Viksit India
2024 | Volume: 02 | Issue: 01

Study onward clustering strategy along with Assorted K-mean Technique

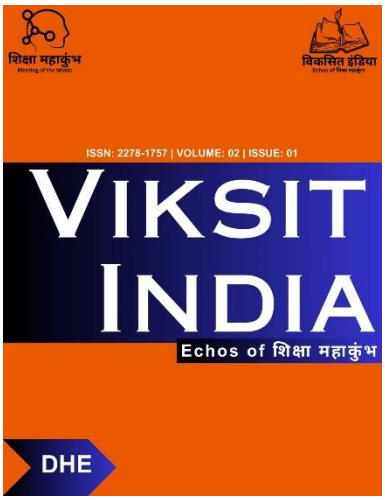
Pardeep Singh Tiwana, Heena Wadhwa, Priya Dogra, Lakhvinder Kaur
Chandigarh Engineering Colleges CGC, Landran, Mohali, Punjab
Chitkara University Institute of Engineering & Technology, Chitkara University, Rajpura, Punjab, India
Chandigarh Engineering Colleges CGC, Landran, Mohali, Punjab
Chandigarh Engineering Colleges CGC, Landran, Mohali,Punjab

Article Info
Received: June 8, 2023
Revised: June 25, 2023
Published: March 31, 2024
Editor: Adv. Vikram Verma

*Corresponding author
Email: tiwana07763@gmail.com
9417307763

Open Access

DOI:



https://vi.rase.co.in/
ISSN: 2278-1757

Abstract
Clustering is a grindstone of severance as scimitar type of data in a group and dissimilar type of data into different groups. It is frequently used technique with k-mean algorithm on assorted type of data with various techniques. In this paper, describe the various techniques of k-mean (i.e., adaptive k-mean, distributed k-mean, moving k-mean, parallel k-mean, Reverse k- mean, sequential k-mean). Starting with K elements chosen from the input data set, the adaptive K-means clustering method is run. The K elements, which are chosen at random, serve as the seeds for clusters. These techniques can help to optimize the resources and data efficiently. This methodology can help to generate sustainable way for the usability of resources to save energy.

Keywords: Clustering, K-mean, Adaptive K-mean, Distributed K-mean, Moving K-mean, Parallel K-mean, Reverse K-Mean, Sequential K-Mean.

Introduction
The most crucial technique to assist consumers efficiently browse, summarized, and organize data is clustering. [6].In the clustering process objects are organized into groups of similar members. Clustering include two type of similarity that is intra-cluster similarity and inter cluster similarity in “Fig1”.

Intra-cluster similarity defines the similarity between the data elements in one group. There are number of researchers who are working on different machine learning algorithms [26][27]. In which data elements have same properties with each other in same group. In a perfect cluster Intra-cluster similarity is high. Inter-cluster similarity defines the similarity between the data elements of different groups. In which data elements of one group have different properties with the data elements of another group. In a perfect cluster Inter-cluster similarity is low.

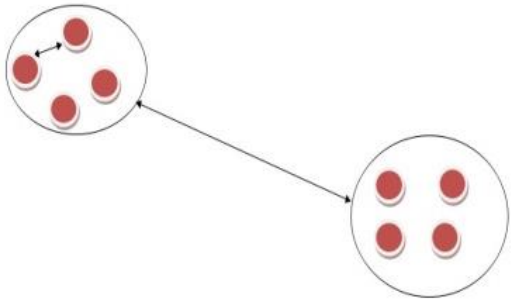


Fig.1. Clustering Similarity

Clustering problems are solved by the k-mean algorithm. This k-mean is a simple procedure to classify data into certain number of clusters on basis of distance to centroid. The various methods that are proposed by using k-mean are described in this is paper are: adaptive k-mean, distributed k-mean, moving k-mean, parallel k-mean, Reverse k-mean, sequential k-mean on different type of data in different areas.

It is very crucial to make cluster perfect to retain effective result.

K-mean
K-mean as one of the most well-known and straightforward unsupervised learning algorithms that divides a set of any form of data into k clusters in order to solve the clustering issue. To determine the k centers, one for each cluster, is the major goal. It organizes data pieces into a predetermined number of clusters. The clever placement of these centroids due to various clusters produces various outcomes. Therefore, it is preferable to situate them apart from one another. When a data collection has n items, indicated by e^1, $e^{2,}$ and en, clustering the process of optimising their placement into k clusters in order to satisfy a global criteria function.

$$\sum_{J=1}^{k}\sum_{i=1}^{n} \text{sim}(e_i, c) \quad (1)$$

Depending on how the term "sim" is defined, it either reduced or maximized (ei,cj). For j = 1, k, Cj is the centroids of the cluster, and sim(ei,cj) measures how similar two elements are to a centroids. The following list of k-stages means is described:

1. Choose k initial centroids "c" that each stand for a cluster.
2. Measure the separation between each data point and the centroids, then place each element on the nearest centroid.ci

$$v_i = (1 / ci) \sum_{J=1} x_i \quad (2)$$

3. Recalculate the distance between each element and new obtained centroids.
4. If no element was reassigned then stop, otherwise repeat 3 and 4.

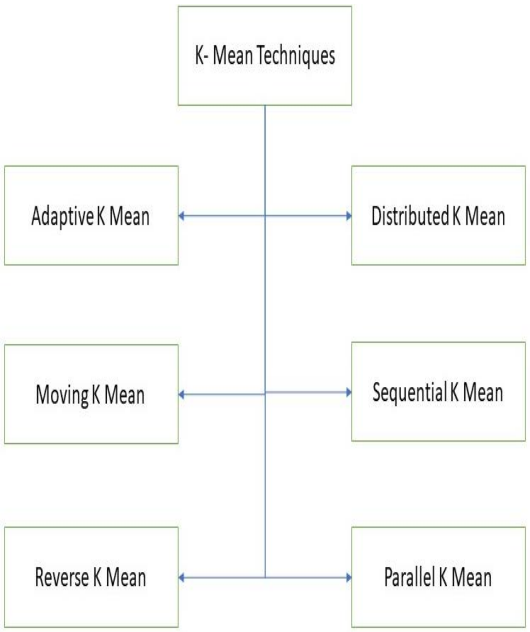


Fig 2. K-Mean Techniques

1. Adaptive K-Mean

Regions with evenly variable intensity distributions can be divided up using adaptive k-mean clustering. For picture segmentation in digital image processing, this approach is employed [1]. Traditional statistical picture segmentation techniques, as straightforward as thresholding or as complex as grouping pixels only by their intensities Adaptive k-mean is used to solve this issue. An intensity-based strategy is used in the adaptive k-mean method.

2. Moving K-Mean

Automatic segmentation of colour images is accomplished using moving k-mean clustering [3]. Moving k-mean is used in image digital processing to separate the undesired background—which is still red even after decolonization—from the TB (tuberculosis bacilli) by employing the green component of the RGB (radial basis function) colour model. Fuzzy c-mean clustering and k-mean clustering are often used for automatic photo segmentation. However, neither technique consistently works because of dead centers, duplicate centers, and being confined to local minimas. Produced satisfactory results. To mitigate these problems, moving k-mean clustering, a modified version of k-mean clustering was created. The technique was first employed to place the centre of an RBF network, but was subsequently discovered to be appropriate for picture segmentation.

3. Reverse K-Mean

Reverse k-mean clustering is a method that searches for a place in geographical data by operating in the opposite direction of regular k-mean clustering to identify centroids [5]. In order to carry out a location search that is not reliant on network connectivity, the reverse k-mean algorithm was developed as a solution. This method was created utilizing the idea of k-mean clustering, and it is called reverse k-mean since it begins by entering the computed centroids values and then searches the areas nearby the centroids values before conducting a zoom-level search.

4. Distributed k-mean

The distributed k-mean clustering approach immediately outpaced the stand-alone alternative by adding more computer nodes to the cluster computing environment. Each computer node was subjected to a distributed k-mean clustering algorithm with an unrestricted number of classes for better results [2]. The categorization of tree species is crucial for the preservation of forests because trees are essential to preserving the circumstances required for life to exist on Earth. The distributed k-mean built for cluster computing environment supports a cluster of servers that consists of a single master node and one or more computer nodes. The k-mean and the merging method are the two essential parts of the distributed k-mean. K-mean clustering is employed on the compute node, and Sequential k-mean. On the compute node, the k-mean clustering technique is used, and the head node is where the merging process takes place. To identify the type of tree in a picture with great spatial resolution, distributed k-means is used.

Data mining uses the sequential k-mean technique to group text data into clusters. The issue was with the cosine function, which solely takes into consideration pairwise similarity when determining whether a text belongs in a cluster or not. Due to the possibility that certain documents in distinct clusters may be near to one another, the cluster wasn't really clearly divided based on pairwise similarity [7]. It is suggested to use sequential k-means based on neighbors to solve this

issue. The neighbor matrix is used in this approach. the neighbor matrix created prior to the commencement of the clustering procedure. When assigning the k-mean procedure, sequential k-mean employs the neighbor matrix to choose the first cluster centroid and calculates the similarity between two texts using the cosine-link function.

Two documents must be created, their leads to the conclusion must be computed, and they must be evaluated against a user-defined threshold in order to construct a neighbour matrix for a given piece of data. The steps in the sequential k-mean method based on neighbour are as follows:

- Generate n ×n neighbor matrix for n number of documents.
- Pick k first cluster centroids from the rankings.
- Determine the similarity value between each cluster and each document using the cosine-link function, then set the centroid of each document at that point..
- Based on the documents that were allocated to them, recalculate the k centroids.
- If no element was reassigned then stop, otherwise repeat 3 and 4.

5. Parallel k-mean

For the clustering of text documents based on the neighbour matrix, parallel k-means is advised. We take into account any two texts that are sufficiently similar to one another. It is a sequential neighbor-based parallel variation of k-means. The sequential k-mean based on neighbours has the drawback of taking a lengthy time to construct the neighbour matrix; the parallel k-mean fixes this issue. Sequential k-means' data parallelism is exploited by parallel k-means' special parallel pair-generating technique to build the neighbour matrix [7].

Parallel k-means based on neighbours are used for message-passing multiprocessor systems, which is very efficient and has great scalability with respect to the number of processors and the amount of data. The parallel k-means approach is used to provide data parallelism instead of the traditional sequential k-mean operations of building the neighbour matrix, selecting initial centroids, and iterating the loop of document assignment. This method uses the pair-generating PG-New technique to produce the neighbour matrix. This uses a configurable number of processors and fewer rounds than either PG-A or PG-B. Following are the procedures for Parallel k-mean based on neighbours:

Create an n-by-n neighbour matrix for an infinite number of documents using p processors.

1. Using p processors, choose k initial cluster centroids in parallel on the rankings.
2. Using the cosine-link function, each processor determines how similar each of its documents is to each centroid of k clusters, and then assigns each document to the closest centroid.
3. Recalculate the cluster centroids using the data from the processors, then broadcast the results to the processors.
4. If no element was reassigned then stop, otherwise repeat 3 and 4.

Approaches of Different Techniques

TABLE I. Different K-mean Techniques

Sr. No	Different K-Mean Techniques		
	Algorithm name	***Area***	***Approaches***
1.	Adaptive k-mean	Digital image processing	Intensity based approach
2.	Moving k-mean	Digital image processing	Redundancy based approach
3.	Reverse k-mean	Computer networks	Independent to network connectivity
4.	Distributed k-mean	Digital image processing	High spatial resolution image
5.	Sequential k-mean	Data mining	Well-defined Clustering based on neighbors
6.	Parralel k-mean	Data mining	Speedup in sequential k-mean

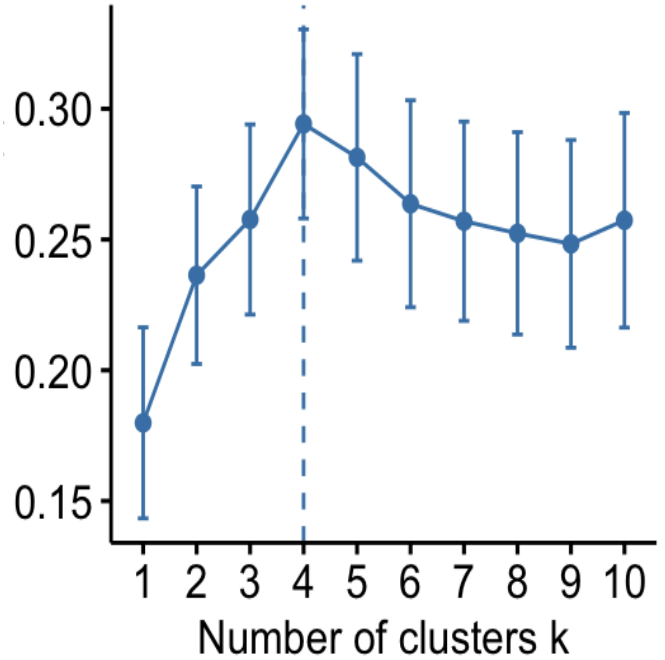


Fig 3: Analysis of Various Clusters

Conclusions

This paper describes the various k-mean techniques. These Various techniques give the good results in different areas. Clustering plays crucial role in every field with various k-mean techniques for good results. One of the many data operations and techniques utilised in the field of data science is the unsupervised machine learning methodology K-means clustering. It is the quickest and most efficient method for

grouping data points even while there is very little information about the data.

References

i. *C. W. Chen, J. Luo and K. J. Parker, "Image segmentation via adaptive K-mean clustering and knowledge-based morphological operations with biomedical applications," in IEEE Transactions on Image Processing, vol. 7, no. 12, pp. 1673-1683, Dec. 1998, doi: 10.1109/83.730379.K. T. Fan, Y. C. Tzeng, Y. J. Su, K. S. Chen, "Tree Identification Using a Distributed K-Mean Clustering Algorithm," IEEE 2010.*
ii. *M. K. Osman, M. Y. Mashor, Z. Saad and H. Jaafar, "Colour Image Segmentation of Tuberculosis Bacilli in Ziehl-Neelsen-Stained Tissue Images Using Moving K-Mean Clustering Procedure," 2010 Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation, 2010, pp. 215-220, doi: 10.1109/AMS.2010.51..*
iii. *Jayabharathy, Jayaraj, and Selvadurai Kanmani. "Correlated concept based dynamic document clustering algorithms for newsgroups and scientific literature." Decision Analytics 1.1 (2014): 1-21.*
iv. *A. Singh and D. Somwanshi, "Offline location search using Reverse K-Mean Clustering & GSM communication," 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), 2015, pp. 1359-1364, doi: 10.1109/ICGCIoT.2015.7380678.*
v. *M. P. Naik, H. B. Prajapati and V. K. Dabhi, "A survey on semantic document clustering," 2015 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), 2015, pp. 1-10, doi: 10.1109/ICECCT.2015.7226036.*
vi. *Li, Yanjun, Congnan Luo, and Soon M. Chung. "A parallel text document clustering algorithm based on neighbors." Cluster Computing 18.2 (2015): 933-948..*
vii. *P. S. Tiwana and N. Mann, "Jellyfish reorder attack on hybrid protocol in manet dissection on variegated parameters," 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2016, pp. 96-100, doi: 10.1109/PDGC.2016.7913122.*
viii. *Tiwana, Pardeep Singh, et al. "An scrutiny of run-time ramification for 5-proviso busy beaver proving empirical composition." International Journal of Innovative Technology and Exploring Engineering 8.9 (2019): 726-732.*
ix. *Sachin, Majithia, and Dinesh Kumar. "Implementation and Analysis of AES, DES and Triple DES on GSM Network." IJCSNS International Journal of Computer Science and Network Security 10 (2010): 298-303.*
x. *Dogra, Priya, and Rakesh Kumar. "Fuzzy K-Medoid Clustering Strategy for Heterogeneous and Dynamic Data for IoT Scenario." Proceedings of International Conference on IoT Inclusive Life (ICIIL 2019), NITTTR Chandigarh, India. Springer, Singapore, 2020.*
xi. *Sharma, Gurmandeep Kaur1 Atul, and Pardeep Singh Tiwana. "Internet of Things: Purview, Challenges and Probability."*
xii. *Nidhi, Nidhi, Sachin Majithia, and Neeraj Sharma. "Predictive Model for Students' Academic Performance Using Classification and Feature Selection Techniques." 2021 2nd International Conference on Computational Methods in Science & Technology (ICCMST). IEEE, 2021.*
xiii. *Gupta, Astha, et al. "AFLP based genetic relationship and population structure analysis of Canna—An ornamental plant." Scientia Horticulturae 154 (2013): 1-7.*
xiv. *Singh, Jupinder, and Sachin Majithia. "Impact of a Node mobility in two Mobile WiMAX Networks under different speeds." International Journal on Recent and Innovation Trends in Computing and Communication 1.11 (2013): 820-824.*
xv. *Kaur, Amanpreet, et al. "Hybrid Approach for Virtual Machine Optimization using BAT Algorithm in cloud." 2021 2nd International Conference on Computational Methods in Science & Technology (ICCMST). IEEE, 2021.*
xvi. *Wallin, Jeffrey J., et al. "Atezolizumab in combination with bevacizumab enhances antigen-specific T-cell migration in metastatic renal cell carcinoma." Nature communications 7.1 (2016): 1-8.*
xvii. *Kumar, Vinod, et al. "Surface modification of WC-Co alloy using Al and Si powder through WEDM: A thermal erosion process." Particulate Science and Technology 36.7 (2018): 878-886.*
xviii. *Sharma, Neeraj, et al. "Wear Behaviour of NiTi SMA Fabricated by P/M: A Taguchi Approach." Nano Hybrids and Composites. Vol. 16. Trans Tech Publications Ltd, 2017.*
xix. *N. Aggarwal, B. Tripathi, D. Chottani and P. S. Tiwana, "Attacks Opportunities in the Cyber Physical Space and the Role of Cybersecurity," 2021 2nd International Conference on Computational Methods in Science & Technology (ICCMST), 2021, pp. 276-280, doi: 10.1109/ICCMST54943.2021.00064.*
xx. *Tiwana, Pardeep Singh, and Nafiza Mann. "Performance Analysis of Jellyfish Reorder Attack on Zone routing protocol in MANET." no. June 2018 (2016).*
xxi. *Rani, Shikha, and Shanky Rani. "Data Security in Cloud Computing Using Various Encryption Techniques." International Journal of Modern Computer Science, ISSN (2016): 2320-7868.*
xxii. *Tiwana, Pardeep Singh, Nafiza Mann, and Gurmandeep Kaur. "Study the impact of different attacks on Zone routing protocol in MANET."*
xxiii. *Sharma, Shalini, et al. "User-Interactive Recommender System for Electronic Products Using Fuzzy Numbers." Advanced Computing and Communication Technologies. Springer, Singapore, 2018. 73-81.*
xxiv. *Solanki, Shano, and Shalini Batra. "Recommender system using collaborative filtering and demographic characteristics of users." International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC) 3 (2015): 4735-4741.*
xxv. *P. S. Tiwana and N. Mann, "Jellyfish reorder attack on hybrid protocol in manet dissection on variegated parameters," 2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2016, pp. 96-100, doi: 10.1109/PDGC.2016.7913122.*
xxvi. *Ochin Sharma, Kiran Mehta, and Renuka Sharma. "Significant Support (SISU): A New Interest Measure in Association Rule Mining." 2021 International Conference on Computational Performance Evaluation (ComPE). IEEE, 2021.*
xxvii. *Mandeep Kaur and Rajni Aron. "An Energy-Efficient Load Balancing Approach for Scientific Workflows in Fog Computing." Wireless Personal Communications (2022): 1-25.*
xxviii. *Daffu, Preeti, and Amanpreet Kaur. "Mitigation of DDoS attacks in cloud computing." 2016 5th International Conference on Wireless Networks and Embedded Systems (WECON). IEEE, 2016*
xxix. *Kaur, Amanpreet, et al. "Load balancing optimization based on deep learning approach in cloud environment." International Journal of Information Technology and Computer Science 12.3 (2020): 8-18.*
xxx. *Kaur, Amanpreet, et al. "Hybrid Approach for Virtual Machine Optimization using BAT Algorithm in cloud." 2021 2nd International Conference on Computational Methods in Science & Technology (ICCMST). IEEE, 2021.*
xxxi. *Rathore, P. S., Chatterjee, J. M., Kumar, A., & Sujatha, R. (2021). Energy-efficient cluster head selection through relay approach for WSN. The Journal of Supercomputing, 77, 7649-7675.*
xxxii. *Rani, Shalli, Syed Hassan Ahmed, and Ravi Rastogi. "Dynamic clustering approach based on wireless sensor networks genetic algorithm for IoT applications." Wireless Networks 26 (2020): 2307-2316*