IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 6, August 2014.
www.ijiset.com
ISSN 2348 – 7968

Data Mining For A Web-Based Educational System

R.ADAIKKALAM MCA.,

Assistant Professor in Computer Applications, Sri Venkateshwara College Of Arts & Science, Peravurani, Thanjavur District.

INTRODUCTION

Web-based educational technologies allow educators to study how students learn (descriptive studies) and which learning strategies are most effective (causal/predictive studies). Since web-based educational systems are capable of collecting vast amounts of student profile data, data mining and knowledge discovery techniques can be applied to find interesting relationships between attributes of students, assessments, and the solution strategies adopted by students. The ever-increasing progress of network-distributed computing and particularly the rapid expansion of the web have had a broad impact on society in a relatively short period of time. Education is on the brink of a new era based on these changes. Online delivery of educational instruction provides the opportunity to bring colleges and universities new energy, students, and revenues.

Many leading educational institutions are working to establish an online teaching and learning presence. Several different approaches have been developed to deliver online education in an academic setting First, this Paper presents an approach to classifying student characteristics in order to predict performance on assessments based on features extracted from logged data in a web-based educational system. We show that a significant improvement in classification performance is achieved by using a combination of multiple classifiers. we have successfully improved the accuracy of the combined classifier performance. Such classification is the first step towards a “recommendation system” that will provide valuable, individualized feedback to students.

The web browser represents a remarkable enabling tool to get information to and from students. That information can be textual and illustrated, not unlike that presented in a textbook, but also include various simulations representing a modeling of phenomena, essentially experiments on the computer. Its greatest use however is in transmitting information as to the correct or incorrect solutions of various assigned exercises and problems. It also transmits guidance or hints related to the material, sometimes also to the particular submission by a student, and provides the means of communication with fellow students and teaching staff.

DATA MINING

The amount of data stored in databases is increasing at a tremendous speed. This gives rise to a need for new techniques and tools to aid humans in automatically and intelligently analyzing huge data sets to gather useful information. This growing need gives birth to a new research field called Knowledge Discovery in Databases (KDD) or Data Mining, which has attracted attention from researchers in many different fields including database design, statistics, pattern recognition, machine learning, and data visualization. Our motivation is gaining the best technique for extracting useful information from large amounts of data in an online educational system.

The main goals to obtain an optimal predictive model for students within such

ISSN 2348 – 7968

systems, help students use the learning resources better, based on the usage of the resource by other students in their groups, help instructors design their curricula more effectively, and provide the information that can be usefully applied by instructors to increase student learning. Data Mining is the process of analyzing data from different perspectives and summarizing the results as useful information. It has been defined as "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data"

Providing an understanding of the application domain, the goals of the system and its users, and the relevant prior background and prior knowledge.

- Selecting a data set, or focusing on a subset of variables or data samples, on Which discovery is to be performed.
- Pre-processing and data cleansing, removing the noise, collecting the necessary information for modeling, selecting methods for handling missing data fields, accounting for time sequence information and changes.
- Data reduction and projection, finding appropriate features to represent data, using dimensionality reduction or transformation methods to reduce the number of variables to find invariant representations for data.
- Choosing the data mining task depending on the goal of KDD: clustering, classification, regression, and so forth.
- Selecting methods and algorithms to be used for searching for the patterns in the data
- Mining the knowledge: searching for patterns of interest
- Evaluating or interpreting the mined patterns, with a possible return to any previous steps.
- Using this knowledge for promoting the performance of the system and resolving any potential conflicts with previously held beliefs or extracted knowledge.

DATA MINING METHODS

The objective of data mining is both prediction and description. That is, to predict unknown or future values of the attributes of interest using other attributes in the databases, while describing the data in a manner understandable and interpretable to humans. Predicting the sale amounts of a new product based on advertising expenditure, or predicting wind velocities as a function of temperature, humidity, air pressure, etc., are examples of tasks with a predictive goal in data mining. Describing the different terrain groupings that emerge in a sampling of satellite imagery is an example of a descriptive goal for a data mining task. The relative importance of description and prediction can vary between different applications. These two goals can be fulfilled by any of a number data mining tasks including: classification, regression, clustering, summarization, dependency modeling, and deviation detection.

PREDICTIVE TASKS

The following are general tasks that serve predictive data mining goals:

• **Classification** – to segregate items into several predefined classes. Given a collection of training samples, this type of task can be designed to find a model for class attributes as a function of the values of other attributes.

• **Regression** – to predict a value of a given continuously valued variable based on the values of other variables, assuming either a linear or nonlinear model of dependency. These tasks are studied in statistics and neural network field.

• **Deviation Detection** – to discover the most significant changes in data from previously measured or normative values. Explicit information outside the data, like integrity constraints or predefined patterns, is used for deviation detection.

DESCRIPTIVE TASKS

- **Clustering** – to identify a set of categories, or clusters, that describe the data
- **Summarization** – to find a concise description for a subset of data. Tabulating the mean and standard deviations for all fields is a simple example of summarization. There are more sophisticated techniques for summarization and they are usually applied to facilitate automated report generation and interactive data analysis .
- **Dependency modeling** – to find a model that describes significant dependencies between variables. For example, probabilistic dependency networks use conditional independence to specify the structural level of the model and probabilities or correlation to specify the strengths (quantitative level) of dependencies Mixed tasks.
- **Association Rule Discovery** – Given a set of records each of which contain some number of items from a given collection, produce dependency rules which will predict the occurrence of an item based on patterns found in the data.
- **Sequential Pattern Discovery** – Given a set of objects, where each object is associated with its own timeline of events, find rules that predict strong sequential dependencies among different events. Rules are formed by first discovering patterns followed by event occurrences which are governed by timing constraints found within those patterns.

ONLINE EDUCATION SYSTEMS

Several Online Education systems such as Blackboard, WebCT, Virtual University (VU), and some other similar systems have been developed to focus on course management issues. The objectives of these systems are to present courses and instructional programs through the web and other technologically enhanced media. These new technologies make it possible to offer instruction without the limitations of time and place found in traditional university programs. However, these systems tend to use existing materials and present them as a static package via the Internet.

CONCLUSION

We introduce the basic concepts of data mining as well as information about current online educational systems, a background on Intelligent Tutoring Systems.. A body of literature has emerged, dealing with the different problems involved in data mining for performing classification and clustering upon web-based educational data. This paper positions itself to extend data mining research into web-based educational systems – a new and valuable application. Results of data mining tools help students use the online educational resources more efficiently while allowing instructors, problem authors, and course coordinators to design online materials more effectively.

Bibliography

[1] Aeberhard, S., Coomans D., and de Vel, O. (2002) “Comparison of Classifiers in

High Dimensional Settings”, Tech. Rep. no. 92-02, (2002), Dept. of Computer

Science and Dept. of Mathematics and Statistics, James Cook University of North

Queensland.

[2] Agrawal, R., Imielinski, T.; Swami A. (2003), "Mining Associations between Sets

of Items in Massive Databases", Proc. of the ACM-SIGMOD 2003 Int'l

Conference on Management of Data, Washington D.C., May 2003.

[3] Agrawal, R.; Srikant, R. (2004) "Fast Algorithms for Mining Association Rules",

ISSN 2348 – 7968

Proceeding of the 20th International Conference on Very Large Databases,

Santiago, Chile, September 2004.

[4] Agrawal, R. and Srikant. R. (2005) “Mining Sequential Patterns”. In Proceeding

of the 11th International Conference on Data Engineering, Taipei, Taiwan, March

2005.

[5] Agrawal, R., Shafer, J.C. (2006) "Parallel Mining of Association Rules", IEEE

Transactions on Knowledge and Data Engineering, Vol. 8, No. 6, December

2006.