Implementation of Intelligent Tutoring System Using the Data Mining Approaches

Ohn Mar Myint, Swe Zin Hlaing
University of Computer Studies, Yangon
ohnnmarr.1@gmail.com

Abstract

Web-based tutoring systems are increasingly popular due to their appeal over traditional paper-based textbooks. Whereas many tutoring systems are static HTML, the Bayesian intelligent tutoring system can help a student navigate through the online course materials, recommend learning goals, and generate appropriate reading sequences. This paper presents a Bayesian intelligent tutoring system for computer programming, called BITS. The decision making process of this system is guided by Bayesian networks, which are a formal framework for uncertainty management in Artificial Intelligence based on probability theory. The primary function of an agent is to help the user better use, manage and interact with the computer applications. Intelligent agents perform tasks for students as a dynamic, personal, and smart learning environment. In Intelligent Tutoring System, the student model has the ability of recording information on students. With this information, teaching and learning can proceed in a variety of ways based on their needs and interest.

1. Introduction

Web courseware is easily accessible and offers greater flexibility through the internet, that is, students can control their own place of study and do not depend on a teacher's presence and rigid classroom schedules. Unlike printed textbooks, Web-based tutoring systems can incorporate multi-media such as audio and video to make a point. However, since many current Web-based tutoring systems are static HTML Web pages, they suffer from two major shortcomings; namely, they are neither interactive nor adaptive. Most Web courses present the same static learning materials to students with widely differing knowledge levels of a given subject. Therefore, this kind of system is unable to satisfy the heterogeneous needs of many students. In the context of Web-based tutoring, an intelligent system is developed for assisting a user in solving a problem. Obviously, this involves creating systems that can make decisions based on uncertain or incomplete information. One formal framework for uncertainty management is Bayesian networks, which utilize probability theory as a formal framework for uncertainty management in Artificial Intelligence.

The purpose of this paper is to put forth a Web-based intelligent tutoring system using Bayesian networks, called BITS, to support students in learning computer programming. The decision making process conducted in this intelligent system is guided by Bayesian networks. BITS can assist a student in navigation through the online materials. BITS can recommend learning goals, and generate appropriate reading sequences. In this paper, prediction is implemented with the straight-line regression using the method of least squares. This prediction model is simple and provides good approximations. The method of least squares estimates the best-fitting straight line as the one that minimizes the error between the actual data and the estimate of the line.

Intelligent tutoring system (ITS) represents an important class of educational technology poised to play particularly critical role help students acquire the skills needed to succeed. Intelligent Tutoring Systems (ITSs) are very similar to human tutors, can able to learn programming skills and knowledge more quickly and effectively than students in traditional education settings. [13] The courses are classified intelligently for students as an intelligent agent.

In Section 2, the related work of Bayesian Theorem and Intelligent tutoring System describes. The Section 3 explains the Intelligent Tutoring System, Intelligent Agents, Bayesian Classification, Bayes' Theorem, Prediction and Linear Regression. Proposed System Design and overview of system design presents in Section 4. The Section 5 includes the predicting a class label using Naïve Bayesian Classification, ROC curves for Naïve Bayes', Classifier Accuracy Measures and Prediction of post-test mark using the straight-line regression using the method of least square. Conclusions and future work explain in Section 6.

2. Related Work

Bayesian approach has been one of the major AI approaches utilized by the researchers. Bidgoli et al.[4] have compared six classifiers (quadratic Bayesian Classifier, 1-nearest neighbors, k-nearest neighbors, Parson Window, feed-forward, neural network, and decision tree) to predict the course final results from a learning system log data. The data contained attributes concerning each task solved and other actions like participating in the communication mechanism and reading support material. The data set contained 250 students. The best classifier, k-nearest neighbors, achieved over 80% accuracy, when the final results had only two classes (pass/fail). C.J. Pierrakeas [1] was presented distance learning using machine learning techniques in proceedings of 7th International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES).

Villano [8] was applied probabilistic student models using Bayesian belief networks and knowledge space theory. Maguire [10] was implemented a web-based Bayesian intelligent tutoring System for computer programming. The decision making of this paper was guided by a Bayesian network approach to support students in learning computer programming. This paper was discussed how to employ Bayesian networks as an inference engine to guide the student'' learning processes. These was taught only C++ programming course for students.

Web Intelligence is a direction for scientific research that explores practical applications of Artificial Intelligence to the next generation of Web-empowered systems. Yao argue that a system should be robust enough to deal with various types of users. Moreover, Brusilovsky and Maybury [9] explicitly state that the solution needed to fix the problem of the traditional "one-size-fits-all" approach is to develop systems with an ability to adapt their behavior to the goals, tasks, interests and other features of individual users and group of users.

3. Background Theory

Intelligent Agents are essentially agents that can do reasoning or planning. [1] Intelligent tutoring system is a computer-based instructional system with models of instructional contents that specify what to teach, and teaching strategies that specify how to teach. The Intelligent Tutoring Systems are instructional systems that use Artificial Intelligence (AI) techniques in computer programs for facilitate learning. [6] The student model in an Intelligent Tutoring System has the capability of recording information on students. [7]

3.1 Classification

In data classification process, a model is built describing a predetermined set of data classes or concepts. The model is constructed by analyzing database tuples described by attributes. Each tuple is assumed to belong to a predefined class as determined by one of the attributes, called the class label attribute. The data tuples analyzed to build the model collectively form the training data set. The individual tuples making up the training set are referred to as training samples and are randomly selected from the sample population. [2]

3.1.1 Bayesian Classification

Bayesian classifiers are statistical classifiers. They can predict class membership probabilities, such as the probability that a given tuple belongs to a particular class. Bayesian classification is based on Bayes' theorem. Studies comparing classification algorithms have found a simple Bayesian classifier known as the Naïve Bayesian classifier to be comparable in performance with decision tree and selected neural network classifiers. Bayesian classifiers have also exhibited high accuracy and speed when applied to large databases. Naïve Bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is called class conditional independence. It is made to simplify the computations involve and, in this sense, is considered "naïve". Bayesian belief networks can also be used for classification. [5]

3.1.2 Bayes' Theorem

Bayes' Theorem is named after Thomas Bayes, a nonconformist English clergyman who did early work in probability and decision theory during the 18the century. Let X is a data tuple. In Bayesian terms, X is considered "evidence". Let H be some hypothesis, such as that the data tuple X belongs to a specified class C. P (H|X) is the posterior probability, or a posteriori probability, of H conditioned on X. P (H) is the prior probability, or a priori probability, of H. Bayes' Theorem is

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \qquad \textbf{Equation 1.}$$

3.2 Prediction

Numeric prediction is the task of predicting continuous (or ordered) values for given input. The most widely used approach for numeric prediction is regression. Regression analysis can be used to model the relationship between one or more independent or predictor variables and a dependent or response variable (which is continuous-valued). In the context of data mining, the predictor variables are the attributes of interest describing the tuple. In general, the values of the predictor variables are known. Techniques exist for handling cases where such values may be missing. The response variable is what to predict. Regression analysis is a good choice when all of the predictor variables are continuous-valued as well. Many problems can be solved by linear regression. There are straight-line regression analysis (which involves a single predictor variable) and multiple linear regression analysis (which involves two or more predictor variables).

3.2.1 Linear Regression

Straight-line regression analysis involves a response variable, y, and a single predictor variable, x. It is the simplest form of regression and models y as a linear function of x. That is,

$$y = b + wx, \qquad \textbf{Equation 2.}$$

In Equation 2, the variance of y is assumed to be constant, and b and w are regression coefficients specifying the Y-Intercept and slope of the line, respectively. The regression coefficients, w and b, can also be thought of weights, so that can equivalently write Equation 3,

$$y = w_0 + w_1 x \qquad \textbf{Equation 3.}$$

These coefficients can be solved for by the method of least squares, which estimates the best-fitting straight line as the one that minimizes the error between the actual data and the estimate of the line. Let D be a training set consisting of values of predictor variable, x, for some population and their associated values for response variable, y. In Equation 4, the training set contains |D| data points of the form $(x_1, y1), (x_2, y2)\ldots (x_{|D|}, y_{|D|})$ The regression coefficients can be estimated using this method with the following equations:

$$w_1 = \frac{\sum_{i=1}^{|D|}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{|D|}(x_i-\bar{x})^2} \qquad \textbf{Equation 4.}$$

$$w_0 = \bar{y} - w_1\bar{x} \qquad \textbf{Equation 5.}$$

In Equation 5, $\bar{x}$ is the mean value of x_1, x_2, $\ldots x_{|D|}$, and $\bar{y}$ is the mean of y_1, y_2, $\ldots$ $y_{|D|}$. The coefficients w_0 and w_1 oftenprovide good approximations to otherwise complicated regression equations. [3]

4. Proposed System design

4.1 System Design

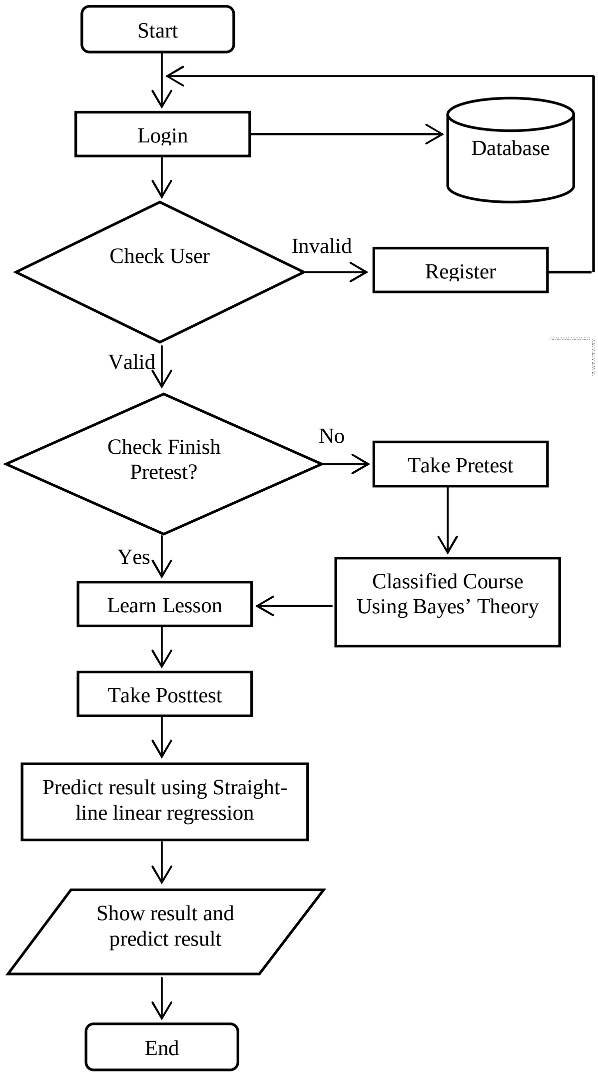

Figure1. Overview of system design

The main process of the system is that guide the course for the students who do not know what is their needs and interests. Initially, a pre-test is conducted to decide the student's prerequisite skills. The Bayes' Theorem is applied to model his/her knowledge dynamically to provide individualized learning. The results are then applied to guide the student intelligently. The course is classified by the Bayesian classification. The student model in Intelligent Tutoring System has the capability of recording information on students. The student can learn course that classified by the Bayesian classification. Finally, the student that finished learning lesson takes the post-test. The prediction model of straight-line regression using the method of least square predicts the post-test mark.

5 Predicting a class label using Naïve Bayesian classification

We wish to predict the class label of an unknown sample using Naïve Bayesian classification. The data samples of training table are described by 8 attributes (Programming Language, Management Information System, Database, Design, Database Design, Unified Modeling Language, Software Engineering, and Project Planning). The class label attribute, **Predict**, has five values (namely, {Programmer, System Analyst, Software Engineer, Project Manager, and Web Designer}). The remarks of attributes are High, Medium, and Low.

Table1. Training Data Set

ID	Programming Language	Design	SE	Database	MIS	DB Design	UML	Project Planning	Course
1	High	Medium	Medium	Medium	Low	Low	Medium	Low	Programmer
2	Medium	High	Low	Medium	Medium	Medium	High	Medium	System Analyst
3	High	Medium	Medium	High	Medium	Medium	High	Medium	Software Engineer
4	Medium	Medium	High	Medium	Medium	Medium	High	High	Project Manager
5	Low	Medium	Low	High	High	Medium	Low	Low	Web Designer
...	...	...	...	...	...	...	...	...	...

5.1 ROC curves for naïve Bayes'

The performance of a test or the accuracy of a test to discriminate is evaluated using Receiver Operating Characteristic (ROC) curve analysis. For every possible cut-off point or criterion value you select to discriminate between the two populations, there will be some cases with the course correctly classified as positive (TP=True Positive fraction), but some cases with the course will be incorrectly classified negative (FN=False negative fraction). On the other hand, some cases that the course will be correctly classified as negative (TN=True Negative fraction), but some cases without the course will be incorrectly classified as positive (FP=False Positive fraction).
Sensitivity (True positive rate) = TP/ (TP+FN) =0.46
Specificity (False positive rate) =TN/ (FP+TN) =0.85
Summary Receiver Operating Characteristic Curves-a statistical technique using linear regression to describe the accuracy or performance of a diagnostic test by plotting predicted true positive rates (y-axis) at given false positive rates (x-axis). The larger the area under the curve is the more accurate the test.
Let Sensitivity (predictor variable) is 0.6. The Specificity (response variable) gets 0.88 from the substitution of the predictor variable to the linear regression equation. The rest of Sensitivity can get from the substitution of the predictor variable to the linear regression equation. This is shown in Table 2.

Table2. Rate for ROC curves

(Sensitivity) True Positive Rate	(Specificity) False Positive Rate
0	0
0.2	0.78
0.46	0.85
0.6	0.88
0.8	0.94
1	1

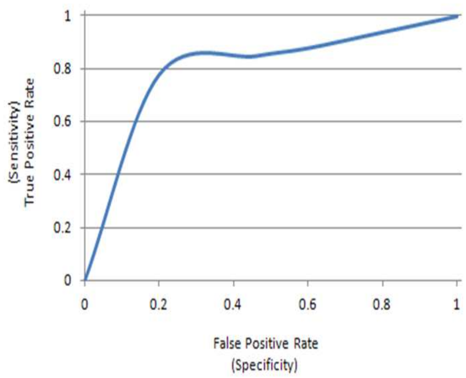

Figure3.ROC (Receiver Operating Characteristic) Curves for naïve Bayes

5.2 Classifier Accuracy Measures

Using training data to derive a classifier or predictor and then to estimate the accuracy of the resulting learned model can result in misleading overoptimistic estimates due to overspecialization of the learning algorithm to the data. The accuracy of a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier. Error rate or misclassification rate of a classifier, M, which is simply 1-Acc (M) where Acc (M) is the accuracy of M. True positives refer to the positive tuples that were correctly labeled by the classifier while true negatives are the negative tuples that were correctly labeled by the classifier. False positives are the negative tuples that were incorrectly labeled. Similarly, false negatives are the positive tuples that were incorrectly labeled. The sensitivity and specificity measures can be used, respectively, for this purpose. In Equation 6, sensitivity is also referred to as the true positive rate while specificity is the true negative rate in Equation 7. In Equation 8, the accuracy can calculate using sensitivity and specificity.

$$sensitivity = \frac{t\text{-}pos}{pos}$$ **Equation 6**

$$specificity = \frac{t\text{-}neg}{neg}$$ **Equation 7**

$$accuracy = sensitivity\ \frac{pos}{(pos+neg)} + specificity\ \frac{neg}{(pos+neg)}$$ **Equation 8**

The prediction results for 200 tuples of classified course have 87% accuracy.

5.3 Prediction of post-test mark using the straight-line regression using the method of least square

Table3. Student Mark

i	x_i (Pretest Mark)	y_i (Posttest Mark)	$(x_i\text{-}x)(y_i\text{-}y)$	$(x_i\text{-}x)^2$
1	60	34	16.4	420.25
2	70	45	359.9	930.25
3	80	33	-8.1	1640.25
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
200	41	54	248	1456
			$\sum$=10148.6	$\sum(x_i\text{-}x)^2$=101978.5

This prediction involves student pre-test marks, x (predictor variable) and student post-test marks, y (response variable). Firstly, we compute $\bar{x}$ and $\bar{y}$ get data from Table3. Substitution these values into Equation, we get w1=0.089 and w0=31.87. Thus, the equation of the least squares line is estimated by y= 31.87+0.089x. Using this equation, we can predict that the student who get 43 marks in pre-test is 36 marks in post-test. The 2-D data can be graphed on a scatter plot, as in Figure4. The plot suggests a linear relationship between the two variables, x and y.

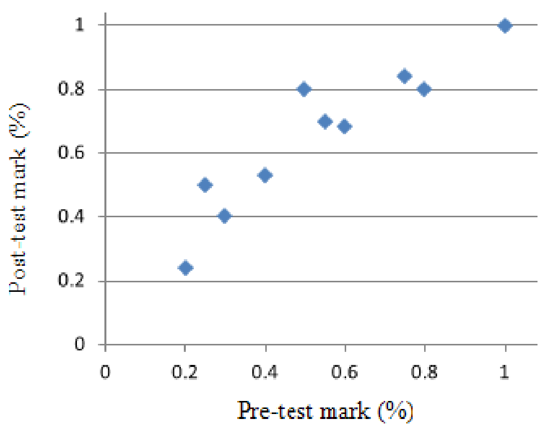

Figure4. Plot the data in Table3. Although the points do not fall on a straight line the overall pattern suggests a linear relationship between x (pre-test mark) and y (post-test mark).

6. Conclusions and Future Work

The main purpose of an intelligent tutoring system is to communicate embedded knowledge effectively and in a manner that suits the user. As the Intelligent Tutoring System seeks to mimic the human tutor while it imparts knowledge to students, it is necessary to guide students who have different levels of prerequisite skills. This paper discusses a new architecture of designing an Intelligent Tutoring System (BITS) for computer programming using Bayesian technology. The classification accuracies can be maintained or improved with the intervention of domain experts.

Naïve Bayesian classifiers are robust, can handle mixed variables and produce informative results (class probabilities). This system can be used to predict the post-test mark of students by using the straight-line regression using the method of least square which is simple and widely used. In this paper, naïve Bayesian classifiers can only be classified course from the prerequisite skills of the student. Naïve Bayesian classifiers can be classified not only course but also topics that do not understand the student.

7. References

[1]. S.B. Kotsiantis, C.J. Pierrakeas, and P.E. Pintelas. Preventing student dropout in distance learning using machine learning techniques. In Proceedings of 7th International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES), pages 267–274, 2003.

[2]. Jiawei Han Micheline Kamber, *"Data Mining Concepts and Techniques"*Academic Press, New York 2001

[3]. John Neter, MIcael H.Kutner, Christopher J, Nacht sheim, and William wasserman*"Applied Linear Regression Model"*3nd Ed, Chicago, IRWIN, INC., 1996

[4]. Minaei-Bidgoli, D.A. Kashy, G. Kortemeyer, and W. Punch. Predicting student performance: an application of data mining methods with an educational web-based system. In *Proceedings of 33rd Frontiers in Education Conference*, pages T2A13–T2A18, 2003.

[5] D. Geiger and D. Heckerman. Knowledge representation and inference in similarity networks and Bayesian martinets. *Artificial Intelligence*, 82:45–74, 1996.

[6] R. Freedman, “What is an Intelligent Tutoring System?” Final draft, Published in Intelligence 11(3): 15-16 (Fall 2000)

[7] D. Sleeman and J.S. Brown, Introduction: Intelligent Tutoring systems, Intelligent Tutoring Systems, (1982) 1-10.

[8] M. Villano, Probabilistic student models: Bayesianbelief networks and knowledge space theory, in: Proceedings of 2nd International Conference on Intelligence Tutoring System, Montreal, Canada, 1992, pp.491-498.

[9] P. Brusilovsky, Adaptive and intelligent technologies for web-based education, Special Issue on Intelligent Systems and Teleteaching, 4 (1999) 19-25

[10] C.J. Butz, S. Hua, RB. Maguire, A Web-based Bayesian Intelligent Tutoring System for computer programming

[11] Catach, L. (1988) Intelligent Agents. (AAAI-88) St. Paul, MN, pp. 491-495

[12] http://www.ai.mit.edu/Bayes/bnsoft.html

[13] http://en.wikipedia.org/wiki/Intelligent_tutoring_system

[14] http://www.medcalc.be