Data Mining in Education: A Review

Authors

Raj Kumar¹, Akshita Sharma²

¹Assistant Professor, Department of CSE Jind Institute of Engineering & Technology, Jind (Haryana)
²M.Tech student, Department of CSE Jind Institute of Engineering & Technology, Jind (Haryana)

Email- bhardwajakky@gmail.com, rajshira@gmail.com

ABSTRACT:

Education mining is process of analyzing data from different perspectives & summarizing it into educational information. Information that could be used to increase revenue, cut costs, or both. In data mining each & everything becomes intelligent & perform task as human. For that purpose there are various tools, techniques & methods are proposed. Support vector machine is a model for statistics & computer science, to perform supervised learning, methods that are used to make analysis of data & recognize patterns. SVM is mostly used for classification & regression analysis. In same way k-nearest neighbor algorithm is a classification algorithm used to classify data using training examples. In this paper we have reviewed used different classification algorithm in field of education to classify & predict student data. Here we use educational nominal data to classify & discover data pattern to predict future courses, Uses to education mining which is use to classify text analysis in future.

Keywords- SVM , k-NN, Patterns, Analysis, Classification, knowledge discovery.

1. INTRODUCTION

Data mining (sometimes called data or knowledge discovery) is process of analyzing data from different perspectives & summarizing it into useful information. Data mining software is a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, & summarize relationships identified. Technically, data mining is process of finding correlations or patterns among dozens of fields in large relational databases. Data mining tools predict future trends & behaviors, allowing businesses to make proactive, knowledge-driven decisions. Automated, prospective analysis offered by data mining move beyond analysis of past events provided by retrospective tools typical of decision support systems. Data mining tools could answer business questions that traditionally were too time consuming to resolve [1].

KDD Process: The term Knowledge Discovery in Databases, or KDD for short, refers to broad process of finding knowledge in data, & emphasizes "high-level" application of particular data mining methods. It is of interest to researchers in machine learning, pattern statistics, artificial intelligence, knowledge acquisition for expert systems, & data visualization. Unifying goal of KDD process is to extract knowledge from data in context of large databases. Using data mining methods (algorithms) to extract (identify) what is deemed knowledge, according to specifications of measures & thresholds, using a database along within any required preprocessing, sub sampling, & transformations of that database.

Fig 1 An Outline of Steps of KDD Process.
Classification is defined as the process of finding a set of models (or functions) that describe and distinguish data classes & concepts, within a goal being to use models to predict classes of objects whose class labels are unknown[2]. Thus, classification is a supervised learning problem where the task is to predict value of a discrete output variable given a set of training examples & a test sample where each training example is a pair consisting of input object & desired class. Selection of the best classification algorithm for a given dataset is a very widespread problem. In this sense it requires to make several methodological choices. Educational Data Mining is an emerging field that could be applied to the field of education, it concerns within developing methods that discover knowledge from data originating from educational environments [3]. Education mining, extraction of hidden predictive information from large databases, is a powerful new technology within great potential to help companies focus on most important information in their data warehouses. Most companies already collect & refine massive quantities of data. Education mining techniques could be implemented rapidly on existing software & hardware platforms to enhance value of existing information resources, & could be integrated within new products & systems as they are brought on-line.

2. EDUCATION DATA MINING

Educational Data Mining refers to techniques, tools, & research designed for automatically extracting meaning from large repositories of data generated by or related to people learning activities in educational system settings. Quite often, this data is extensive, fine-grained, & precise. For example, several learning management systems (LMSs) track information such as when every student accessed each learning object, how many times they accessed it, & how many minutes learning object was displayed on user’s computer screen [4]. As another example, Intelligent tutoring systems record in the data every time a learner submits a solution to a problem; they may collect time of submission, whether or not solution matches expected solution, amount of time that has passed since last submission, order in which solution components were entered into interface, etc[5]. Precision of this data is such that even a fairly short session within a computer-based learning environment may produce a large amount of process data for analysis.

In other cases, data is less fine-grained. For example, a student's university transcript may contain a temporally ordered list of courses taken by student, grade that student earned in each course, & when student selected or changed his or her academic major. EDM leverages both types of data to discover meaningful information about different types of learners & how they learn, structure of domain knowledge, & effect of instructional strategies embedded within many learning environments. These analyses provide new information that would be difficult to discern by looking at raw data. For example, analyzing data from an LMS may reveal a relationship between learning objects that a student accessed during course & their final course grade. Similarly, analyzing student transcript data may reveal a relationship between a student's grade in a particular course & their decision to change their academic major [6]. Such information provides insight into design of learning environments, which allows students, teachers, school administrators, & educational policy makers to make informed decisions about how to interact with, provide, & manage educational resources.

3. k-NN IN EDUCATION MINING

In pattern recognition, k-Nearest Neighbors algorithm (or k-NN for short) is a non-parametric method used for classification & regression. In both cases, input consists of k closest training examples in feature space. Output depends on whether k-NN is used for classification or regression [7].
In \textit{k-NN classification}, output is a class membership. An object is classified by a majority vote of its neighbors, within object being assigned to class most common among its \(k\) nearest neighbors (\(k\) is a positive integer, typically small). If \(k = 1\), then object is simply assigned to class of that single nearest neighbor.

In \textit{k-NN regression}, output is property value for object. This value is average of values of its \(k\) nearest neighbors.

4. SUPPORT VECTOR MACHINE (SVM) IN EDUCATION MINING

Support Vector Machine is a supervised machine learning algorithm which could be used for both classification & regression challenges in educational sector. However, it is mostly used to classification problems. In algorithm, we plot each Educational mining data item as a point in \(n\)-dimensional space (where \(n\) is number of features you have) within value of each feature being value of a particular coordinate [8]. Then, we perform classification by finding hyper-plane that differentiates two classes very well. Support Vectors are simply co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates two classes (hyper-plane/line).

5. CONCLUSION & FUTURE SCOPE

The various data mining techniques could be effectively implemented on educational data. From results it is clear that classification techniques could be applied on educational data for predicting student’s outcome & improve their results. Efficiency of various classification algorithms could be analyzed based on their accuracy & time taken to drive result. Since application of data mining brings a lot of advantages in higher learning institution, these techniques could be applied in other areas of education to optimize resources, to predict number of students who are likely to get a placement.

In future we use different classification algorithms for text analysis or web contains mining over educational data analysis.

REFERENCES

6. Lloyd-Williams, M. —Case studies in the data mining approach to health information analysis, Knowledge Discovery and Data Mining (1998/434), IEEE Colloquium on, 8May1998.