

Using Data Mining in E-Learning Environment for Student Modelling

Satyawan Hembade & Dr. Ajit More

Institute of Management and Entrepreneurship Management,
Bharati Vidyapeeth (Deemed to Be) University, Pune, India

Received: December 06, 2018

Accepted: January 24, 2019`

ABSTRACT: *E-Learning systems are just traditional based web portals or Intelligent Tutoring Systems (ITS) or Learning Management Systems (LMS). During the usage of such systems, various kind of data is generated, this includes navigation sequence, kind of contents accessed, participation in various activity and performance of them. This data can be mined to find some hidden patterns and result of mining process is used to redesign the course contents according to student's need. There is need to focus use of data mining to understand how student learns in such environment. This article describes various methods of student modelling and use of data mining for it.*

Key Words: *Data Mining, Student Modelling, Educational Data Mining, Bayesian Network, Fuzzy Logic*

1. Introduction

E-learning utilizes a network (LAN, WAN or Internet) for delivery, interaction, or facilitation. Evolutions in the technology also evolved eLearning. These evolutions make use of wide range of technologies in eLearning systems. Now days there are many number of eLearning systems; which are either open source or owned by individual organisations. These are web based applications referred as Intelligent Tutoring Systems (ITS), Content Management Systems(CMS) or Learning Management Systems(LMS). Ellis (2009) defines LMS as a “software application for the administration, documentation, tracking, reporting and delivery of educational courses or training programs”. The eLearning systems are need to be provide more benefit to the student and teacher; to make sure this the data generated during the usage of e-learning system can be mined and referred as e-learning data mining. Data generated during usage of eLearning system is analysed and a model of learning behaviour called ‘student model’. This model can be developed using traditional data analysis techniques or data mining techniques can also be used.

2. Data Mining for Student Modelling

In education today teaching and learning process is changed from traditional face to face interaction to web based educational systems. Evolutions of internet make it possible to use web technologies in online courses and this use web based educational systems grown exponentially in last few years. In eLearning environment neither a teacher nor student is bound to present at a specific location. Teacher and student both can make use of internet to float courses and learn course contents remotely and courses and process are independent of machine platforms (Brusilovsky & Peylo, 2003).

In online learning or e-learning each student has different style in learning process. Process of learning by a student is much more important to understand how student learns in the eLearning system. There is need to adapt behaviour of learner to make learning process better. Modelling this behaviour is called as student modelling or in general terms user modelling. The User model is need to be build and update, this entire process is called user modelling. The primary goals of user modelling may be listed as follows (A Jameson, C Paris, 1997):

- To assist the user in locating information.
- To customize the information presented to the user.
- To modify an interface according to the user.
- To select appropriate instructional exercises or interventions.
- To provide feedback to the user on the level of their knowledge.
- To reinforce collaboration.
- To predict future behaviour of users.

The user in eLearning system is the student. Therefore, in eLearning systems, the concept of “student model” is used instead of user model.

Frias-Martinez, Magoulas, Chen and MacRedie (2005) classifies student models are in four groups depends on the use of model:

- Prediction Model: it has capability to predict needs and scores of student using past behavior of student.
- Recommendation Model: this model has capability of suggesting interesting elements to student.
- Classification Model: this model classifies items (students, course contents, questions and other items) into predefined group or classes.
- Filtering Model: it is the filtering items from the original set and provide it to student corresponding to her/his preferences.

3. Student Modelling Process

Zhou and Evens(1999) specifies that to build a student model, the first decision is to select aspects to be modelled. Kobsa (2001)describes the factors to be considered when student model is to be build. He specifies that “the student model should be constructed with student’s’ goals, knowledge, capabilities, preferences, multimedia experience, interests, personality in mind, which are effective in their learning”. Somyürek (2009) proposes the student modelling process as shown in Figure 1. Student modelling Process (Somyürek, 2009).

Somyürek (2009) states that there are three basic steps in student modelling process:

- Collect or gather data about the student characteristics: data required to construct model is need to be gathered; it contains goals, plans, attitudes, capabilities, knowledge, beliefs of the learner, etc. This data can be static data or dynamic data. Static data includes personal data and cognitive style while dynamic data is how student interact with, grades of student at various attempts and performance improvement of a student over the time. To get this data various methods such direct questions, assumptions and student system interaction.
- Build or Construct the student model: to constructa student model information collected is represent learning process and this representation is used to make decisions. These decisions include predicting learning path of a student, and grades of student in courses. Student model is also used to choose most suitable pedagogical strategy for new subject design to be presented to the students.
- Refine/update the student model: models designed are not valid for all types of students. These models are need to update as more data is collected through usage of eLearning systems.

4. Student Modelling Techniques

There are various student modelling techniques are used to implement student models and these are knowledge and behavioural based (Kobsa, 2001). The knowledge based student models are builds on data collected through direct questions to student through questionnaire and studies of student. The data collected has a purpose to obtain initial heuristics. Behavioural model is build using data collected from the observations of the student during usage of system. Somyürek (2009) mentions various techniquesconstruct student models; these are

- Stereotype model
- Overlay model
- Perturbation model
- Fuzzy logic model
- Machine learning techniques (Bayesian network).

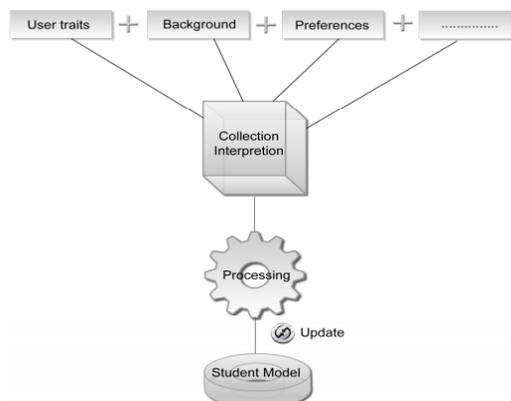


Figure 1. Student modelling Process (Somyürek, 2009)

4.1 Stereotype model

Stereotype model is a model which is easiest to construct. The stereotype model is used to classify users/students in groups and used to generalize student characteristics to that group (Kobsa, 2001). In this model it is necessary to define set of characteristics of groups for classification and granularity degree need to maintain. Groups are formed based on common characteristics of individuals such as user is beginner or expert(Figure 2. Stereotype Model).

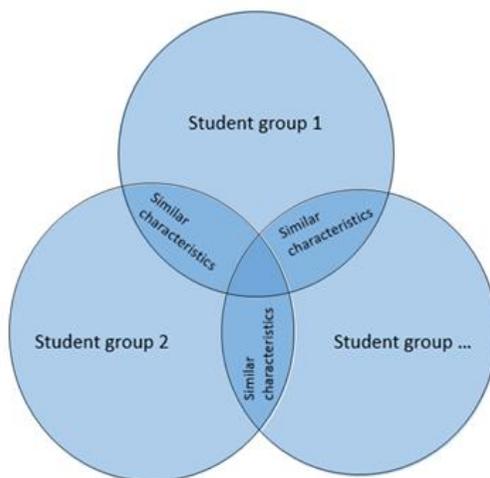


Figure 2. Stereotype Model(Somyürek, 2009)

In situations when there need to adapt user interface and learning type, stereotype model is sufficient. When there is need to have individual adaptations or guidance specific need to provide to individuals or suggestion systems by system are stereotype model might fall short (Koch, 2000, p: 51).

4.2 Overlay Model

Beck, Stern, and Haugsjaa (1996) describes overlay models to be used to represent student model in which student knowledge is represented as subset of expert knowledge. Representation of overlay model by Martins, Faria, De Carvalho, and Carrapatoso (2008) is shown in Figure 3. This model represents relation of student knowledge as to domain/expert knowledge as layer to layer and student knowledge is always a subset of system knowledge. This model is based on the principle that the contents provided by eLearning system systems will always ensure that there is correlation between student knowledge and systems knowledge. One of the fact of disadvantages of this model is that it does not considers that student may have some beliefs and hence that beliefs are not part of the model.

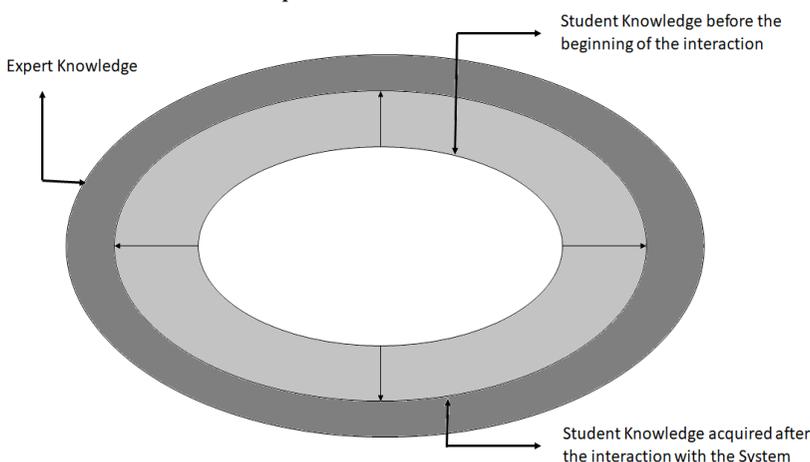


Figure 3. Representation of Overlay Model (Martins et al., 2008)

Martins et al.(2008)describes that this model represents domain model as individual topics and concepts. The success if this model is more depends on how domain model is divided into small and topics and concepts. Complexity of this model is depends on granularity of domain model structure and the assessment of knowledge acquired by student through interaction or usage of the system.

4.3 Perturbation Model

Students are always having some misconceptions about subject or subject topics. In overlay model there is no chance to represent student's misconceptions or misbeliefs. There is need to have modifications in overlay model to represent wrong information. Perturbation model constructed with modification to overlay model which overcomes limitation of representing misconceptions (Beck et al., 1996). Webber (2004) states that the perturbation model differs from the overlay model since it doesn't perceive student's knowledge as a simplification of expert knowledge, but rather like perturbations over the expert knowledge. This model is much more useful in adaptive systems as students learn concepts by mistakes or we can say that the students miss concepts are corrected with proper explanations and annotations to make learning easy or we can say that this model helps learners to correct errors (Nguyen & Do, 2008). This model helps in much better way to identify errors made by student and transform this errors or false information into meaningful data (Beck et al., 1996).

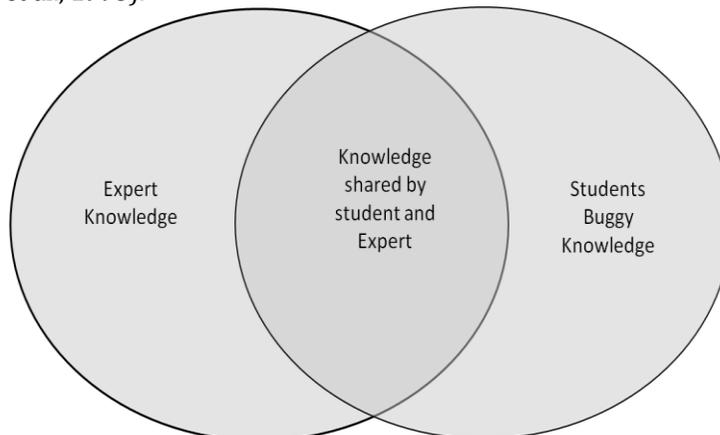


Figure 4. Perturbation Model (Beck et al., 1996)

As in this model learning starts with some misconceptions and later on during learning process these misconceptions are corrected; this model is called as buggy model. It allows to associate each wrong rule; which are obtained from the tutors pedagogical and domain experience. The student model is obtained by replacing each wrong rule by correct rules. As number of learners uses system to learn concepts with their wrong answers to specific concepts. The adaptive eLearning system proceeds by discriminating problems and present them to the student to know where student has wrong rules or misconceptions.

4.4 Fuzzy logic model

This model is based on the concepts of fuzzy set and certain mechanism is required to correctly classify class. Fuzzy Logic makes use of natural language to represent set and If-Then rules with natural language statement are used to demonstrate knowledge of domain experts. Fuzzy logic method is used to generates decisions and facilitates human reasoning based on imprecise information coming from the student-computer interaction (Goel, Lallé, & Luengo, 2012).

Few of the applications of fuzzy logic methods are discussed in section **Error! Reference source not found..** One of the approach of using fuzzy logic method with multi agent based e-learning framework is proposed by Goel et al.(2012). Here author proposes framework based on the fuzzy logic method to further suggestions to learner to increase his/her satisfaction and provide enhanced and personalized learning experience. Huapaya (2012) describes a model that captures the expert knowledge of experienced professors and is used to design a cognitive diagnostic model based on Fuzzy Logic.

4.5 Bayesian Network

"A Bayesian Network is a directed acyclic graph $G = \langle V, E \rangle$, where every vertex v in V is associated with a random variable X_v , and every edge (u, v) in E represents a direct dependence from the random variable X_u to the random variable X_v . Let $\text{Deps}(v) = \{u \mid (u, v) \text{ in } E\}$ denote the direct dependences of node v in V . In a Bayesian Network, each node v in V of the graph is associated with a conditional probability distribution $\text{CPD}(v)$, which denotes the probability distribution of X_v conditioned over the values of the random variables associated with the direct dependences $D(v)$." (Gordon, Henzinger, Nori, & Rajamani, 2014)

Bayesian network is network of probabilistic variables or observables and relationships between them. Variables in the network are often defined by expert and relations are dependency between them.

5. Research Scope in eLearning Data Mining

It is found that the majority of the research about using eLearning systems, gathering data from it and using it for applying DM techniques is carried out by universities in the Europe, USA, China, and developed countries. In India there is scope to implement an eLearning system to experiment about acceptance of such systems by teachers, student and of course by universities. There also scope for identifying eLearning systems and designing tools to extract data from such system. As these systems are new and students are using first time. There is ample scope for identifying type of course contents and their impact on concept understanding. As student models are helping to understand them well; there is scope for experimenting on modelling a learner model. To model student behaviour data mining can contribute and hence there is scope for identifying data mining techniques can be used for the same.

References

1. A Jameson, C Paris, and C. T. (1997). "Reader's Guide," in User Modeling. In Proc of the Sixth International Conference UM97. New York: Springer-Verlag.
2. Beck, J., Stern, M., & Haugsjaa, E. (1996). Applications of AI in education. *Crossroads*, 3(1), 11–15.
3. Brusilovsky, P., & Peylo, C. (2003). Adaptive and Intelligent Web-based Educational Systems. *International Journal of Artificial Intelligence in Education (IJAIED)*, 13, 159–172.
4. Frias-Martinez, E., Magoulas, G., Chen, S., & MacRedie, R. (2005). Modeling human behavior in user-adaptive systems: Recent advances using soft computing techniques. *Expert Systems with Applications*, 29(2), 320–329.
5. Goel, G., Lallé, S., & Luengo, V. (2012). Fuzzy logic representation for student modelling: Case study on geometry. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7315 LNCS, 428–433.
6. Gordon, A. D., Henzinger, T. A., Nori, A. V., & Rajamani, S. K. (2014). Probabilistic Programming. *Proceedings of the 36th International Conference on Software Engineering (ICSE 2014, Future of Software Engineering Track)*, 167–181.
7. Huapaya, C. R. (2012). Proposal of fuzzy logic-based students' learning assessment model. XVIII Congreso Argentino de Ciencias de La Computaci{ó}n.
8. Kobsa, A. (2001). Generic user modeling systems. *User Modeling and User-Adapted Interaction*, 11(1–2), 49–63.
9. Koch, N. (2000). Software engineering for adaptive hypermedia systems *Software Engineering for Adaptive Hypermedia Systems Reference Model , Modeling Techniques and Development Process* Nora Parcus de Koch Dissertation des Doktors der Naturwissenschaften. Ph.D Thesis, Ludwig-MaximiliansUniversity of Munich, (August), 51.
10. Martins, A. C., Faria, L., De Carvalho, C. V., & Carrapatoso, E. (2008). User Modeling in adaptive hypermedia educational systems. *Educational Technology Society*, 11(1), 194–207.
11. Nguyen, L., & Do, P. (2008). Learner Model in Adaptive Learning. *World Academy of Science Engineering and Technology*, 45(November), 395–400.
12. Somyürek, S. (2009). Student modeling: Recognizing the individual needs of users in e-learning environments. *International Journal of Human Sciences*, 6, 429. Retrieved from
13. Webber, C. (2004). From Errors to Conceptions – An Approach to Student Diagnosis. In *ITS '04* (pp. 710–719).
14. Zhou, Y. Z. Y., & Evens, M. W. (1999). A practical student model in an intelligent tutoring system. *Proceedings 11th International Conference on Tools with Artificial Intelligence*, 13–18. <https://doi.org/10.1109/TAI.1999.809759>