Educational Data Mining in Chin State

U Moe Thant

Lecturer, University Of Computer Studies, Kalay

ABSTRACT

Now a day Educational data mining is new emerging technique of data mining that can be applied on the data related to the field of education. In this proposed method si Based on the clustering methods such as centroid based, distribution based and density based clustering. Cluster includes groups in with small distance among the cluster members. The student's school requirement level of optimization is formulated by using clustering. In centroid based clustering, clusters are represented by a central vector. The number of clusters is fixed to k, k-means clustering gives a formal definition as an optimization problem. The clustering model most closely related to statistics is based on distribution model. Experiments attempts to improve the accuracy by using the method of data mining using R Tool.

Keyword : *cluster,k-means*

1.INTRODUCTION

The main objective of primary education institutes is to provide education development to Chin State rural area and to improve the quality of managerial decisions. One way to achieve progress of school buildings and their students easily go to school in primary education system is by discovering knowledge from educational data to study the main attributes that may affect the education performance [3]. The discovered knowledge can be used to offer a helpful and constructive recommendations to the academic planners in primary education institutes or School to enhance their decision making process, to improve primary school building and where area to build school building to provide decision maker to improve rural situation and many other benefits Improved educational data mining uses many techniques such as centroid based, and density based clustering. Cluster includes group with small distance among the cluster of Group members.

2. METHODS AND MATERIAL

Data Clustering is unsupervised and statistical data analysis technique. It is used to classify the same data into a homogeneous group of primary school students it is used to operate on a large data-set to discover hidden pattern and relationship helps to make decision quickly and efficiently. Cluster analysis is used to break down a large set of data into subsets called clusters. Each cluster is a collection of data objects that are similar to one another. They are placed within the same cluster but are dissimilar to objects in other clusters. Following algorithms are used in education mining in Clustering.

2.1 K-Mean Clustering Algorithm

K-Means is a non-hierarchical clustering method that seeks to partition the data into the form of one or more clusters [2, 6]. This method partitions the data into clusters so that the data having the same characteristics are grouped into one cluster and the data that have different characteristics grouped into another cluster.

2.2 K-Mean Clustering Algorithm Steps

Step 1: Input. Dataset, Clustering Variables and Maximum Number of Clusters (K in Means Clustering)
Step 2: Initialize cluster centroid.
Step 3: Calculate Euclidean Distance. Euclidean is one of the distance measures used on K Means algorithm.

3.EXCEPTATION MAXIMIZATION (EM) ALGORITHM

Expectation Maximization Algorithm (EM-clustering) is used to cluster the given data. An EM algorithm is a mixture based algorithm that finds maximum likelihood estimates of parameters in probabilistic models. EM-clustering to group students according to their attend school township [4, 6]. It gives Mean of each cluster for each primary attribute. Using these results we can divide students into groups. Expectation Maximization Algorithm showed how useful in educational data mining. It can be in primary education particularly to improve adult education. The proposed system used education data from database result improvement. Proposed system collected all available data including their usage of our proposed area study of Chin State. This system applied data mining techniques. Also the system clustered the student into group using EM-clustering. Each one of this knowledge can be used to improve the performance of adult education.

4. DATA SET AND ATTRIBUTE SELECTION

This data set consist 140 instances and each instance consists of 6 attributes in Table 1. Initially data is collected in excel sheet.

Fig -1 Education Data of Chin State

Table -1 Attribute Description

Attribute	Description
Children_<16yearsages	Children number
Children_gotoschool	Students number
Children_Female	Female student number
Children_Attendschool_OutsideTownshsip	Other Township's students
Where_Childrengoto_School1	School location
Class	Home town or Outside

5. K-means CLUSTERING ALGORITHM BASED ON SCHOOL LOCATION

The proposed system clustered the data based on the student go to school location. It means that the students in the school are near home or far home.
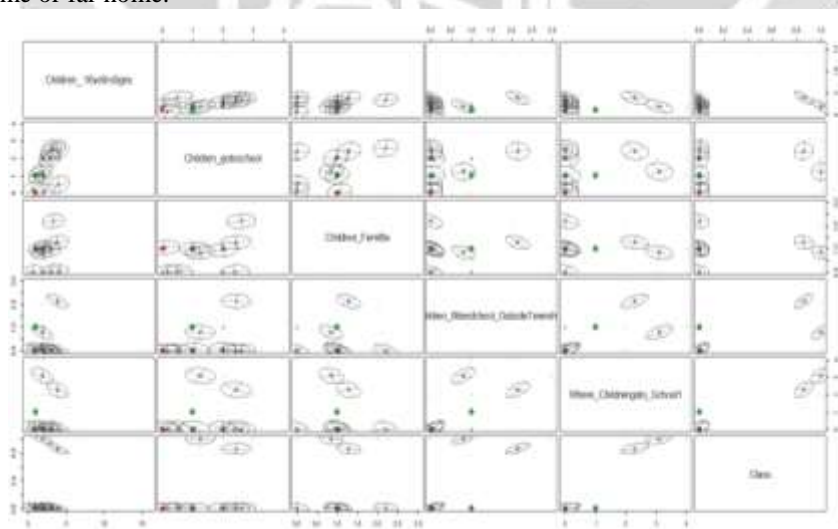

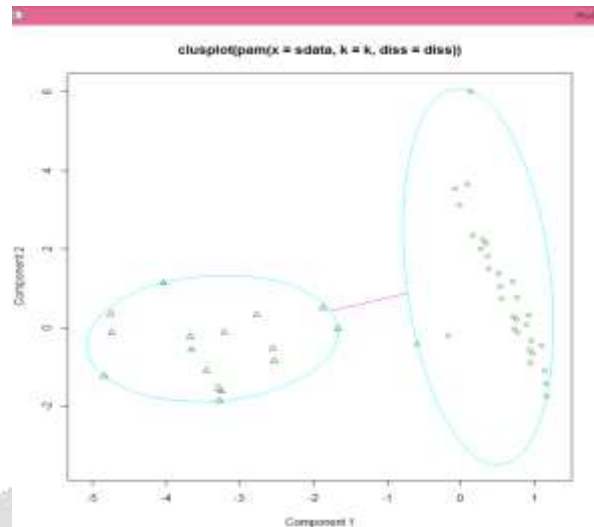**Fig -2 Education Data of Chin State**

Fig -3 These two components explain 78.68% of the point variability

Table – 2 K-means and EM algorithm (Accuracy)

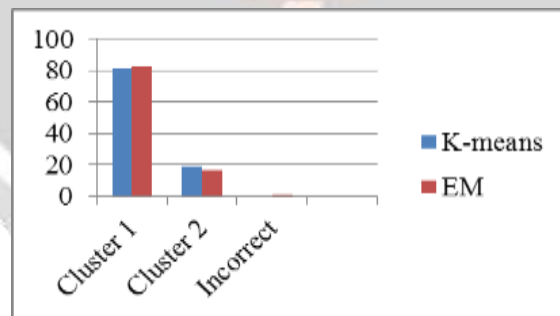|	Cluster 1	Cluster 2	Incorrect Cluster
K-means	81%	19%	0.0%
EM	83 %	17%	1.4%

Fig -4 Performance Compression of K-means and EM algorithm

   Data objects as initial centroids and applying the k- means algorithm and EM Algorithms. The proposed system obtained the following results from the education data set consisting of 140 instances. Only 1.4 percentages of instances incorrect cluster category when the system used EM algorithm. In cluster 2, 19 percentages of the students and 17 percentages cluster 2 category of (Children attend school outside Township), 81 percentages of the students attendance and 83 percentages attendance of student to their hometown school fall in cluster 1 the category of (Children go to school in Township). Analyze the problem of education data and overcome the lacking points so that the a few students go to school outside of their Township. At that time decision maker performance not need well quickly to build school in rural area. In the future school building requirements overcome the lacking points.

6. CONCLUSIONS

This above study shows that data mining will be considered most useful in educational field. Predicting student's academic building is of great concern to the education institutes. By applying data mining techniques and tools in clustering of student attended location is helpful to identify academic building requirement. K-means and EM algorithm are used for clustering. K-means provide 100 % accuracy and EM Provide 99% accuracy.

REFERENCES

[1] Ayesha S, Mustafa T, Sattar A.R., and.Khan, M.I 2010 Data Mining Model for Higher Education

[2] B.K.Baradwaj and SaurabhPal, 2011 "Mining Educational Data to analyses Student's performance", (IJACSA) International Journalof Advanced Computer Science and Applications, Vol. 2, No. 6, 2011, p 63-69.

[3] Banerjee, Abhijit, R. Banerji, E. Duflo, and M.Walton. 2012. Effective Pedagogies and a Resistant Education System:Experimental Evidence on Interventions to Improve Basic Skills in Rural India. MIT

[4] C. M.Vera, C.R.Morales, and S. V.Soto. 2013, "Predicting School Failure and Dropout by Using Data Mining Techniques" , IEEE Journal Of Latin-American Learning Technologies, Vol. 8, No. 1, Feb 2013 .

[5] Delavari, N.Beikzadeh, M.R. (2004). A new model for using data mining in higher education system, 5th international Conference on Information Technology based Higher education and training: ITEHT (04), Istanbul, Turkey, 31st May-2nd June 2004.

[6] Dréze, Jean and Kingdon, Geeta, 2000. "School Participation in Rural India," Review of Development Economics.

[7] James, Estelle, E.M.King, and A.Suryadi. 1996. "Finance, Management, and Costs of Public and Private Schools in Indonesia." Economics of Education Review, 15(4), pp. 387-98

[8] Jayaraman, Rajshri, D. Simroth, and F.D. Vericourt. 2010. The Impact of School Lunches on Primary School Enrollment: Evidence from India's Mid- Day Meal Scheme. Indian Statistical Institute.

[8] G.Abu, D, and S. Klasen (2004), The Costs of Missing the Millennium Development Goal on Gender Equality." World Development 32 (7): 1075–107

[9] Han, J.,Kamber, M. (2006). Data mining: Concepts and techniques (2nd ed.). Boston, MA: Elsevier.

[10] J.Das,S.,Dercon , J.Habyarimana ,P.Krishna ,K.Muralidharana and V.Sundararaman 2013 " School Input, hose hold substitution and test scores" AmericanEconomic Journal Applied Economics 5(2), p-29-29.

[11] Johnstone, James and Jiyono. 1983. "Out-of-school Factors and Educational Achievement in Indonesia." Comparative Education Review, 27(2), pp. 278-295.

[12] J. Han and M. Kamber, 2008 "Data Mining Concepts and Techniques", Second Edition Morgan Kaufmann Publishers.

[13] Khan Z, Khan N 2005. "Scholastic Achievement of Higher Secondary Students in Science Stream", Journal of Social Sciences, Vol. 1, No. 2, pp. 84-87.

[14] Kotsiantis, S. Pierrakeas, C. Pintelas, P. 2004. Preventing student dropout in distance learning systems using machine learning techniques, In International Conference on Knowledge-Based Intelligent Information and Engineering Systems, Oxford, 3-5.

BIOGRAPHIES (Not Essential)

U MOE THANT
LECTURER
UNIVERSITY OF COMPTER STUDIES, KALAY