is shown below.

	classified as			
	Below Satisfactory	Satisfactory	Above Satisfactory	Total No.
Below_	83	19	5	107
Satisfactory	23	14	4	41
Above_	4	7	13	24

Table 1: Output of the 3 fold cross validation.

The confusion matrix depicts that out of the total records provided to the program, about 64% of the records were classified correctly. A closer look showed us that 78% of the below satisfactory performance category was correctly classified, only 34% of the satisfactory performance category was correctly classified and 54.2% of the above satisfactory performance Category is correctly classified. Students' negligence to show their true feelings for some of the items in the questionnaire and the sample size used may be attributed to the prediction errors.

Moreover, as observed in the confusion matrix, the chance of classification of a below performance category student into a high performer, which may be the most dangerous in the learning process is only 0.046.

In relation to the relevance of the attributes to the prediction of performance, the Random search algorithm in the Weka data mining tool[37] revealed that gender, interest for math and English performance are the most relevant for prediction of performance while all the other search algorithms revealed that only interest for math and English performance are the most relevant attributes.

3.2. Individual Performance Prediction

The best learned network which yielded the minimum prediction error was used to classify a new student to one of the three math performance categories. A java program was written in order to automatically analyze the values of the seven attributes and consult the Bayesian network graph to predict the category of performance. The 75 questions from the already developed questionnaire items were entered in the computer. For the purpose of qualifying the numerical values of the attributes, the summary values (mean and standard deviation) obtained from the empirical data were also used in the program. The features of the interface developed are described below.

When a new student starts the system, he/she will be asked to enter name, id, gender and English fluency as depicted in the following screen shot. (figure 3).

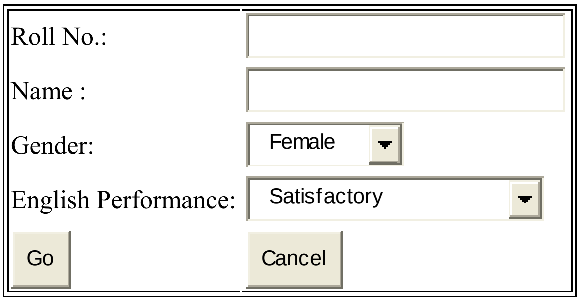

Figure 3: student enters here general information

After the student fills in those data, another screen is displayed (figure 4) where the student fills in his/her extent of agreement for statements to the 75 statements..

From the question answering session of the program, the response of the students for each item was scored by the program. Based on the scores of each item, the system automatically calculates the values of each of the attributes.

The program then consults the belief network tool for the probability of the student having above satisfactory, below satisfactory or satisfactory performance. The system takes the category with the higher probability and stores the information along with values of the other attributes.

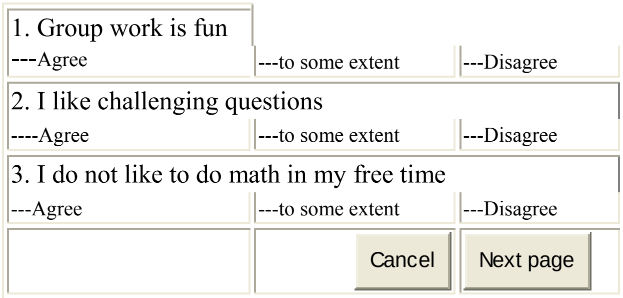

Figure 4 : student indicates here extent of agreement

4. Conclusions and Future Work

In examining the problem of prediction of performance, we have found that it is possible to automatically predict students' performance. Moreover by using an extensible classification formalism such as Bayesian networks, it becomes possible to easily and uniformly integrate such knowledge into the learning task. Our experiments also show the need for methods aimed at predicting performance and exploring more learning algorithms.

We plan to test the prediction performance of the model in a real world experiment where students are given tests and to compare their performance against the prediction of our system. It is believed that, if put to practice, this individualized performance prediction will help the teacher a lot in giving the necessary assistance to a student.

Further experiments are also being carried out to recommend student clusters based on the predicted performance. While the existing clustering algorithms are based on similarity checking, we plan to explore on difference checking so that automatic composition of groups with heterogeneous nature will be possible.

Finally we are also interested in including performance prediction into the student model component of an Intelligent tutoring system, i.e we plan to experiment on extending the functionality of the student model so that it is able to propose clusters of students by using an incremental algorithm that does not require the whole data set in advance.

5. Acknowledgement

We would like to thank Professor Dr. Christiane Floyd (University of Hamburg) and Professor Dr. Darge Wole (Addis Ababa University) for their unreserved supervision and assistance during the course of this work. Our appreciation also goes to the instructors of Yekatit 12 senior secondary school in Addis Ababa who cooperated in facilitating matters and to the students for filling out the rather lengthy questionnaire.

6. References

[1] Atkinson, E., An investigation into the relationship between teacher motivation and pupil motivation, *Educational Psychology, 20*(1), 2000,45-57.

[2] Caplan, S. et al, Socio emotional factor contributing to adjustment among early entrance to college students, *Gifted Child Quarterly, 46*(2), 2002, 124-134.

[3] Antonio Lozano Diaz, Personal, family, and academic factors affecting low achievement in secondary school, *Electronic Journal of Research in Educational Psychology and Psycho Pedagogy. 1*(1),2003, 43-66.

[4] Georgiou, S. et al., Teachers attributions of student failure and teacher behavior toward the failing student, *Psychology in the Schools, 39*(5), 2002, 583-596.

[5] Gonzalez, J.A. et al,. A structural equation model of parental involvement, motivational and aptitudinal characteristics, and academic Achievement, *Journal of Experimental Education,70*(3),2002,257-287.

[6] Zaudneh Y., Darge W., and Nardos A., *A survey of the teaching – learning situation in institutions of higher learning in Ethiopia*, 1989, (unpublished document)

[7] Darge W. *The reactions of social sciences first year students in Addis Ababa University to moral dilemmas related to Academic Matters*, 1989, (unpublished document).

[8] Huitt, W., Individual differences. *Educational Psychology Interactive,* Valdosta, GA: Valdosta State University,1997,[retrieved on 13 Sept, 2004 from http://chiron.valdosta.edu/whuitt/col/instruct/indiff.html

[9]http://hubel.sfasu.edu/courseinfo/TS/indivdiff.html [retrieved on 11 Sept. 2004].

[10] Tekeste Negash, *The crisis of Ethiopian education: some implications for nation building*(Uppsala:Uppsala University,1990).

[11] Fung, Robert & Kuo-Chu Chang, Weighting and integrating evidence for stochastic simulation in Bayesian networks, *Proc. 5th Conf. on Uncertainty in Artificial Intelligence*, Elsevier Science Publishing Company, Inc., New York, NY, 1990, 209-219.

[12] B. Stewart, Predicting project delivery rates using the Naïve-Bayes Classifier, *Journal of Software Maintenance and Evolution: Research and Practice. Vol.14.*2002. 161-179.

[13] Breiman L, et al *Classification and regression trees* (Wadsworth: Belmont CA, 1984)

[14]Bishop, CM. *Neural networks for pattern recognition* (Oxford University Press: New York NY, 1995).

[15]Goldberg DE. *Genetic algorithms in search, optimization and machine learning* (Addison Wesley: Reading MA, 1989).

[16]Friedman N, Geiger D, Goldszmidt M., Bayesian network classifiers, *Machine Learning. 29*(2-3): 1997,131-163.

[17] Pearl J., *Probabilistic reasoning in intelligent systems: networks of plausible inference*, (Morgan Kaufmann: San Mateo CA, 1988).

[18] F.V. Jensen., *An introduction to Bayesian network* (London. U.K: University College London Press, 1996).

[19] F.V. Jensen, Bayesian network basics, *AISB Quarterly: Vol 94*, 1996, 9-22.

[20] Henrion, Max., Some practical issues in constructing belief networks, *Proc. 3rd Conf. on Uncertainty in Artificial Intelligence*, Elsevier Science Publishing Company, Inc., New York, NY, 1987, 161-173.

[21] Cheng, J., Bell, D.A. and Liu, W., An algorithm for Bayesian belief network construction from data. *Proc. 6th International Workshop on Artificial Intelligence and Statistics*, Florida, 1997a, 83-90.

[22] MacKay, David J. C., *Information theory,inference and learning algorithms* (United Kingdom: Cambridge University Press, 2003).

[23] Pearl, Judea, *Causality : Models, Reasoning, and Inference* (United Kingdom: Cambridge University Press, 2000).

[24] Lauritzen, Steffen L. & David J. Spiegelhalter, Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of the Royal Statistical Society, Series B (Methodological), 50*(2), 1988. 157-224

[25] Spiegelhalter, David J. Incorporating Bayesian Ideas into Health-Care Evaluation, *Journal of Statistical Science, 19* (1),2004, 156-174.

[26] Gale, Douglas and Kariv, Shachar, *Bayesian learning in social networks*, 2003 [retrieved on 30th Oct., 2004 from http://ist-socrates.berkeley.edu/~kariv/Gale_Kariv_I. pdf]

[27] Collins, J.A. et al. Adaptive assessment using granularity hierarchies and Bayesian nets, *Proc. 3rd International Conference on Intelligent Tutoring Systems.* Montreal, Canada. 1996, 578-586.

[28] Jensen, Finn V. et al, Bayesian updating in recursive graphical models by local computations. *Computational Statistics Quarterly, 4,*1990, 269-282.

[29] Olga Goubanova, Predicting segmental duration using Bayesian belief networks, [Retrieved on 9 Sept. 2004 from **Fehler! Hyperlink-Referenz ungültig.**]

[30] Sahami, Mehran et al, A Bayesian approach to filtering junk e_mail,[Retrieved on 12 Sept. 2004 from http://research.microsoft.com/~horvitz junkfilter.htm].

[31] VanLehn, Kurt et al., Minimally invasive tutoring of complex physics problem solving. *Proc. 5th International Conf. on Intelligent Tutoring Systems,* 2002,367-376.

[32]Conati, C. & VanLehn, K., A student modeling framework for probabilistic on-line assessment of problem solving performance, *Proc. 5th International Conference on User Modeling,* Kailua-Kona, HI: User Modeling Inc., 1996, 75-82.

[33] VanLehn, Kurt & Martin Joel, Student assessment using Bayesian nets., *International Journal of Human Computer Studies . 42*,1995,575-591.

[34]Murray, William RA, Practical Approach to Bayesian Student Modeling, *Proc. 4th International Conference on Intelligent Tutoring System,* 1998, 424-433.

[35] Barnett, V. *Sample Survey principles and methods.* (Hodder publisher,1991)

[36] http://sourceforge.net/projects/bnj/

[37] http://www.cs.waikato.ac.nz/~ml/weka/