Data Mining Techniques for Maintenance of Instances for Universities

J.Kishore Kumar[1],Dr A.Ravi Prasad[2], S.Ramakrishna[3]

[1,2]*S.G.Govt. Degree College, Piler, India*
[3]*S.V.University,Tirupathi,India*

Abstract-Knowledge discovery in education organization have been built and operated mainly to support decision making using knowledge as strategic factor. In this paper, we investigate the use of various data mining techniques for knowledge discovery in education organization. Existing software are inefficient in showing such data characteristics. We introduce different exhibits for discovering knowledge in the form of association rules, clustering, classification and correlation suitable for data characteristics. Proposed data mining techniques, the decision- maker can give a clear idea of whether an organization is eligible for continuing its services or it has to be dropped.

Keywords-Result, Association rules, Clustering, Classification, Correlation, Data mining.

INTRODUCTION

Presently in some places due to difficulty of maintenance of large number of students in a single university, its grants permission for colleges to follow same rules and regulations. These colleges are known as affiliated colleges which can be said to be instances of university which not only follow the rules and regulations but also follow syllabus.

But there are cases such that the affiliated colleges exist but not possess the minimum requirements, which not only affect student's life but also cheat parents, universities and society. Such situation can be found out by using Data mining techniques.

Data mining can be defined as the process of selecting, exploring and modelling large amounts of data to uncover previously unknown patterns. In the organization, data mining can help university gain advantage. For example, by applying data mining techniques, university can fully exploit data about instance as well as gaining a greater understanding of their loop holes if exist in rules and regulations and to help reduce fraud, improve underwriting and enhance risk management.

This paper discusses how education organizations can benefit by using modern data mining methodologies and thereby reduce risk, increase quality, retain current customers and develop new rules and regulations. Data mining methodology often can improve upon traditional statistical approaches to solving business solutions. Specifically, data mining can help university firms in quality improving by practices such as:
• Acquiring new colleges.
• Performing sophisticated classification.

Acquiring New Colleges

An important requirement for educating more students is acquisition of new colleges. A traditional approach is to increase the number of instances that meet policy constraints. A drawback to this approach is that increase in number of colleges will reduce quality of education.

Hence in this situation it is important to identify those colleges among already existing colleges which does not possess minimum requirements .For identifying this situation, a statistical technique called "cluster analysis," is used. Clustering is a technique of partitioning or segmenting the data into groups that might or might not be disjointed. The clustering usually accomplished by determining the similarity among the data on predefined attributes. The most similar data are grouped into clusters. Since clusters are not predefined, a domain expert is often required to interpret the meaning of the created clusters.

Definition:
Given a database $D = \{t1,t2,\ldots,tn\}$ of tuples and an integer value k, the clustering problem is to define a mapping
$f : D \_ \{1,\ldots,k\}$where each ti is assigned to one cluster kj, $1\_ j \_k$.A cluster kj, contains precisely those tuples mapped to it that is,$kj = \{ ti \mid f(ti) = kj, 1\_ i \_ n, and ti\_ D\}$

Algorithm 1.1 k-means Clustering
K-means is an iterative clustering algorithm in which items are moved among sets of clusters until the desired set is reached

Input:
D= {t1, t2, t3,...,tn} //Set of elements
k //Number of desired clusters

Output:
K //set of clusters

Algorithm:
assign initial values for means m1, m2, …, mk;
repeat
assign each item ti to the cluster which has closest mean;
calculate new mean for each cluster;
until convergence criteria is met.

Note that the initial values for means are arbitrarily assigned and the algorithm could stop when no or very small number of tuples are assigned to different clusters. As per the algorithm, first we have to find mean of each cluster. If there is any average difference then place that element to closest cluster.

Classification: Segment Databases

To improve predictive accuracy, databases are segmented into homogeneous groups. Then the data of each group can be explored, analysed and modelled. Depending on the quality improving questions, segmentation can be done using variables associated with requirements or risk factors. Segments based on these types of variables often provide sharp contrasts, which can be interpreted more easily. Classification maps data into predefined groups or segments.

Classification algorithms require that the classes be defined based on data attributes values. They often describe these classes by looking at the characteristics of data already known to belong to the classes. As a result, university can more accurately predict the likelihood of a quality and requirements of college.

Definition

Given a database D = {t1,t2,…,tn } of tuples (items, records) and a set of classes C = {C1,…,Cm}, the classification problem is to define a mapping f : D _ C where each it is assigned to one class. A class Cj, contains precisely those tuples mapped to it that is, Cj = { ti | f(ti) = Cj, 1_ i _ n, and ti_ D}

Algorithm 1.2 K Nearest Neighbour

When classification is to be made for new item using K Nearest neighbour algorithm, its distance to each item in the training set must be determined. The new item is then placed in the class that contains the most items from the (K) closest set.

Input:

T //Training data
K //Number of neighbours
t //Input tuple to classify

Output:

c //class to which t is assigned

Algorithm:

N = ø
//Find the set of neighbours, N, for t
For each d _ T do
If |N|_ K, then
N = N _ {d};
else
if u _ N such that sim(t,u) _ sim(t,d), then
begin
N = N – {u};
N = N _ {d};
end
//Find class for classification
C=class to which the most u _ N are classified;

For example, if there are three groups like first group is college with minimum requirements, like minimum infrastructure minimum faculty, minimum transportation and minimum laboratory. second group of college is with medium requirements like medium infrastructure, medium faculty ,medium transportation, medium laboratory and third group of college is with good infrastructure, good faculty, good transportation, good laboratory then when a college is having good infrastructure, medium laboratory, medium faculty, medium transportation then that college will come under second
group i.e., medium college.

CONCLUSION

Data mining can help universities gain advantage mainly to support decision making. The university needs to know the essentials of decision making and data mining techniques to compete in the quality education.