

Man Power Allotment In Industry Sector Using Kmean Algorithm Special Reference On Rane Trw Steering Systems

P.Anitha Vairamany, Dr.K.Subramaniyan

Research Scalar, Bharathiyar University,Coimbatore

Professor,Department Of Computer Science, Govt Arts College Of Arts And Science.Pudukkottai.

Abstract

Staff allotment is a crucial job in industry sector. Human resource plays leading rule to achieve success of the industry. The Data mining is analytical method to allot the right job for right man. This is not easiest job to match the work schedule in large industrial sector. There is a large databases have to be maintain each and every department. There must be a need for perfect knowledge discovery model to retrieve the information about the staff. From that the management can make perfect schedule. The data mining approaches will make suitable to make the work schedule. how can be Data mining techniques and k mean algorithm process the effective job will be the theme of this article.

Data mining may be regarded as an evolving approach to data analysis in very large databases that could become a useful tool to management professionals. Data mining involves extracting knowledge based on patterns of data in very large databases. Yet, data mining goes beyond simply performing data analysis on large data sets. Organizations that employ thousands of employees and track a multitude of employment-related information might find valuable information patterns contained within their databases to provide insights in such areas as employee retention and compensation planning. To develop the staff planning and allotment, the k mean clustering algorithm can be used for this job. K mean is a method is popularly 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 nearest mean serving as prototype of cluster. k mean algorithm is can be grouped employees as a different cluster with nearest mean.

Key words: kmean cluster formations. Set theory.

Introduction:

The most distinct characteristic of data mining is that it deals with very large and complex data sets (gigabytes or even terabytes). The data sets to be mined often contain millions of objects described by tens, hundreds or even thousands of various types of attributes or variables (interval, ratio, binary, ordinal, nominal, etc.). This requires the data mining operations and algorithms to be scalable and capable of dealing with different types of attributes.

However, most algorithms currently used in data mining do not scale well when applied to very large data sets because they were initially developed for other applications than data mining which involve small data sets. In terms of clustering, we are interested in algorithms which can efficiently cluster large data sets containing both numeric and categorical values because such data sets are frequently encountered in data mining applications. clustering large data sets or can handle large data sets efficiently but are limited to numeric attributes. K means is the one of the un supervised analysis. Which aims to partition n observations into k clusters in which each observation belongs to cluster with nearest mean.

Literature review:

The term k means was first used James MacQueen in 1967. The standard algorithm was first proposed by Stuart Lloyd used this pulse code modulation Bell Labs used this from 1982 to still now. More efficient version proposed and published in Fortran by Hartigan and Wong in 1975/79. In 1957.

DBSCAN (Ester et al., 1996) and BIRCH (Zhang et al., 1996). These algorithms are often revisions of some existing clustering methods. By using some carefully designed search methods (e.g., randomised search in CLARANS), organising structures (e.g., CF Tree in BIRCH) and indices (e.g., R⁺-tree in DBSCAN), these algorithms have shown some significant performance improvements in clustering very large data sets. Again, these algorithms still target on numeric data and cannot be used to solve massive categorical data clustering problems.

Existing work:

The need for research with respect to the k mean is used in this work. These algorithms with objectives and methodologies have been stated in broad way in this research. This paper describes explains the broader way of knowledge extraction from k mean . in the previous system have not clear idea about cluster . Key management and Origin of the seed(employee detail).The key is more essential that is the mean of the clusters. Using this we can easily form the Cluster .

Need of proposed work:-

Search of relevant records or similar data search is a most popular function of database to obtain knowledge. There are certain similar records that we want to fall in one category or form one cluster. Query redirection is one of the good approaches to retrieve data from different databases on different servers.

K mean suits Right manpower to Right job:

The k-means algorithm is well known for its efficiency in clustering large data sets. However, working only on numeric values prohibits it from being used to cluster real world data containing categorical values. Here k mean used to categorise into employee data set in top different clusters like whole employees section employees ,Permanent employees ,CAP (company apprentice Trainee Programmer) Using this cluster how the machine can run without stop. Every Production System have to be operated a full day without stop .The section employee who is the permanent employees of the company. They are well trained to operate

The system. Some time there must be lack of trained people and if they went for a holiday The CAP should be will that gap instead of trained employee. For that k mean clustering is used to allot the job.

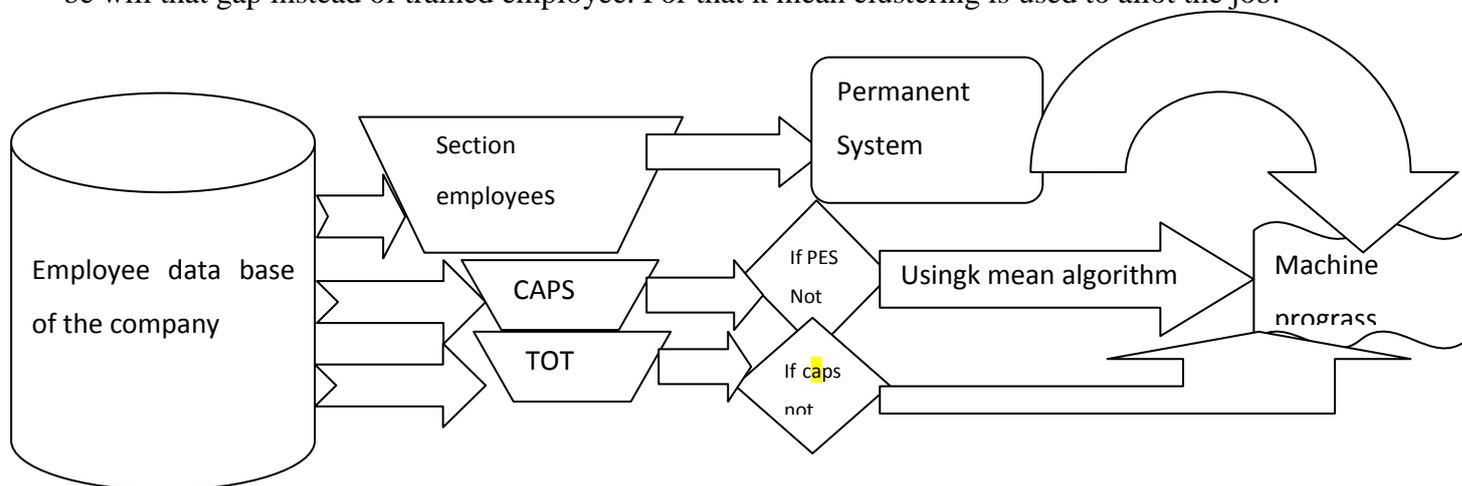


Figure shows the progress of the system

Extraction of right employee for right job using k means clustering Algorithm.

In this approach we define k sets one of the each cluster $k \leq n$. The next step is to organise data in appropriate data set and associate it to nearest set. here again recalculate knew sets. As the loop has been generated until right solution to get.

1. Place Initial Actual number of Employees as a cluster under one mean that is k mean .
2. Then place the permanent employee in another cluster using same kmean in before step. That means the have same or nearest the actual previous mean. Name the cluster Pset

3. place again new cluster CAP (Company apprentices programmer) this cluster has different data That cluster name is C set. Then reform the CAP(Company apprentices programmer) new cluster in to permanent employee kmean. That means who know the particular job that man power pick out from the cluster then they will form new cluster.

$CAP_{set} \cup P_{set}$ then merge or join
 $|CAPS| \cup |Regular|$

$Caps_i < Regular_j$

$$j = \sum_{i=1}^k \left(\sum_{i=1}^n |CAPS_i| \cup |Regular_j| \right)$$

4. Loop have will be generated in the cluster.

5. Using First in First out the p set data will be allotted that means the permanent employee will be allotted for the job. when the actual number cannot be satisfied the next CAP(Company apprentices programmer) will be allotted.

6. this procedure will be repeated until the cluster coverage.

Explanation of the Algorithm:

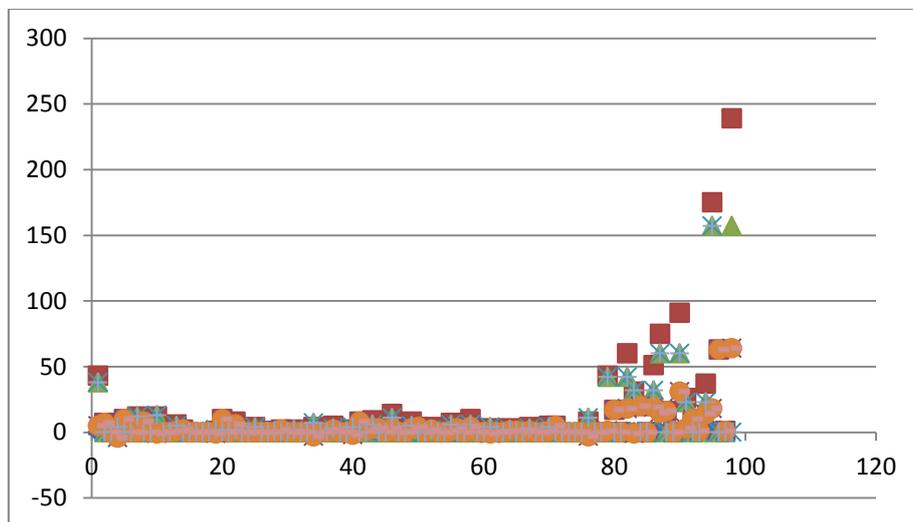
Here we first group or form the set or cluster using particular point or kmean. First group whole employees of the section. Then form another cluster for permanent employees. for example group permanent employees as who belongs to particular section. How they are efficient to run particular program in the machine. Then form the next cluster for CAPS. Again reform the CAPs group using the same k mean for permanent employee. that means who know to run a program in a particular machine. Join or merge this group in to pset than the loop will be continued according to the actual need for the section. Using First in First out Permanent Employee will be allotted for the job The CAPS will be allotted. In the following table describes over all need of the proposed work. first we form a rough cluster or approximate number of seeds (that means how many men need for the job). Then divide in clusters for the job. There must be gap between need and regular employees cluster. Then we need CAPs to fill the gap. Again cluster reformed using with CAPS cluster. Then find a Centetriad Of the cluster then reform the cluster gain this cluster much match need of the plan. Repeat the process until we get the process.

FIG - MFG Man power status @							11-Jul-15			
CELL	GLOBAL PLAN	LINE	CATEGORY	CURRENT			AFTER 7 DAYS		AFTER 14 DAYS	
				PLAN	ACTUAL	GAP	ACTUAL	GAP	ACTUAL	GAP
ASSY	85	Assy Comm	Regular	43	38	5	38	5	38	5
			Caps	7		7	7	7	7	
			Tot							
		A/c - Assy	Regular		4	-4	4	-4	4	-4
			Caps	10		10	10	10	10	
			Tot							
A	69	64 Hsg	Regular	12	12		12		12	
			Caps	8		8	8	8	8	
			Tot							
		54 Hsg	Regular	12	13	-1	13	-1	13	-1

		Vhsg & Covers	Caps							
			Tot							
			Regular	6	5	1	5	1	5	1
		Turning	Caps	2		2		2		2
			Tot							
			Regular							
		A/c - a	Caps							
			Tot							
			Regular	1	2	-1	2	-1	2	-1
		T54	61	O/S - soft	Caps	10		10		10
Tot										
Regular	8				2	6	2	6	2	6
G/W	Caps									
	Tot									
	Regular			4	4		4		4	
54 R/p	Caps									
	Tot									
	Regular			1	1		1		1	
Drive ring	Caps			2		2		2		2
	Tot									
	Regular	2	2		2		2			
O/S - Hard	Caps									
	Tot									
	Regular	4	7	-3	7	-3	7	-3		
T54 - Valves	Caps									
	Tot									
	Regular	5	3	2	3	2	3	2		
A/C - T54	Caps									
	Tot									
	Regular	2	4	-2	4	-2	4	-2		
B	155	R/P - soft	Caps	8		8		8		8
			Tot	1		1		1		1
			Regular	9	6	3	6	3	6	3
		R/P - hard	Caps	3		3		3		3
			Tot							
			Regular	14	11	3	11	3	11	3
		S/S - soft	Caps							
			Tot							
			Regular	8	5	3	5	3	5	3
		S/S - hard	Caps	4		4		4		4
Tot										
Regular	4		2	2	2	2	2	2		
W/S - soft	Caps									
	Tot									
	Regular	7	6	1	6	1	6	1		
W/S - hard	Caps	3		3		3		3		
	Tot									
	Regular	10	6	4	6	4	6	4		
64 I/s - soft	Caps									
	Regular	3	4	-1	4	-1	4	-1		

		<i>Tot</i>							
	64 I/s - hard	<i>Regular</i>	3	2	1	2	1	2	1
		<i>Caps</i>	1		1		1		1
		<i>Tot</i>							
	64 V/S	<i>Regular</i>	4	3	1	3	1	3	1
		<i>Caps</i>							
		<i>Tot</i>							
	Drop Arm	<i>Regular</i>	5	4	1	4	1	4	1
		<i>Caps</i>	5		5		5		5
		<i>Tot</i>							
	Turning	<i>Regular</i>							
		<i>Caps</i>							
		<i>Tot</i>							
	A/C - B	<i>Regular</i>	8	11	-3	11	-3	11	-3
		<i>Caps</i>							
		<i>Tot</i>							
85	ASSY	<i>Regulars</i>	43	42	1	42	1	42	1
		<i>Caps</i>	17		17		17		17
		<i>Tot</i>							
		<i>All</i>	60	42	18	42	18	42	18
69	A	<i>Regulars</i>	31	32	-1	32	-1	32	-1
		<i>Caps</i>	20		20		20		20
		<i>Tot</i>							
		<i>All</i>	51	32	19	32	19	32	19
155	B	<i>Regulars</i>	75	60	15	60	15	60	15
		<i>Caps</i>	16		16		16		16
		<i>Tot</i>							
		<i>All</i>	91	60	31	60	31	60	31
61	T54	<i>Regulars</i>	26	23	3	23	3	23	3
		<i>Caps</i>	10		10		10		10
		<i>Tot</i>	1		1		1		1
		<i>All</i>	37	23	14	23	14	23	14
370	TOTAL	<i>Regulars</i>	175	157	18	157	18	157	18
	TOTAL	<i>Caps</i>	63		63		63		63
	TOTAL	<i>Tot</i>	1		1		1		1
		GRAND TOTAL	239	157	64		64		64

The following graph describes the clusters of different employees. And how can they grouped together.



Conclusion:

K mean is the simplest algorithm. Here in this paper is try to define the simplest problem .Manpower allotment with in the industry sector . so here the number of cluster can be specified as an input to the algorithm, so it is a easiest procedure with in the sector. When large scale of employee selection This approach can not be suited . because there is no right number of cluster. The cluster mean can be changed in the employee selection ,experience reference e etc. we cannot fit into particular cluster.

References:

- [1] Anderberg, M.R. 1973. Cluster Analysis for Applications. Academic Press.
- [2] Ball, G.H. and Hall, D.J. 1967. A clustering technique for summarizing multivariate data. Behavioral Science, 12:153–155.
- [3] Bezdek, J.C. 1981. Pattern Recognition with Fuzzy Objective Function. Plenum Press.
- [4] Bobrowski, L. and Bezdek, J.C. 1991. c-Means clustering with the l_1 and l_1 norms. IEEE Transactions on Systems,
- [5] Man and Cybernetics, 21(3):545–554.
- [6] Cormack, R.M. 1971. A review of classification. J. Roy. Statist. Soc. Serie A, 134:321–367.
- [7] Dubes, R. 1987. How many clusters are best? An experiment. Pattern Recognition, 20(6):645–663.
- [8] Dubes, R. and Jian, A.K. 1979. Validity studies in clustering methodologies. Pattern Recognition, 11:235–254.
- [9] Ester, M., Kriegel, H.P., Sander, J., and Xu, X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, USA: AAAI Press, pp. 226–231.
- [10] Everitt, B. 1974. Cluster Analysis. Heinemann Educational Books Ltd.
- [11] Fisher, D.H. 1987. Knowledge acquisition via incremental conceptual clustering. Machine Learning, 2(2):139– 172.
- [12] Gowda, K.C. and Diday, E. 1991. Symbolic clustering using a new dissimilarity measure. Pattern Recognition, 24(6):567–578.
- [13] Gower, J.C. 1971. A general coefficient of similarity and some of its properties. BioMetrics, 27:857–874.
- [14] Huang, Z. 1997a. Clustering large data sets with mixed numeric and categorical values. Proceedings of the First Pacific Asia Knowledge Discovery and Data Mining Conference, Singapore: World Scientific, pp. 21–34.
- [15] Huang, Z. 1997b. A fast clustering algorithm to cluster very large categorical data sets in data mining. Proceedings of the SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, Dept. of Computer Science, The University of British Columbia, Canada, pp. 1–8.
- [16] IBM. 1996. Data Management Solutions. IBM White Paper, IBM Corp.
- [17] Jain, A.K. and Dubes, R.C. 1988. Algorithms for Clustering Data. Prentice Hall.
- [18] Kaufman, L. and Rousseeuw, P.J. 1990. Finding Groups in Data—An Introduction to Cluster Analysis. Wiley.

- [19] Klosgen, W. and Zytkow, J.M. 1996. Knowledge discovery in databases terminology. *Advances in Knowledge*
- [20] *Discovery and Data Mining*, U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy (Eds.), AAAI Press/The MIT Press, pp. 573–592.
- [21] Kodratoff, Y. and Tecuci, G. 1988. Learning based on conceptual distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(6):897–909.
- [22] Lebowitz, M. 1987. Experiments with incremental concept formation. *Machine Learning*, 2(2):103–138.
- [23] MacQueen, J.B. 1967. Some methods for classification and analysis of multivariate