

Implementation clustering approach for prediction of Academic Performance

Anil Kumar Pandey¹, Sachin Saxena²

^{1,2}Computer science & Engineering, Invertis University Bareilly

Abstract- Today finding the student's academic performance is a most important issue for a education system. An education system for analyzing student's on groups based are give better quick result according to the level of their performance as compare to individual . In this paper we purpose an approach is combined with the deterministic clustering to analyze and groups the given student's list. Here we apply clustering approach to evaluate academic performance of students.

Keywords-K-Mean clustering, Deterministic model , education.

I. INTRODUCTION

k-means clustering is a method of combo of items into k groups. The grouping is done by minimizing the sum of squared distances (Euclidean distances) between items and the centroid[1]. k-means clustering is a method of vector division, basically from signal processing, that is allowed for cluster search in data mining. k-means grouping aims to partition n view into k clusters in which each view match to the cluster with the close mean, plate as a imago of the cluster[2]. Cluster search is a task that seek to find same class of object and it is also one of the main method in data mining. K-mean is the famous section clustering method[3]. Clustering is a action of combo of data objects into divided clusters so that the data in the clone cluster are kin, but data belong to different cluster vary[4]. A cluster is a group of data object that are similar to one more are in same cluster and unlike to the objects are in other clusters[5]. Cluster search of data is an basic task in knowledge detect and data mining. Cluster search goal to group data on the basis of parallel and difference among the data items. The case can be performed in a managed, semi-supervised or freely address[6].

Clustering is a process in which a group of untagged patterns are partitioned into a number of sets so that akin patterns are select to the same cluster, and various patterns are assigned to different clusters. [7]. There are two aims for a clustering algorithms: certain good clusters and doing so expeditiously. Cluster analysis method is one of the most rational methods of data mining. The method will directly power the consequence of clustering[8]. K-mean clustering and see the lack of standard k-means such as k-means algorithm calculates span of each data point from each cluster sentry[9]. Clustering style are large methods for the audit of data, clear thinking based on the audit and for drain the disagreement observed in them. Grouping is unsupervised lore and do not rely on pretend classes[10]. In clustering we part the varied between item by grading the way between each pair of item. These part include the Euclidean, Manhattan and Minkowski distance[11]. K-mean is the most approved dividing method of clustering. It was firstly pressed by MacQueen in 1967[12]. K mean is a one by one, non-deterministic, numerical, together method of clustering. In k-mean each cluster is tag by the tight cost of item in the cluster. Here we share a set of n object into k cluster so that intercluster parallel is low and intracluster likes is high[13]. Similarity is uniform in term of mean value of item in a cluster[14]. Clustering techniques are broad methods for the audit of data, forecast based on the audit and for eliminating the breach attended in them. Iterative techniques are used to group datas which forms part of a cluster as per parallel and like attribute[15].

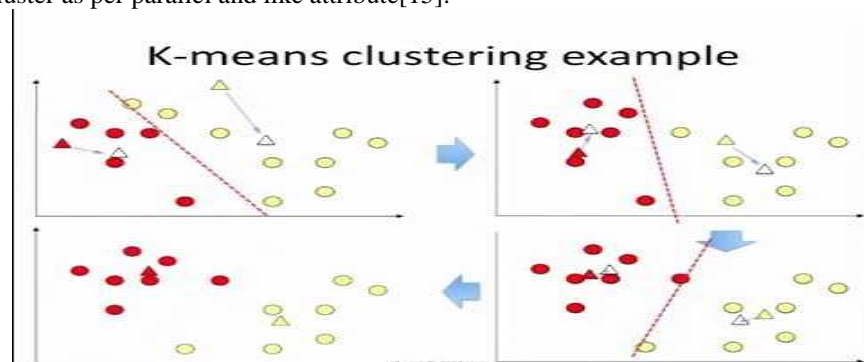


Fig:1 Kmean Clustering

Academic achievement represents performance outcomes that indicate the size to which a person has hep specific aims that were the focus of action in instructional environments, specifically in school, college, and university. School systems mostly define cognitive aims that either apply beyond multiple subject areas include the heritage of knowledge and understanding in a specific rational domain. Therefore, academic achievement should be examined to be a varied build up that comprises various domains of learning. Because the field of academic achievement is very wide-ranging and covers a broad variety of educational outcomes, the definition of academic achievement trust in on the indicators used to part it. Base on this hypercritical issue, layout of students into various class allow to their performance has become a complicated task. To calculate the Academic Performance of a class. We first find the following values:

With classic form of students based on their average scores, it is difficult to obtain a comprehensive view of the state of the students' performance and simultaneously discover large details from their time to time performance. With the help of data mining methods, such as clustering algorithm, it is possible to discover the key tone from the students' performance and possibly use those tone for future prediction. There have been good results from applying K-Mean clustering with the Euclidean distance measure, where the distance is computed by data the square of the distance between each scores, summing the squares and data the square root of the sum[16]. K-means clustering is a type of unattended learning, which is used when you have unlabeled data (i.e., data without defined categories or class). The goal of this algorithm is to find class in the data, with the number of class represented by the variable K. This paper presents K-Mean clustering algorithm as a simple and efficient tool to monitor the progression of students' performance in higher institution. Grouping is a well known data mining technique which is used to group together data items based on similarity property.

II. K-MEANS CLUSTERING APPROACH FOR ACADEMIC PERFORMANCE:

k-means clustering is a method of vector division originally from signal refine, that is popular for cluster analysis in data mining. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a model of the cluster. K-Means is one of the simplest unsupervised learning algorithms used for clustering. Given D, a data set of n objects, and k, the number of clusters to form, a partitioning algorithm fit the objects into k partitions ($k \leq n$), where each partition represents a cluster.

The clusters are formed to advance an detached wall criterion, such as a difference function based on distance. Flow chart of K-Mean clustering of objects given blew in Fig2

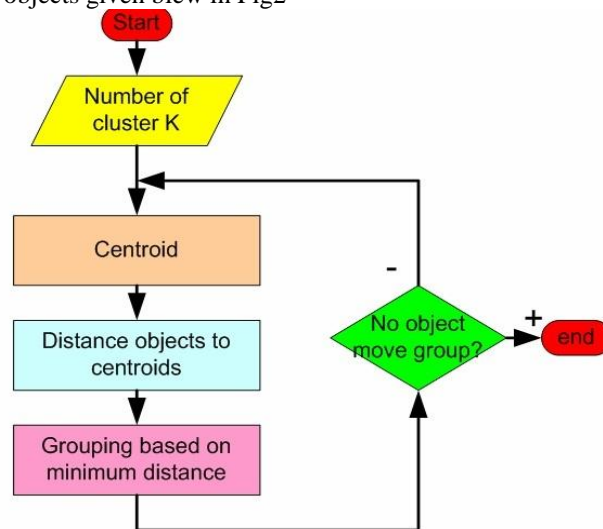


Fig:2:Kmean Clustering object flow

There are different steps used for find the group into given different objects. As per given above flow chart the K mean clustering algorithm define as

Step 1: input the number of clusters for grouping data.

Step 2: divide the first K clusters

Step 3: Take first k case or Take Unplanned pick of k elements

Step 4: Calculating means of each group formed in the dataset.

Step 5: K-means shell out each record in the dataset to only one of the initial clusters
 - Each record is allotted to the nearest cluster using a measure of distance.

Step 6: K-means search each log in the dataset to the most similar cluster.

Example of Find the Academic Performance of a given classes:

$K = \{55, 60, 65, 66, 70, 75, 77, 78, 84, 88\}$

$K = 4$

Mean(M1)=66 M2=75

$K1 = \{55, 60, 65, 66, 70\}$

M1=63

$K1 = \{55, 60, 65, 66, 70\}$

$K2 = \{75, 77, 78, 84, 88\}$

M2=80

$K2 = \{75, 77, 78, 84, 88\}$

Again applying K-Mean on K1 cluster

$K1 = \{55, 60, 65, 66, 70\}$

M1=60 M2=66

$C4 = \{55, 60\}$

M1=57

$C4 = \{55, 60\}$

$C3 = \{65, 66, 70\}$

M2=67

$C3 = \{65, 66, 70\}$

Applying K-Mean on K2 cluster

$K2 = \{75, 77, 78, 84, 88\}$

M1=77

$C2 = \{75, 77, 78\}$

$K1 = \{2, 3, 4, 10, 11, 12\}$

M1=77 M2=86

$C2 = \{75, 77, 78\}$ $C1 = \{84, 88\}$

M2=84

$C1 = \{84, 88\}$

$K2 = \{20, 25, 30\}$

Thus 4 clusters are:

$C1 = \{84, 88\}$

$C2 = \{75, 77, 78\}$

$C3 = \{65, 66, 70\}$

$C4 = \{55, 60\}$

C1	Excellent
C2	Very Good
C3	Good
C4	Very Fair

Table 1: Performance index

IV. DISCUSSION AND CONCLUSION

In this study we form use of data mining process in students data base applying K-Means point method in clustering algorithms and decision tree testing to analyze and mend the aspect of engineering study. The bosses can use some art to boost the course outcomes to pick up the knowledge. We hope that the advice hatch after the rub out of data mining and data clustering technique may be helpful for the teacher as well as for students. This work may be proper for teacher to compare and greater the students act; reduction failing ratio by taking right steps at proper time to fix the quality of study. For future work, we hope to better the technnique in form to get more blue-chip and true outputs, useful for teachers to fix the students information outcomes.

V. REFERENCES

- [1] kedar b. sawant "Efficient Determination of Clusters in K-Mean Algorithm Using Neighbor hood Distance" International Journal of Emerg ing Engineering Research and Technology Volume 3, Issue 1, January 2015
- [2] ahamed shafeeq b m 1 and hareesha k s 2 "Dynamic Clustering of Data with Modified K-Means Algorithm" 2012 International Conference on Information and Computer Networks.
- [3] ming-chuan hung, jungpin wu+, jin-hua chang and don-lin yang "An Efficient k-Means Clustering Algorithm Using Simple Partitioning*"journal of information science and engineering 21, 1157-1177 (2005)

- [4] Sadhana Tiwari, Tanu Solanki “ An Optimized Approach for k-means Clustering” 9th International ICST Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness (QShine-2013)
- [5] Cosmin Marian Poteras, Marian Cristian Mih¹aescu, “An Optimized Version of the K-Means Clustering Algorithm” Proceedings of the 2014 Federated Conference on Computer Science and Information Systems.
- [6] Oyelade, O. J, Oladipupo, O. O, Obagbuwa, I. C “Application of k-Means Clustering algorithm for prediction of Students’ Academic Performance” (IJCSIS) International Journal of Computer Science and Information Security.
- [7] Deepali Virmani¹,Shweta Taneja²,Geetika Malhotra³ “ Normalization based K means Clustering Algorithm” Department of Computer Science,Bhagwan Parshuram Institute of Technology,New Delhi.
- [8] Kalpit G. Soni^{1,2} and Dr. Atul Patel³ “ Comparative Analysis of K-means and K-medoids Algorithm on IRIS Data” International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 13, Number 5 (2017)
- [9] K. Alsabti, S. Ranka, and V. Singh, “An efficient k-means clustering algorithm,” in Proceedings of 1st Workshop on High performance Data Mining, 1998.
- [10] D. L. Yang, J. H. Chang, M. C. Hung, and J. S. Liu, “An efficient k-means-based clustering algorithm,” in Proceedings of 1st Asia-Pacific Conference on Intelligent Agent Technology, 1999, pp. 269-273
- [11] Y. M. Cheung, “k*-means: a new generalized k-means clustering algorithm,” Pattern Recognition Letters, Vol. 24, 2003, pp. 2883-2893.
- [12] Z. Huang, “Extensions to the k-means algorithm for clustering large data sets with categorical values,” Data Mining and Knowledge Discovery, Vol. 2, 1998, pp. 283-304.
- [13] Tapas Kanungo, Senior Member, IEEE, David M. Mount, Member, IEEE, Nathan S. Netanyahu, Member, IEEE, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu, Senior Member, IEEE “An Efficient k-Means Clustering Algorithm: Analysis and Implementation” IEEE transactions on pattern analysis and machine intelligence, vol. 24, no. 7, July 2002.
- [14] Unnati R. Raval¹, Chaita Jani² “Implementing & Improvisation of K-means Clustering Algorithm” IJCSMC, Vol. 5, Issue. 5, May 2016.
- [15] Jyoti Yadav^{#1}, Monika Sharma^{*2} “A Review of K-mean Algorithm” International Journal of Engineering Trends and Technology (IJETT) – Volume 4 Issue 7- July 2013
- [16] Khaled Alsabti, Sanjay Ranka, Vineet Singh” An Efficient K-Means Clustering Algorithm”
- [17] Unnati R. Raval, Chaita Jani “Implementing and Improvisation of K-means Clustering” IJCSMC, Vol. 4, Issue. 11, November 2015
- [18] Ahamed Shafeeq B M, Hareesha K S “Dynamic Clustering of Data with Modified K-Means Algorithm” International Conference on Information and Computer Networks vol. 27, pp.221-225, 2012
- [19] Shi Na, Liu Xumin, Guan yong “Research on k-means Clustering Algorithm” Third International Symposium on Intelligent Information Technology and Security Informatics,IEEE 2010, pp. 63-67.
- [20] D. Napoleon, P. Ganga lakshmi “An Efficient K-Means Clustering Algorithm for Reducing Time Complexity using Uniform Distribution Data Point” IEEE 2010, pp. 42-45.
- [21] Al-Daoud, M. B., Venkateswarlu, N. B., and Roberts, S. A. Fast K-means clustering algorithms. Report 95.18, School of Computer Studies, University of Leeds, June 1995.
- [22] Kanungo, T., Mount, D. M., Netanyahu, N., Piatko, C., Silverman, R., and Wu, A. The efficient K-means clustering algorithm: analysis and implementation. IEEE Trans. Pattern Analysis Mach. Intell. 2002, 24(7), 881–892.