J. Basic. Appl. Sci. Res., 2(12)12017-12023, 2012
© 2012, TextRoad Publication

ISSN 2090-4304
Journal of Basic and Applied Scientific Research
www.textroad.com

Intelligent E-Learning Systems Using Student Behavior Prediction

Morteza Saberi Anari[1], Maryam Saberi Anari[2]

MA in Artificial Intelligence, Tehran, Iran

ABSTRACT

There were an increasing application of electronic and web systems for education in recent years. To increase efficiency of these systems and improve the education, artificial intelligence techniques and design of systems based on the student model are used. Using the student behavior prediction, the system can be customized for the individual. In this paper a method is provided for predicting the behavior of students using the Markov hidden model. Our input is a set of actions performed by the user during the training in interaction with the electronic system. First, by preprocessing the data collected from an e-learning system, the parameters of Markov hidden models were adjusted. Then Markov hidden models were thought by the help of Baum Welch algorithm using the training data. Then for detecting the new user behavior the pioneer algorithm was used. The accuracy of prediction in the proposed method was better than the designed artificial neural networks.

KEYWORDS: Behavior prediction, the e-learning systems, Markov hidden model, Baum - Welch algorithm, pioneer algorithm

INTRODUCTION

The idea of using computers and computer networks for an educational task goes back to the twentieth century and early1960's. Although the main goal of the ARPANET project was to establish a network for military information exchange and protection of the U.S. from the consequences of the Cold War with the former Soviet Union, yet three of four centers for the computer networks implementation developed in the course of school education. Since the early 70's when the project was yielded, exchange of scientific information and data between the U.S. training centers began. The e-learning as it is today and in this range of application was formed in the early 1990s with the advent of the worldwide internet, and given the many capabilities of web, e-learning has grown rapidly, and today has stabilized its place in the educational structure of many countries. Today, schools and universities use electronic systems to teach their students. This web based systems also provided an effective step towards distance education to the extent that it led to creation of virtual universities. To improve and increase the efficiency of e-learning systems, artificial intelligence techniques and system designs have been used based on the student model. Prediction the behavior of the learner is a useful tool for customization and providing a better training [Mousavi etal,2008].

Most of the intelligent e-learning systems have three major sections: an expert system which includes the domain of system knowledge and the method for reasoning knowledge for problem solving, the student's model which includes information about the learner, and teaching technique which includes appropriate teaching strategy [Yacef,2004].

E-learning systems save details of student activities and their interactions with the educational system. Thus it provides a massive collection of training data which can be used for statistical processing. Different algorithms and different methods of artificial intelligence have led to finding hidden patterns and knowledge from educational systems data and have helped considerably the decision makers in the field of higher education to promote and improve educational processes such as planning, registration, evaluation and counseling. Using the set of collected data, the learner behavior can be predicted during the education. The behavior means the amount of user work according to the works that can be done in the e-learning system [Yaghini,2010].

Using the activities of e-learning system learners and using hidden Markov model which is a useful statistical method, this paper tries to predict the future behavior of a new learner. Depending on the user behavior, a system can be personalized for the learner and provide suggestions to improve the student's learning.

2 - Research Background

Educational systems gain rich data about students' behavior during learning. Using artificial intelligence algorithms data mining techniques, it is possible to describe, interpret and predict the behavior of the learner, and

***Corresponding Author**: Morteza Saberi Anari, MA in artificial intelligence, Tehran, Iran.
E-mail:morteza.saberi88@ gmail.com

also evaluate the progress process in relation to learning outcomes. Detecting patterns in the behavior of learners at problem level and prediction in between educational meetings using data related to educational systems are other researches in this field.

Given that the probability of dropping out of school in the first year is more than any other educational year, educational institutions by predict this kind of students are able to properly adjust their marketing strategies and other programs and thereby improve future student retention rate. There is a study by Ming Yang to predict first-year university students' withdrawal using k-means algorithm [Yang,2006].

Predicting enrollment of students is fundamentally important in the mission of a university. There is a study by Aksenova in 2006 to predict the student registry using support vector machines and rule-based predictive models for students of computer science at California State University. Data used include population, unemployment rates, fees and taxes, family income, high school graduation rates and historical data of registry is the past [Aksenova etal, 2006].

In a study by Ming Yang, Poisson regression model is used to design a prediction model for registration in a particular subject (General Mathematics). Data used to design a model of the data related to fall 2004 include the number of students in each study subject and level from the first year to fourth year students, and the number of people registered from each study subject and level for general mathematics [Yang,2006].

Prediction of the next actions of the students is presented using weighted Markov model in 2008 by Huang et al. As Markov models with low order have low accuracy and Markov models with high-order have computational complexity, in this methods using weighting method the expressed problems were resolved. Weights show the similarity between the actions taken by the students in this study and other students who have used the system. To calculate the similarity, first a sequence of actions the student has done so far is displayed, and then using dynamic programming techniques finds the largest share between two learning paths and then the consistency between the learning path of the students and the largest share is calculated [Huang etal,2008].

3 - Hidden Markov model

Hidden Markov models first were introduced in a series of statistical papers by Leonard E. Baum and other authors in mid 1960s. Its first application was speech recognition that was started in mid 1970s. At mid-1980s it was used for the analysis of biological sequences, particularly DNA. Since then it was considered as a field of Bioinformatics. Statistical methods of Markov source or hidden Markov model became increasingly popular in recent years. There are two very powerful reasons for this popularity; firstly, these models are very rich in the mathematical structure, and thus can form a theoretical basis to be used in a wide range of applications. Secondly, when the models are used correctly, they work very well for important applications [Rakesh, Desai,1996].

The term hidden is used for the subject, because we are talking about the issues that the way of doing them is hidden from our view, and have the parametric nature. This means that not only we do not know what the outcome will be, but the type of that happening and the probably of that happening should be concluded from available parameters. It is like throwing a coin in a closed box or in a room away from our view. It means that the resulting model is a stochastic model with an underlying random process which is invisible (hidden) for the observer, and is only deductible by a set of stochastic processes that produce pursuance of observations.

4 - Proposed Methods

In this part data collection method and the way to use data will be introduced. Proceedings in phases of the model values adjustment and prediction are explained. Then procedures are presented in the form of a chart.

4-1 - Data Collection

Data has been collected from the events recorded in the e-learning system at Khajeh Nasir University which is a Moodle open source system. Moodle learning system is a web-based e-learning project which presents itself as a course management system. It helps students to be able to make effective educational activities on-line. Model claims that now with one site and a professor could be able to facilitate a University with 200,000 students. It is also introduced as a web content system for e-learning, so as long as there is e-learning, the model will be.

Then by evaluating applicable actions in the system, the works are reduced to 4 items and are labeled and actions done by each student in the class during the term are obtained, segregated by the related study subject. These actions include enrolment to the study subject, observation of lesson archive, participation in the discussion, and participating in online classes, and they are respectively numbered 1 to 4.

4-2 - Hidden Markov model parameters

A hidden Markov model can be created with the following parameters [Lawrence, Juang, 1986 and Mark, 2004]

• N the number of hidden states of model

• M the number of observation symbols in the alphabet. If the observations are discrete, then M will have an unlimited amount.
• $A = \lfloor a_{ij} \rfloor$ Transition state matrix: a set of transition probabilities between states

$$(1)\quad a_{ij} = p\{q_{t+1} = j \mid q_t = i\}, 1 \le i, j \le N$$

In which q_t represents the current state. Transition probabilities should fulfill the normal limitation of a random probability distribution. These limits include the following.

$$(2)\quad a_{ij} \ge 0, 1 \le i, j \le N$$

$$(3)\quad \sum_{j=1}^{N} a_{ij} = 1, 1 \le i \le N$$

• Observation probability distribution: a probability distribution for each of the states:

$$(4)\quad b_j(k) = p\{o_t = v_k \mid q_t = j\}, 1 \le j \le N, 1 \le k \le M$$

In this relation v_k indicates k^{th} that is the observed in alphabets and o_t reflects the current input vector parameters.
• $\pi = \{\pi_i\}$ The probability distribution of the initial state in which

$$(5)\quad 1 \le i \le N \quad \pi_i = p\{q_1 = i\},$$

Thus a hidden Markov model with discrete probability distribution using the following three can be detected.

$$(6)\quad \lambda = (A, B, \pi)$$

In this study the hidden states are considered as student activity levels in three levels of low, medium and high. Observed states will be actions taken by the students in interaction with e-learning systems.

The learning Markov model is the optimal estimation of (λ) model parameters and is done by one of two Viterbi or Baum - Welch methods. Model training course can also be a combination of the two methods. In this paper, the method of Baum - Welch is used in estimating the parameters. One of the special features of this algorithm is that its convergence is guaranteed.
Forward and backward variables defined in Hidden Markov Models respectively in relations 7 & 8:

$$(7)\quad \alpha_t(i) = p\{o_1, o_2, \ldots, o_t, q_t = i \mid \lambda\}$$

$$(8)\quad \beta_t(i) = p\{o_{t+1}, o_{t+2}, \ldots, o_T \mid q_t = i, \lambda\}$$

To describe the Baum - Welch algorithm that is also known as forward - backward algorithm, in addition to forward and backward variables, more covariates shall be defined.
The first covariate is the probability of being in state *i* at time *t* and in state *j* at time t +1, which is defined as follows.

$$(9)\quad \varepsilon_t(i, j) = p\{q_t = i, q_{t+1} = j \mid O, \lambda\}$$

The above covariate is defined using forward and backward variables as follows.

$$(10)\quad \varepsilon_t(i, j) = \frac{\alpha_t(i)\alpha_{ij}\beta_{t+1}(j)b_j(o_{t+1})}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)\alpha_{ij}\beta_{t+1}(j)b_j(o_{t+1})}$$

The second covariate indicates the late possibility of i state with sequence observations and hidden Markov models and is expressed as follows.

$$(11)\quad \gamma_t(i) = p\{q_t = i \mid O, \lambda\}$$

This variable using the forward and backward variables is as follows.

$$(12)\ \gamma_t(i) = \left[\frac{\alpha_t(i)\beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)} \right]$$

The relationship between these two variables is expressed as follows.

$$(13)\ \gamma_t(i) = \sum_{j=1}^{N} \varepsilon_t(i, j),\ 1 \le i \le N, 1 \le t \le M$$

Baum - Welch learning algorithm is obtained with maximizing the $P(O \mid \lambda)$ value. The next step is to update the model parameters with respect to the following re-estimation relationships.

$$(14)\ \overline{\pi}_i = \gamma_1(i), 1 \le i \le N$$

$$(15)\ \overline{a}_{ij} = \frac{\sum_{t=1}^{T-1} \varepsilon_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}\ 1 \le i \le N, 1 \le j \le N$$

$$(16)\ \overline{b}_j(k) = \frac{\sum_{\substack{i=1 \\ o_i = v_k}}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)},\ 1 \le j \le N, 1 \le k \le M$$

In this study a semester was divided into two periods, and given the three levels of activity, there will be nine different models (Table 1). To determine the student's level of activity, first a weight is determined for each practice and in each period, the total weight of practices are calculated and by specifying two thresholds the activity levels are determined (Table 2).

Table 1: The nine different categories of students

Number of state	Activity levels first part	Activity levels second part
1	Low	**Low**
2	Low	**Moderate**
3	Low	**High**
4	Moderate	**Low**
5	Moderate	**Moderate**
6	Moderate	**High**
7	High	**Low**
8	High	**Moderate**
9	High	**High**

Table 2: The obtained ranges related to the three levels of student status level

High activity level	Moderate activity level	Low activity level
More than 250	Between 80 to 250	**Less than 80**

Using Baum - Welch relationships and training data, 9 parameters of Hidden Markov Model are obtained.
In the following figure a part of the hidden Markov model training method is proposed.

Figure 1: Learning steps and the method to create the hidden Markov model

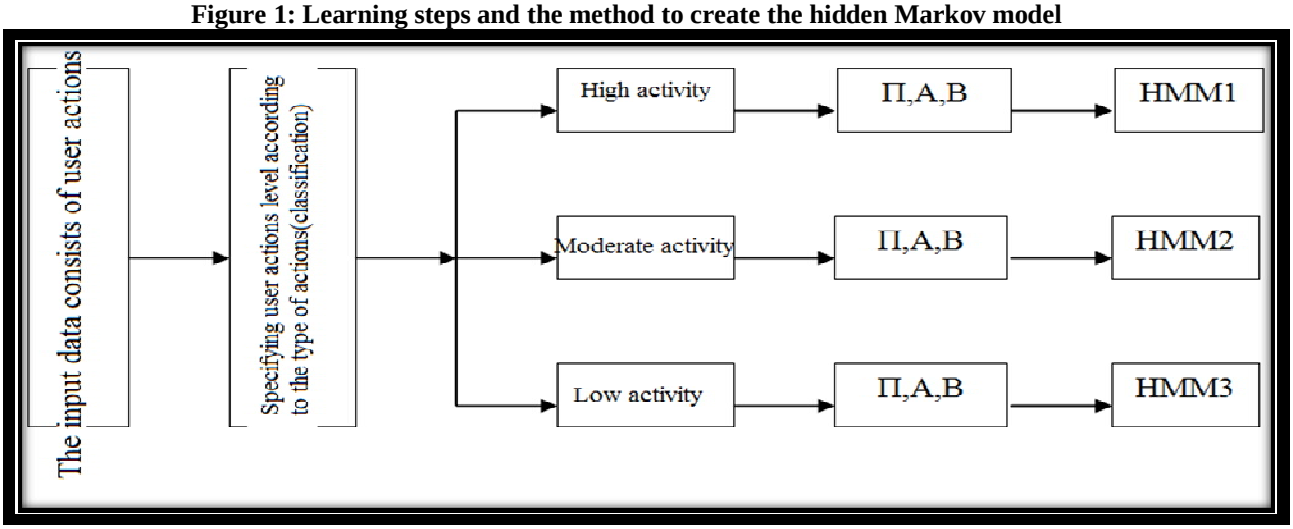
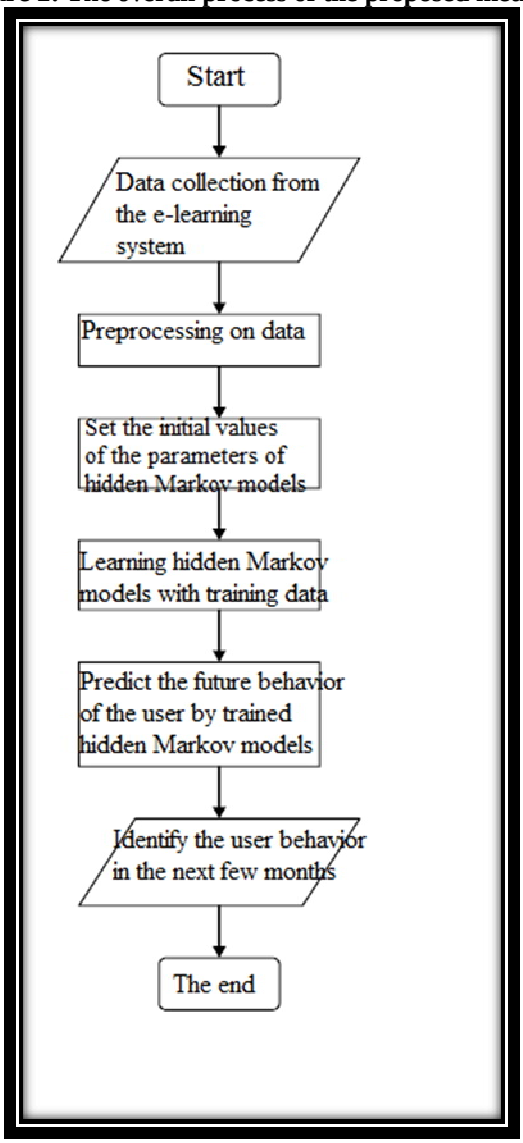


Figure 2: The overall process of the proposed method.



4-3 - Testing models

To evaluate the model, the use forward and backward processes have offered. In this method all sequences of states with t length are considered. For example, the probability for generated sequence of observations $O = \{O_1,...,O_t\}$ is calculated from the sequence of states $Q = \{q_1,...,q_t\}$ from the model λ as follows.

$$(17)\quad P(O \mid Q, \lambda) = \prod_{i=1}^{t} P(O_i \mid q_i, \lambda)$$

That can be written as the following:

$$(18)\quad P(O \mid Q, \lambda) = b_{q_1}(O_1) b_{q_2}(O_1) ..... b_{q_t}(O_t).$$

The probability of sequence of states $Q = \{q_1,...,q_t\}$ is calculated as follows:

$$(19)\quad P(Q \mid \lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} ..... a_{q_{t-1} q_t}$$

However, the probability that both events O and Q occur simultaneously is shown with $P(O, Q \mid \lambda)$:

$$(20)\quad P(O, Q \mid \lambda) = P(O \mid Q, \lambda) P(O, \lambda)$$

Here the probability of O occurrence is obtained with sum of all possible sequences.

$$(21)\quad P(O \mid \lambda) = \sum_{allQ} P(O \mid Q, \lambda) P(Q, \lambda)$$

Figure 3 presents stages related to test of prepared models. At this stage, from 100 available students' data, 10 are considered as experimental data. Respectively each of students' actions sets were sent to the system, and then are tested with 9 new models that each have their 3 A, B, π parameters produced in the training phase. To test the dhmm_logprob function, Matlab software is used. By sending nine new models with the experimental data set, nine probabilities for every one of the nine new models are obtained for the new data. Finally, the data is given to a state with a higher probability.

Figure 3: Stages for testing the design education system using Hidden Markov Models

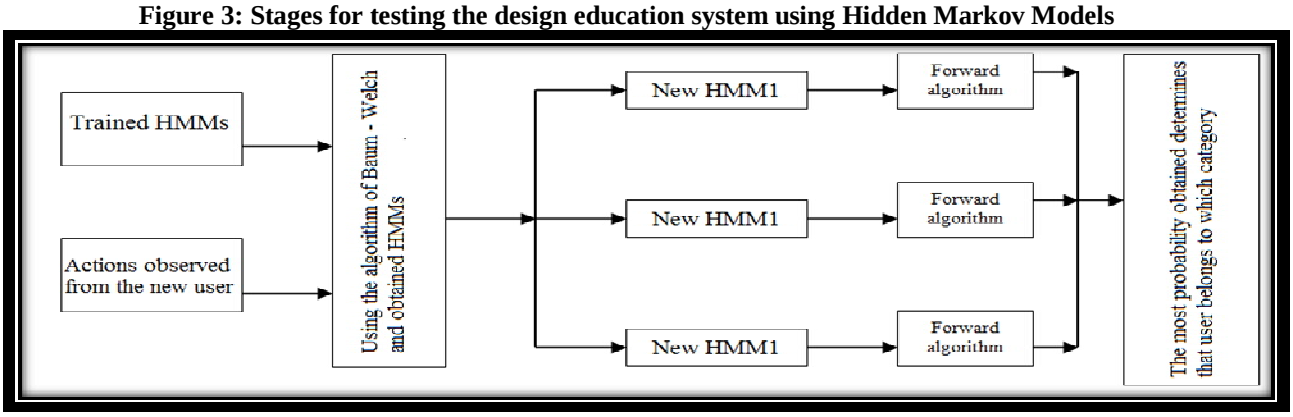
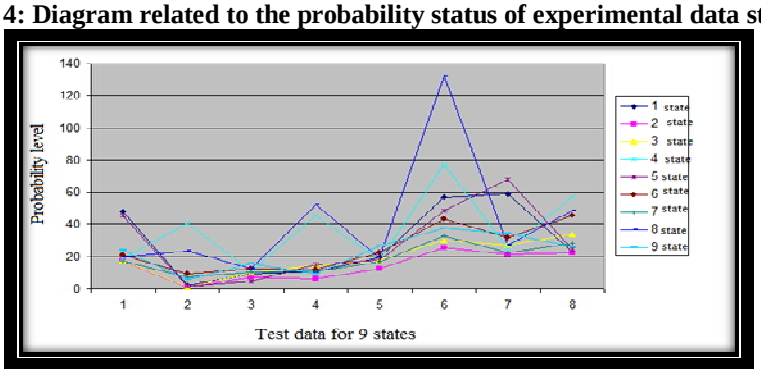


Figure 4 presents the charts for possibilities related to experimental data in state 6. Average of performance after 8 implementations for 9 different states related to student status is equal to 83.59.

Figure 4: Diagram related to the probability status of experimental data state six



Then to compare our work, we designed a Multilayer Perceptron neural network [10] to predict the progress of students in the next month. The input of neural network is student's actions and achievement of student in the first month. Therefore the neural network will have 5 inputs that are respectively enrolment to the course, observing lessons archive, log on to the online classes, participate in class discussions and the total number of actions in that interval. Neural network has one output which is the rate of improvement in the next month.

The designed neural network at best has 3 layers as (1-20-5) and the hidden layer is considered equal to 20. Network was been tested with 100 data and then was tested and evaluated with 30 data and efficacy rate was 81.5% which shows that the proposed method is superior in terms of performance.

5 - Conclusion

According to the related research in the fields of education, there has not been behavior prediction using the user's previous actions. For this project, first with data collection from e-learning systems and labeling data, the actions of each student in a special lesson was obtained. Based on strong mathematical structure of Hidden Markov models and its broad applications in different fields, in order to predict the behavior in e-learning systems this tool was used. With the training and preparation Hidden Markov models associated to the use of Baum - Welch algorithm, and then by using forward algorithms, the behavior of new user was predicted.

Results are acceptable due to lack of training data and prediction percent equals to 83.59. To compare the proposed method the perceptron neural network was designed and the accuracy rate equals to 81.5. It shows that our method has more acceptable performance.

Given that the hidden Markov model in this paper has been used as a model for behavior prediction, and the results are very good and acceptable, it can be concluded that it is a good tool to predict subjects such as weather forecasted, air pollution levels, the web page user calibration, and in the field of tool security for prediction networking attacks and network traffic and such.

REFERENCES

[1] Yaghobi Mousavi, Saeed, Pazuki, Iraj, e-learning and its impact on educational systems, Islamic Azad University Press, South Tehran Branch, 2008.

[2] Yaghini, Masoud, Heydari, Somayeh, "data mining to enhance and improve the higher education system processes", Fourth Conference on Data Mining, Iran, Tehran, 2010.

[3] Ming Yang; "Data Mining Techniques Applied to Texas Woman's University's Enrollment data - What Can the Data Tell us?", MS Thesis, Texaz Woman's University, 2006.

[4] Aksenova SS, Du Zhang, Meiliu Lu; "Enrollment Prediction through Data Mining", IEEE International Conference on Information Reuse and Integration, pp. 510-515, 2006.

[5] Xiaodi Huang, Jianming Yong, Jiuyong Li, Junbin Gao, "Prediction of student actions using weighted Markov models", IT in Medicine and Education, IEEE International Symposium on Digital Object Identifier, 2008.

[6] Kalina Yacef, "Making Large Class Teaching More Adaptive with the Logic-ITA", Sixth Australasian Computing Education Conference, Dunedin, New Zealand18-22 January, 2004.

[7] D. Rakesh., U. Desai, "A tutorial on hidden Markov models", Technical Report No.: SPANN-96. 1, 996.

[8] R. Lawrence and B.H. Juang, "An introduction to hidden Markov models", IEEE ASSP Mag, 1986, vol 3. pp. 4-16.

[9] S. Mark, "A revealing introduction to hidden Markov models", 2004, http://www.cs.sjsu.edu/faculty/stamp/RUA/HMM.pdf.

[10] Balabin, R.M., Safieva.R.Z., Lomakina.E .. "Comparison of linear and nonlinear calibration models based on near infrared (NIR) spectroscopy data for gasoline properties prediction". Chemometr Intell Lab, 88 (2), 183-188, I 2007.