Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 57 Issue: 10:2024
DOI: 10.5281/zenodo.14038149

AN EXAMINING CLUSTER BEHAVIOUR ANALYTICALLY USING K-MEANS, EM, AND K* MEANS ALGORITHM

Dr. MARRIAPPAN E
Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamilnadu, India. Email: mapcse.e@gmail.com

Dr. ANNA LAKSHMI A
Department of Information Technology, RMK Engineering College, Chennai, Tamilnadu, India.
Email: annalaxmi.raj@gmail.com

AMALA PRINCETON X
Department of Computer Science and Engineering, VV College of Engineering, Tirunelveli, Tamilnadu, India. Email: princecardoza@gmail.com

VETRIVEL P
Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamilnadu, India. Email: vetrivel@ritrjpm.ac.in

Dr. RAMASAMY S
Department of Mechanical Engineering, St. Mother Theresa Engineering College, Tuticorin, Tamilnadu, India. Email: ramasamy@mtec.ac.in

ANGEL HEPHZIBAH R
Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamilnadu, India. Email: angelhepzibah@ritrjpm.ac.in

Dr. KALIAPPAN M
Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamilnadu, India. Email: kalsrajan@yahoo.co.in

RAMNATH M
Department of Artificial Intelligence and Data Science, Ramco Institute of Technology, Rajapalayam, Tamilnadu, India. Email: ramnath25@gmail.com

Abstract

An essential component of an intelligent data analysis process is clustering, an unsupervised learning technique. By grouping the patterns into homogeneous clusters, it facilitates the investigation of the links between the patterns. In the realm of information retrieval(IR), clustering has been dynamically applied to arrange of applications. One of the most active areas of study and development nowadays is clustering. By using clustering, one can find the set of significant groups in which members are more linked to each other than to members of other groups. There salting groupings can offer a framework for arranging length by text sections to make browsing and searching easier. Numerous clustering techniques have been thoroughly examined in relation to the clustering problem. Expectation Maximization(EM) and its variations, as well as the well-known link-means algorithm, are examples of iterative optimization clustering algorithms that have been shown to perform rather well for clustering. These algorithms are still among the most popular and effective. In the heart spect dataset, which has the following features: purity, entropy, CPU time, cluster-wise analysis, mean value analysis, and inter-cluster distance, this study examines the partition method clustering approaches, EM, Kmeans and K*Means algorithm. In order to support the

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 57 Issue: 10:2024
DOI: 10.5281/zenodo.14038149

conclusion that the behaviour in clusters produced by the EM algorithm is of a higher calibre than that of the k-means and k*means algorithms, the research finally presents the experimental results from datasets for five clusters.

Keywords: Cluster Analysis, Mean Value Analysis, EM, K- means, K*means, Purity, Entropy, Purity and Entropy.

I. INTRODUCTION

Data is divided into meaningful or useful groups (clusters) by CLUSTER analysis. If meaningful clusters are the goal, then the resulting clusters should capture the “natural” structure of the data. For instance, cluster analysis has been used to group related documents for browsing, find proteins and genes with similar functions, and provide a grouping of spatial locations that are prone to earthquakes. In other cases, however, cluster analysis is only a useful starting point for other purposes, such as saving data or quickly identifying the Closest neighbors between two places. Numerous domains, including psychology and their social sciences, biology, statistics pattern recognition, information retrieval, machine learning and data mining, have long used cluster analysis, whether for comprehension or practical purposes. Partitioning approaches, hierarchical methods, density-based methods, grid-based methods and model-based methods are the general categories in to which clustering algorithms can be divided. This paper's goal is to use the partitioning technique centroid algorithm, k-means, k*means and EM(Expectation-Maximization) algorithm to analyse the influence of clusters and their quality. This is how the paper is structured. Partitioning techniques are covered in Section 2. The K-means algorithm based on centroids is explained in Section3.InSection 4, the k*means method is explained. The EM algorithm, which is based on a probability model with parameters that characterize the likelihood that an instance belongs to a certain cluster is explained in Section5.The findings of the experiment are described in section 6. With fewer discussions the work is concluded in section7.

II. RELATED WORKS

Finding the intrinsic grouping in a set of unlabeled data is the primary goal of clustering. As a result, the user is responsible for providing this criterion in a way that ensures the clustering outcome satisfies both their needs and the requirement soft the relevant user[11and12]. Clustering algorithms fall under the following categories: When using Exclusive Clustering, data are clustered in a way that prevents them from being included in other clusters if a particular datum exists and belongs to one.

k-means algorithm

Fuzzy sets are used in Overlapping Clustering to cluster data, allowing each point to potentially belong to many clusters with varying degrees of membership [14]. In this instance, information will be linked to a suitable membership value.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 57 Issue: 10:2024
DOI: 10.5281/zenodo.14038149

Fuzzy C-means algorithm

The union of the two closest clusters in the provided data set serves as the foundation for a Hierarchical Clustering algorithm [15]. By designating each datum as a cluster, the initial condition is fulfilled. It takes a few attempts to reach the desired final clusters.

Agglomerative algorithm

The last kind of clustering is called probabilistic clustering, and it employs an entirely probabilistic methodology.

Mixture of Gaussians algorithm

The centroid and medoids algorithms are the two main sub categories of partitioning techniques. The gravity centre of each instance is used by the centroid methods to represent each cluster [4]. The instances that are closest to the gravity centre are used by the medoid algorithms to represent each cluster. The k-means method is the most well-known centroid algorithm. The data set is divided into k subsets using the k-means algorithm so that each subset's points are closest to the same centre. To be more specific, k instances are chosen at random to represent the clusters [13]. There maiming instances are allotted to the nearest centre based on the chosen criteria. Then, using the mean of all the data points in a cluster, K-means calculates the new centres. Until the gravity centres remain unchanged, the process is repeated. Storage allocation and energy efficient techniques [21] and [22] are useful in this conceptual version of transverse orientation. An underlying probability model with parameters describing the likelihood that an instance belongs to a certain cluster is assumed by the expectation-maximization(EM) algorithm.

PV_E mobile is used to decorate the maximum load e-capability parameter and voltagee-stability [20] and [21]. The rating of an app is considered by setting a threshold value for evaluation [25]. This algorithm's initial predictions for the mixture model parameters are its method. The cluster probabilities for each instance are then computed using these values. The procedure is then repeated using these probabilities tore-estimate the parameters. How popular an app is and what permissions it requires from the user significantly increase the app's potential security risks [24].

III. METHODOLOGY

Each point is assigned by the K-means algorithm to the cluster whose centroid, or centre, is closest. The coordinates of the centre are the arithmetic means for each dimension divided by the total number of points in the cluster; that is, the centre is the average of all the points in the cluster [2].

One way to conceptualize K-Means as an algorithm is as an algorithm that uses hard data assignment to a pre-determined set of divisions. Each data value is assigned to the closest partition at each algorithm pass based on a similarity metric, like the Euclidean distance of intensity [5]. These difficult assignments are then used to recalculate the divisions.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 57 Issue: 10:2024
DOI: 10.5281/zenodo.14038149

Each time through, a data value can Either create k clusters at random and identify their centres, or create k random locations at random and use them as cluster centres.

- Each point should be assigned to the closest cluster centre.
- Recalculate the new centres of clustering.
- Iterate through the preceding two stages until a convergence requirement is satisfied, which is often the assignment hasn't changed.

The precise number of clusters that must be as separate as feasible must be entered into this kind of algorithm. Thus, the k-means clustering algorithm can handle these kinds of research questions. Generally speaking, the k-means algorithm will generate precisely k distinct clusters with the highest level of distinction [18]. The optimal number of clusters, k, that results in the largest distance between the clusters needs to be calculated from the data and is not known beforehand.

The outcome of a k-means clustering study is evaluated by looking at the means for every cluster on every dimension to see how unique our k clusters are. For most, if not all, of the dimensions utilized in this analysis, very different means are obtained [19]. The degree of discrimination across clusters that each dimension provides is indicated by the magnitude of the results obtained from the analysis of variance carried out on that particular dimension. The algorithm's basic steps are rather straight forward: first a pre-determined number of k clusters are considered and then observations are assigned to those clusters in a way that maximizes the difference between cluster means.

K* Means Algorithm

The Step-wise Automatic Rival-penalized (STAR) k-means algorithm a generalized variant of the conventional k-means clustering technique, is presented in this section of the study [2]. The k*means clustering algorithm starts by allowing each cluster to obtain a minimum of one seed point. You could think of this initial stage as a pre-processing step. Next, using a learning method that automatically penalizes the winning chance of all rival seed points in subsequent competitions and tunes the winning unit to adapt to an input, the units are adaptively fine-tuned. The comprehensive k*means can be distributed as follows:

- The initial step is generally implemented using Frequency Sensitive Competitive Learning such that they can achieve the goal as long as the number of seed points is not less than the exact number k* of clusters. Therefore, the number of clusters is $k \geq k^*$, and randomly initialize the k seed points.

Select a data point x^t at random from the input data collection and for j=1, 2, ...k let,

$$U_j = \begin{cases} 1 & \text{if } j = w = \arg\min_r \lambda_r \|x_t - m_r\|_k, \\ 0 & \text{otherwise.} \end{cases} \quad \text{-- (1)}$$

- Uj: This is likely an indicator variable, which takes the value of 1 if a certain condition is met and 0 otherwise.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 57 Issue: 10:2024
DOI: 10.5281/zenodo.14038149

- $j=w=\text{argmin}_r \lambda_r \|xt-mr\|_k$
- j and wrefer to indices that represent clusters or groups.
- argmin_r finds the value of r that minimizes the expression that follows.
- λ_r: This might represent a weight or scaling factor for each cluster or group r.
- x_t: This could represent a data point (like a vector) in your dataset.
- m_r: This likely represents the mean or centroid of cluster r.
- $\|xt-mr\|_k$: This denotes the distance (in the k -norm, like Euclidean norm if k=2) between the data point $x_t$and the cluster centrem_r.

$$\lambda_v = n_j \sum_{r=1}^{n_r} \quad \text{--(2)}$$

- λ_r: This variable is defined as a product of njand the summation.
- n_j: Likely a constant or scaling factor.
- $\Sigma_{r=1}^{n}$ A summation over r from 1 up to nr.
- n_r: This might represent cumulative values or counts.

Number of occurrences of $u_r=1$.

The next step is to update the winning seed point mw by the following equation

$$m_{new} = m_{old} + \frac{(x - m_{old})}{w} \quad \text{--(3)}$$

- m_{new}: The updated value of m.
- m_{old}: The previous value of m.
- x: The new data point being incorporated.
- w: A weight or scaling factor, possibly representing the count or a specific weight parameter.

where the positive learning rate isthe lowest. Instead of estimating the in exact values, they only seek to distribute the seed points into specific regions, so the input co variance is left out.

The a fore mentioned pair of actions are iterated until the k-seriesof u_j, where j = 1, 2,... k, stays un altered for every x_t.

- If you have a data point xt, use the following equation to calculate I(j|xt)s: I(j|x) ={1, if j =w=argminrρr,0 otherwise.
- Revise the winning seed point mw using the subsequent formula:

$$m_{new}=m_{old}+w\cdot(x-m_{old}) \quad \text{--(4)}$$

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 57 Issue: 10:2024
DOI: 10.5281/zenodo.14038149

Where:

- m_{new} is the updated or new average.
- m_{old} is the previous or old average.
- x is the new data point being introduced.
- w is the weight (often between 0 and 1).

Further the parameters like αj s and Σwshould be updated. The above mentioned two steps are repeated until the K series of I(j|xt) with j=1,2,...k, remain unchanged for all x_{ts}. Some of the drawbacks of the traditional k means clustering process are addressed by the k*means clustering algorithm. The dead-unit issue that existed with the traditional k means clustering approach is resolved by the k*means clustering algorithm.

EMALGORITHM

The general purpose of EM techniques is to detect clusters invariables (or observations) and to group those data in to clusters. An example of a typical application for this kind of analysis is a type of marketing research study where a wide range of variables connected to consumer behaviour are measured representative sample of responders. The goal of theresearchistoidentify"marketsegments,"orgroupingsofrespondents.

A is to identify "market segments," or groupings of respondents that, in comparison to respondents who "belong to" other clusters, are somehow more similar to one another (or to all other members of the same cluster). Finding the differences between the clusters that is finding the particular variables or dimensions that vary and how they differ with respect to individuals in various clusters are typically just as important as identifying the clusters themselves.

This clustering method's basic strategy and logic involve computing a single continuous big variable across a sizable sample of observations [16 and 17]. Furthermore, taking in to account that each sample in the provided dataset comprises of two clusters of observations with distinct means, the distribution of values for the continuous big variable is expected to follow a normal distribution. This fundamental method of clustering is expanded upon by the EM (expectation maximization) algorithm in two significant ways: The EM clustering technique computes probabilities of clusters that are (final). Though categorical variables can also be accommodated by modifying the conventional k-means clustering approach, the general EM algorithm can be applied to both continuous and categorical variables [1].

The K-Means algorithm and the EM algorithm have a very similar configuration. selecting the input partitions Comes first. The Expectation step, which starts the EM cycle, is specified by the equation that follows:

$$E[z_{ij}] = \frac{p(x = x_i|\mu = \mu_j)}{\sum_{n=1}^{k} p(x = x_i|\mu = \mu_n)} \quad \text{--(5)}$$

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 57 Issue: 10:2024
DOI: 10.5281/zenodo.14038149

According to this equation, assuming that μ is partition μ_i divided by the sum over all partitions k of the same previously specified probability, the expectations or weight for pixel z with respect to partition j equals the likelihood that x is pixel x_i. The weights 'reduced expression results from this. The covariance of the pixel data is represented by the sigma squared found in the second expression. The M step, also known as the maximisation step, starts after the E step is complete and each pixel has a weight or expectation for each partition. We define this step using the following equation:

$$\mu_j \leftarrow \frac{1}{m} \sum_{i=1}^{m} E[z_{ij}] x_i \quad \text{--(6)}$$

cluster memberships based on one or more probability distributions, as opposed to grouping cases or observations to maximise the differences in means for continuous variables [3]. The clustering algorithm then seeks to maximise the likelihood or total probability of the data given the According to this equation, the weighted average of the pixel values whose weights are the weights from the E step for this specific partition is used to replace the partition value j. Like with the K-Means technique, this EM cycle is repeated for every new set of partitions until the partition values no longer fluctuate significantly.

IV. EXPERIMENT RESULTS

This section of the paper examines the behaviour of the clusters using the SPECTF dataset to cluster five cluster groups. This is accomplished by comparing the results of the EM algorithm, k means and k*means in order to highlight the experimental findings. After using the purity measure, we determine the fraction of members of the CPU time measurements for the EM, k*means, and k means algorithms are all confirmed in this study. We use the EM, k*means, and k means algorithms to calculate the dataset's mean values for clustering. Based on each of these primary criteria, an overall analysis has been conducted to examine the cluster quality for the values of both the EM, k*means and K-means algorithm. The algorithms have been executed using MATLABversion7.0. The clusters formed by EM and K-means algorithm has been evaluated.

Heart Spect Dataset

SPECTF is a good data set for testing ML algorithms, it has database of 267 SPECT image sets (patients) was processed to extract features that summarize the original SPECT images. All the inputs in the data set where normalized. It can be scaled in the range of0 to1.Purity measures the percentage of the domain an tclass members267 instances that are described by 45 attributes. The dataset describes diagnosing of cardiac Single Proton Emission Computed Tomography(SPECT)-1images. Each of the patients,

Table 1: Is classified in to two categories: normal and abnormal

Purity Measure	Normal	Number of 1's
Kmeans	K*Means	EM
3403	3390	3373

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 57 Issue: 10:2024
DOI: 10.5281/zenodo.14038149

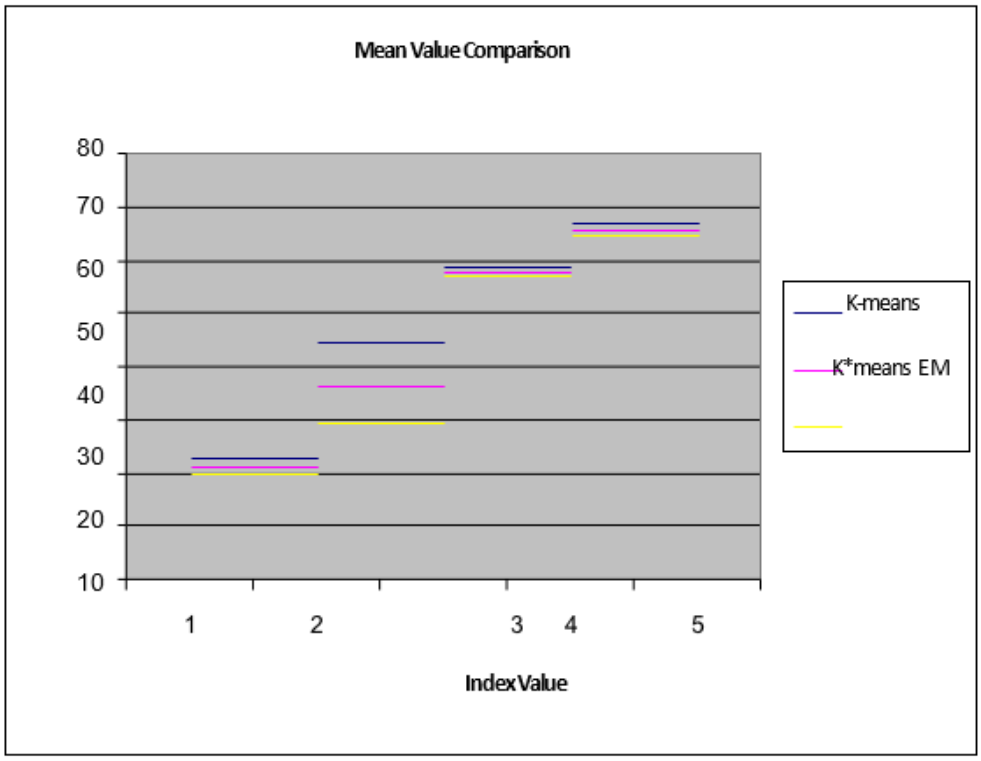
Table 2: CPU Time

Kmeans	K*Means	EM	Kmeans + K*Means
80e+003	1.7385e+003	.7388e+003	3.2424e+003

Table 3: Mean wise Comparison

K Means	K* Means	EM
22.7297	21.2556	19.5883
44.7815	36.4515	29.6015
58.6818	57.9650	57.1785
66.8687	65.6565	64.7206
74.8776	72.9605	71.1364

K*means and EM algorithm for five types of clusters. The graphs how the experimental results how the clusters' actions ink*means, k means and the EM technique. Comparing the EM approach to the k means and k*means clustering methods, it is evident that progress has dominant class, from which entropy is derived. The most widely used criterion for comparing clustered datasets with in a given cluster (larger is better), while entropy looks at the distribution of documents from each reference class within clusters (smaller is better).



The percentage of members of the dominant class who scored a 1after the purity metrics were applied is shown in (Table-1). For every kind of cluster, the entropy is computed and compared. One of the most crucial factors to be considered is CPU time consumption, which is covered in this section. Table 2 lists the CPU times for the EM, K*means, K-

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 57 Issue: 10:2024
DOI: 10.5281/zenodo.14038149

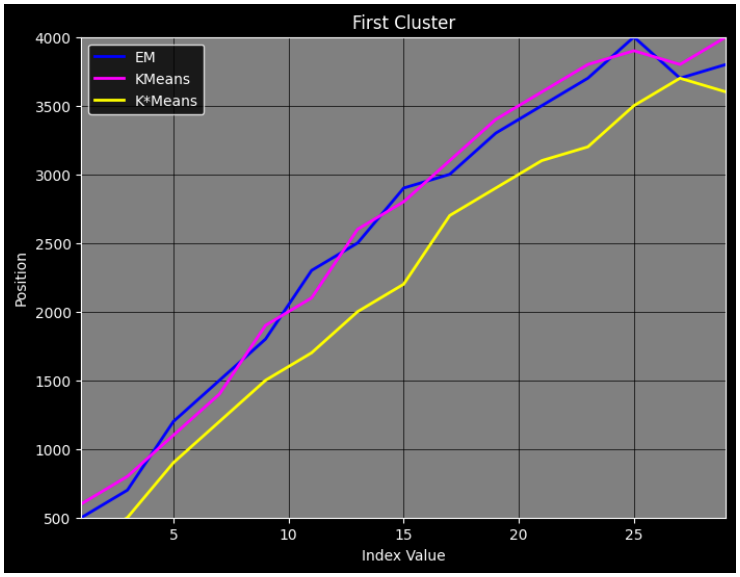
means, and combined EM and K-means algorithms. been made. The dataset has been allocated a cluster limit of five, and each cluster within these five clusters is plotted along with its index and position values. Within these five groups, Figure 2 displays the dataset clustering results.

Figure 2's graphs (a, b, c, d, and e) show how the SPECTF dataset behaves when sorted by first, second, third, fourth, and fifth, in that order. For each of the five clustering groups in each of the two techniques, the graphs are displayed with the index against the position values of each dataset. This demonstrates un equivocally that, each time the dataset is clustered, the EM method produces results that exhibit a notable improvement above those of the Kmeans and K*means algorithms.

VI. CONCLUSION

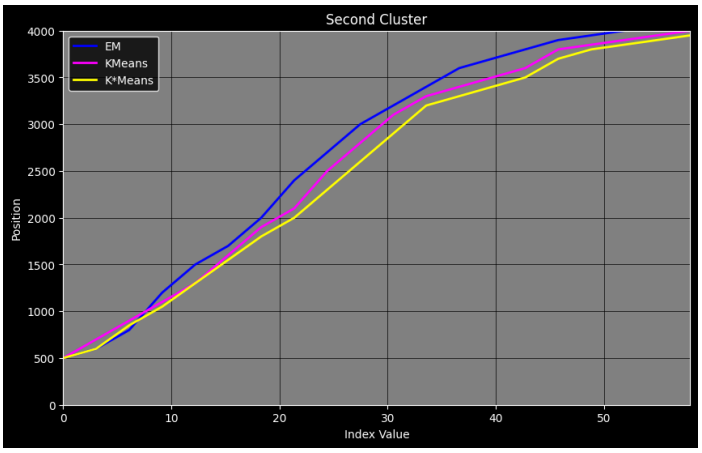
This paper provides a quick overview of the partition methods, including a detailed description of the k-means and k*means methodologies. The cluster quality with purity measure for the EM, K*means and Kmeans algorithms is the demonstrated. Five clusters are used to analyze EM, K*means, and K-means cluster-wise. K-means, K*means, and EM are used to analyze mean value and CPU time performance. Ultimately, the findings of the experiment are determined, demonstrating the enormous improvement in the quality of the behaviour of clusters in the EM algorithm when compared to the k*means and k means algorithm.

There needs a great deal of future work to be done in the field of clustering technique which is active of research which have been applied to a wide variety of research problems.

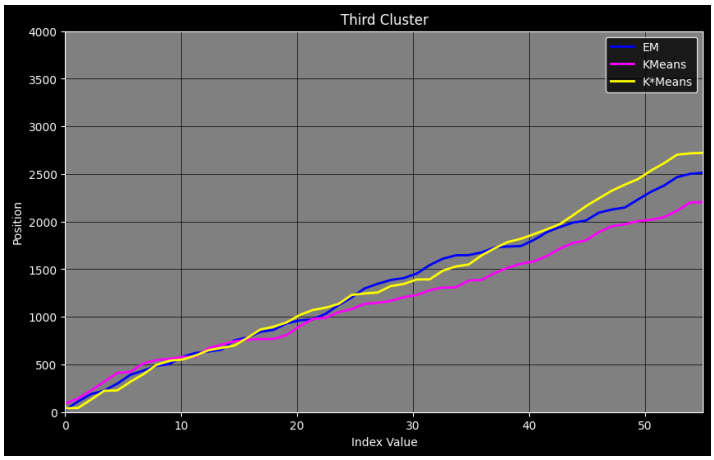


(a)

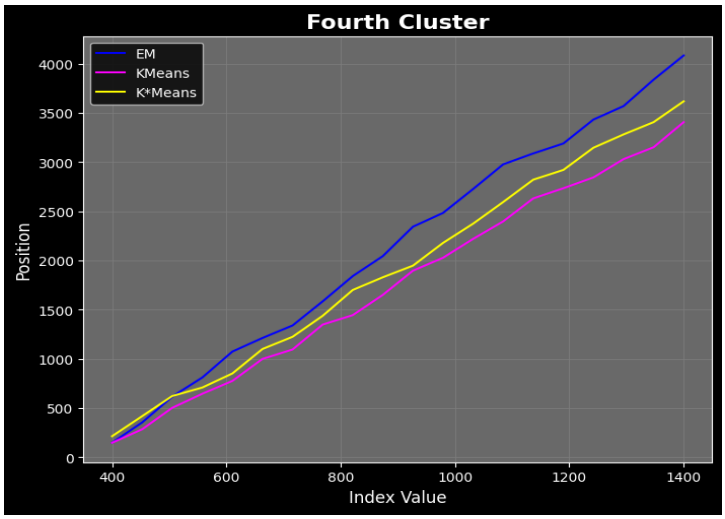
Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 57 Issue: 10:2024
DOI: 10.5281/zenodo.14038149



(b)

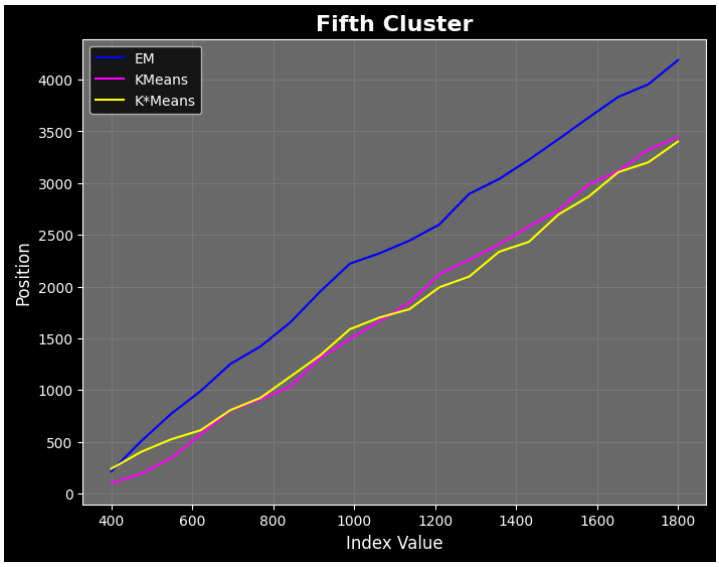


(c)



(d)

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 57 Issue: 10:2024
DOI: 10.5281/zenodo.14038149



(e)

Fig 2: Results on SPECTF dataset: Clustering of (a) SPECTF dataset into first cluster (b) SPECTF dataset into second cluster (c) SPECTF dataset into Third cluster(d)SPECTF dataset into Fourth cluster(e)SPECTF dataset into fifth cluster

In the field of medicine for the clustering diseases, to cures for diseases, or to classify the symptoms of diseases that can lead to very useful taxonomies in the field of psychiatry, the correct diagnosis of clusters of therapy to be effective, symptoms like schizophrenia, paranoia, etc. must be addressed. By using cluster analytic approaches, archaeologists have tried to create taxonomies of stone tools, burial goods etc. Therefore, there is a need for significant advancements in this field of study. Cluster analysis is generallya highly useful tool when one needs to organize a very huge amount of data in to digestible comprehensible piles.

References

1) G.J. McLachlan and T. Krishnan, “The EM Algorithm and Extensions”, Wiley, 1997.

2) Yiu-Ming Cheung, k*-Means: A new generalized k-means clustering algorithm, Pattern RecognitionLetters24, 2003.

3) Michiko Watanabe and Kazunori Yamaguchi, “The EM Algorithm and Related Statistical Models”in2004

4) S.B. Kotsiantis, P. E. Pintelas, “Recent Advances in Clustering: A BriefSurvey”, 2005.

5) Tapas Kanungo, David M. Mount, Nathan S.Netanyahu, ChristineD. Piatko, Ruth Silverman, and AngelaY.Wu, “Efficient Algorithms for K-Means Clustering”.

6) WilliamJ. Palm, University of RhodeIsl and, “A Concise Introduction to MATLAB”, McGraw-Hill, 2008.

7) Oren Kurland Lillian Lee, Clusters, language models, and adhoc information retrieval, Volume27, Issue3-ACM Transactions on Information Systems (TOIS), May2009.

Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/
Journal of Tianjin University Science and Technology
ISSN (Online):0493-2137
E-Publication: Online Open Access
Vol: 57 Issue: 10:2024
DOI: 10.5281/zenodo.14038149

8) Hans-PeterKriegel, Peer Kröger, ArthurZimek, “Clustering high-dimensional data: A survey on subspace eclustering, pattern-based clustering, and correlation clustering”, Article No. 1 Volume 3, Issue 1ACM Transactions on Knowledge Discovery from Data (TKDD), March2009.

9) Changqing Zhou, Dan Frankowski, Pamela Ludford, Shashi Shekhar, Loren Terveen, “Discovering personally meaningful places: An interactive clustering approach”, Volume25, Issue 3ACM Transactions on Information Systems (TOIS), July2007.

10) Achtert, E., Böhm, C., David, J.,Kröger, P., and Zimek, A., “Robust clustering in arbitrarily oriented subspaces”, In Proceedings of the 8th SIAM International Conference on Data Mining(SDM),2008.

11) Strehl, A., Ghosh, J., and Mooney, R.J. Impact of similarity measures on webpage clustering. In Proceedings of AAAI Workshop on AI for Web Search, pages58–64, 2000.

12) Strehl, A., and Ghosh, J. Cluster ensembles acknowledge re use frame work for combining multiple partitions. Journal on Machine Learning Research, 3:583–617, 2002.

13) McCallum, A. K. Bow: A toolkit for statistical language modelling, text retrieval, classification and clustering. http://www.cs.cmu.edu/~mccallum/bow.

14) Rasmussen, E. Clustering algorithms. In W. Frakes and R. Baeza Yates, editors, Information retrieval: data structures and algorithms. Prentice Hall,1992

15) Salton, G., and Buckley, C. Term weighting approaches in automatic text retrieval. Information Processing and Management: An International Journal, 24(5):513–523, 1988.

16) He, J., Tan, A.H., Tan, C.L., and Sung, S.Y. On Quantitative Evaluation of Clustering Systems. In W. WuandH. Xiong, editors, Information Retrieval and Clustering. Kluwer Academic Publishers, 2003.

17) Boley, D., and Borst, V. unsupervised clustering: A fast scalable method for large datasets. CSE Report TR99029, University of Minnesota, 1999.

18) Bradley, P.S., and Fayyad, U.M, “Refining initial points for Kmeans clustering”, In Proceedings of the Fifteenth International Conference onMachineLearning,pages91–99,1998.

19) Boley, D. Principal direction divisive partitioning, Datamining and Knowledge Discovery, 2(4):325–344, 1998.

20) M Kaliappan, E Mariappan, MV Prakash, B Paramasivan, Load Balanced Clustering Technique in MANET using Genetic Algorithms. Defence Science Journal 66 (3), 251-258

21) GS Kumar, M Kaliappan, LJ Julus, Enhancing the Performance of MANET using EESCP, Pattern Recognition, Informatics and Medical Engineering (PRIME), 2012

22) S Vimal, YH Robinson, M Kaliappan, K Vijayalakshmi, S Seo, A method of progression detection for glaucoma using K-means and the GLCM algorithm toward smart medical prediction. The Journal of Supercomputing, PP.1-17

23) M Sivaram, M Kaliappan, SJ Shobana, MV Prakash, V Porkodi Secure storage a location scheme using fuzzy based heuristic algorithm for cloud, Journal of Ambient Intelligence and Humanized Computing, pp.1-9.

24) Ramnath, M., Rubavathi, C.Y. Proceedings of the 2023 2nd International Conference on Augmented Intelligence and Sustainable Systems, ICAISS 2023, 2023, pp. 821–825.

25) Ramnath, M., Rubavathi, C.Y. Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2021, 2021, pp. 1180–1183, 9388544