Improved Novel Clustering Technique for Diverse and Self-motivated Traffic Data Stream for IoT Scenario

Akhila Nadimpalli, R. Shiva Shankar, Chalapathi Raju K, A. Harish Varma

Abstract*: Technologies are changing day by day and IoT is worldwide data and may of great business important to various users. sTo create such reasonable data, majority adaptive and K-mediod clustering techniques are employed in data mining. In research work, it focus on comparing adaptive, K-medisod and novel clustering technique to internet-of-things data collection in ITSs (Intelligence Traffic System). In traffic DataStream is composed form online site, it challenges of 30,000 instances with 9 attributes, clusters formed after evaluation and number of clusters is identified after the evaluation. Proposed techniques are significant too easy than some other clustering techniques with respect to all computation recall and precision parameters. In traffic databases depends on the data separation and cluster enhancement that is quality of clusters. To resolve the major issues that over load the system or Centre's in IoT which consequences the huge kind of data on internet. It evaluated a set of consequences experiments using token and manufacture data from traffic use case view where the traffic considerations from the city monitor. Comparison of clustering methods that helps in determining suitable clustering approach for the offer internet of things database which results in optimal performance metrics.

***Index Terms*: Clustering Approach, IoT, Intelligent Traffic Systems, precision and recall.**

I. INTRODUCTION

The internet of things is used to identify the physical objects which is connected to internet and also with other devices is called as the internet of things. The correlation to computing devices, mechanical, objects, people and digital machine that delivered by the unique identifiers is the internet of things The communication may occur between the network without the presence of the human to human and human to computer interaction[1] .

Data mining is the process of searching the large sets of the data to identify the patterns and the establish relations among them for the analysis of the data. The basic tasks of the data mining are the searching of the data and the finding of the useful data [2]. The main tasks of this method was finding the pattern and the pattern can be used for the growth process in The business organizations. The basic steps involved in the data mining process are described as [3]:-

i) Investigation: - The data Investigation is transformed in to another form of the variables and the problems of the nature based is determined in this approach.

ii) Pattern Recognition: - The data are distinguished for the specific variables and the next procedure is the pattern identification. The patterns are identified and then prediction of the patterns.

iii) Positioning: - The patterns can be positioned for the desired outcome.

II. LITERATURE SURVEY

Daniel Puschmann, et al., 2017[4] discussed on different clusters in a data stream based on data distribution. The numbers of clusters of the incoming data from the data stream are selected using clustering method. The data drift are adjusted based on clustering algorithm with data drift. The stream of data analyzed using data drift and the silhouette coefficient can be measured using the quality of the clusters. The number of clusters based on data distribution computes the adaptive clustering method. In this paper, the calculation of the data stream based represented by high frequency. The clustering mechanism based on data drifts where centroid changes using clustering method. **Hind Bangui et al., 2018[5]** analyzed on specific aspect of clustering algorithms in big data. The key factor of clustering indicates fusion of big data, mobile environment, cloud computing and IOT technologies. The clustering algorithm in big data was used in Internet of Things. The IOT data have large complexity; variety creates a main role in daily life. The different clustering methods supervised structured data, input data and extract information from IOT sources. In this paper, the different clustering algorithm in big data was implemented to overcome the problem. **Praveen Kumar et al., 2016[6]** studied on various clustering algorithms of data stream. With the advent of new technologies there is large boost in the generation of the data streams by IOT applications. The advancement in data stream clustering become more complex because for the processing of data stream only one scanning is possible ad it changes with the time so overcome new clustering data stream algorithm was developed. The traditional data is different from the data steams because of the dynamic processing of the data. The data streams generated by IOT applications proved a great importance for extracting information, algorithms of data mining applied for

Revised Manuscript Received on July 09, 2019

Akhila Nadimpalli, Computer Science Engineering, SRKR Engineering College, Bhimavaram, India.

R. Shiva Shankar, Computer Science Engineering, SRKR Engineering College, Bhimavaram, India.

Chalapathi Raju. K, Electronics and Communication Engineering, SRKR Engineering College, Bhimavaram, India.

A. Harish Varma, Electronics and Communication Engineering, SRKR Engineering College, Bhimavaram, India.

further generation of data streams. The main method of data mining is clustering of data stream with same data objects and detecting outliers. **Fatos Xhafa et al., 2015[13]** reviewed on heterogeneous clustering of big data Stream. In this paper, by using a heterogeneous cluster for Big Data Stream from flight radar24 global monitoring system, a consistency is achieved in the system. Flight radar24 comprises of the thousands of the flights which are tracked simultaneously which provide real time data of aircrafts around the world. In the traditional computing system, huge amount of data may not able to process data in the data stream. So some part of the information may be lost. This is one of the main issues in the processing of the data so there is need of the consistency in the stream of data. In this paper, different frameworks of data stream were proposed to improve the consistency of the data stream. **Yunbo Li et al., 2017[14]** detailed explanation on the production of the on-site renewable energy in different edge clouds for improving solution in IOT systems. In this paper, a model was proposed for computing the objects to the edge which depends on the availability of the renewable energy. In IOT applications the main challenges are the huge amount of data streaming and computing resources. In this paper, they focused on internet of vehicles (IOV) as the convergence of the mobile internet and IOT. The video streams from the camera analyzed for the object detection and tracking of the objects. So to overcome the challenges, to leverage on-site renewable energy production in the different edge cloud nodes to green Internet of Things applications.

2.1 Suggestion Drawn from Literature Survey

After that the various papers studied and analyzed the following inferences are drawn:

(i) Authors have implemented clustering methods to mine the information in internet of thing environment.

(ii) K-mediod clustering technique is consuming a more time for execution, while random clustering is speedily one as compare to all other clustering techniques.

(iii) Rapid growth in Internet-of-thing world, it is complex to manage the data streams so more robust techniques will be required to perform clustering.

(iv) Information is not safe and secure.

III. BACKGROUND

3.1 Clustering in Data Mining

Clustering is the process for the analysis of the data has the wide applications in computer science, biology and statics. The clustering is the grouping of the sets of the data which are differentiated in to classes with the clusters may be same[7] Clustering has a vital part in collection of fields of science and other sociologies, science, measurements, design reaction, data retrieval, machine knowledge and information mining. The Clustering can be divided in to various categories which are, partitioning methods, density-based methods, network based methods, model-based methods, constraint-based approach and clustering algorithms [8].

The Heuristic-Based Clustering is described below:

The parameters used describe the residual energy and the neighbour count in heuristic based approach. The neighbours describe the connection of the nodes. The nodes are deployed in the network and the network can be dynamic and movable [9]. The steps are described as:-

i) In Broadcasting, each node broadcasts the packets with IP address to neighbours nodes in the network.

ii) The multicasting is performed where the nodes sends the residual energy and neighbour count to the neighbour nodes.

iii) In this the cluster head is selected and formatted, every node receive the data from the neighbours nodes and maintain the neighbour list.

3.2 Classification in Data Mining

The functions that map a data items in to different predefined classes based on the set of the attributes. The decision tree matches model without any loss of the accuracy [10]. The classification is described as:-

i) Decision Tree: Decision Tree is flow chart which determines each node has the attribute value and each branch had the outcome of the test and leave describes the class distribution. A decision tree is a predictive model of the input space in to cells. Each node in decision tree has input variables and branches from the nodes maps the result to the test. The decision tree is a predictive model that can be observed as a tree where both branch of the tree is a classification and leaves describes the classification of the data set.

ii) Classification and Regression Tree (CART): The method of constructing the decision tree is the Classification and Regression Tree. The splitting of the nodes at each node based on the function of the single attribute. The diversity of the nodes decreases if there is absence of the nodes. The last phase of the growing phase, the best training set has been allocated to certain leaf of the full decision tree [11].

iii) ID3approach: The algorithm based on the hunts algorithm and the tree builds on the two phases: - The phases are the building blocks measure the splitting attributes of the nodes. The presence of the noise in the nodes can be removes using pre-processing technique. The data can be computed with high data as root nodes and possible values of the attributes. The attributes can be handled using ID3 algorithm by threshold values of the splitting tracks [12].

IV. IMPLIMENTATION WORK

In this section, we design a three main objectives in IoT Big Data Streams described in below are:-

Analyzed and Study the various clustering techniques in Big data Stream in IoT. Develop an adaptive clustering Technique with feature extraction then to calculate the performance which is divided clusters. Develop a K-mediod Clustering algorithm to speedily data divide into cluster format. Implement a Novel Clustering approach to improve the divide data and Devise a technique that energetically offers comparative study with existing clustering techniques. To perform the experimental results, conclude and verify the clustering strategy and calculate the performance metrics and compare it. Initialize, it gathers the database from the Machine Learning repo. Site and download the Traffic dataset in Internet-of-things. Upload the Traffic DataStream and all attributes are shown in table format.

Second phase is in pre-processing, to evaluate the intensity of data steam and calculate the small and large value from the dataset. Implement an Adaptive K-mean clustering technique to divide the data into two phase's i.e., Cluster1 and Cluster2. Main aim is to separate the k-assumptions into n-groups in which individual assumption belongs to the cluster set with the nearest sum, segment as a proto-type of the data cluster. K-mediod Clustering Technique selects the data points to be mediod forms. In novel approach to implement a filter the divide data into again two phases cluster 1 and cluster 2. Evaluate the performance metrics and compare the other parameters.

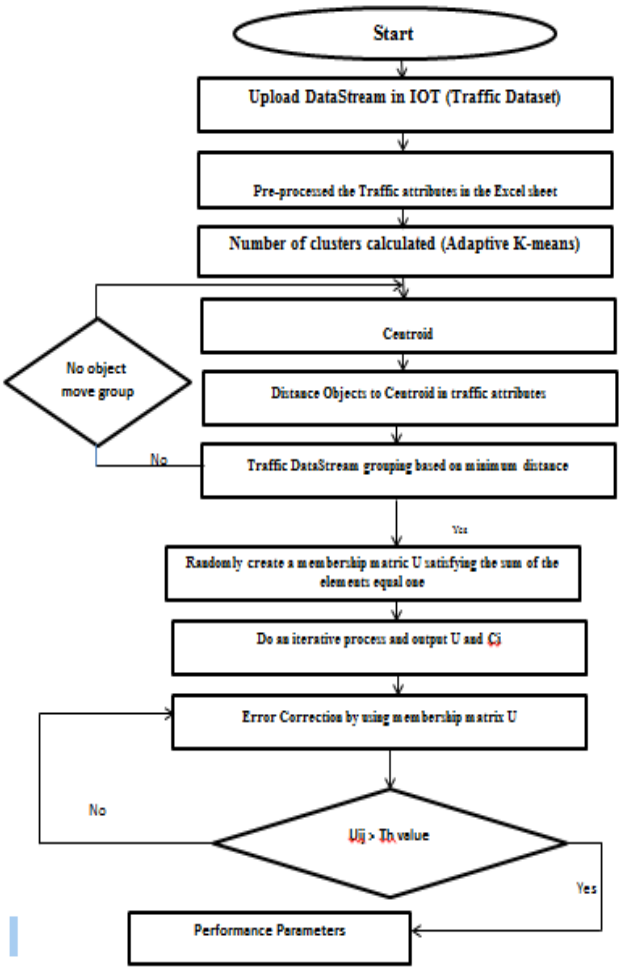

Figure 1: Implementation Flow Chart

V. EXPERIMENTAL RESULTS

In this section, proposed and comparison techniques have been developed using MATLAB 2016a simulation Tool used. Simulation is utilized for pre-processing traffic db and its performance metrics compare with adaptive and K-means data mining techniques. In traffic DataStream is composed form online site, it challenges of 30,000 instances with 9 attributes, clusters formed after evaluation and number of clusters is identified after the evaluation. Proposed techniques are significant too easy than some other clustering techniques with respect to all computation recall and precision parameters.

In this proposed work defined that the framework pattern using GUI tool used in MATLAB 2016a in this traffic manager database download from the UCI machine learning repository site. These phases are divided and explained in above implementation section.

(i) Dataset uploading
(ii) Data Pre-processing
(iii) Base Paper Algorithm implementation (Adaptive Clustering)
(iv) K-mediod Clustering Method and
(v) Design a novel algorithm which is novel clustering algorithm
(vi) Performance Metrics
(vii) Comparison.

Representation the traffic dataset stream and data pre-processing data undesired information happens when no data esteem in put away for the data-variable in a perception. Unwanted information is a typical event and could importantly affect the conclusions that can be drawn from the database.

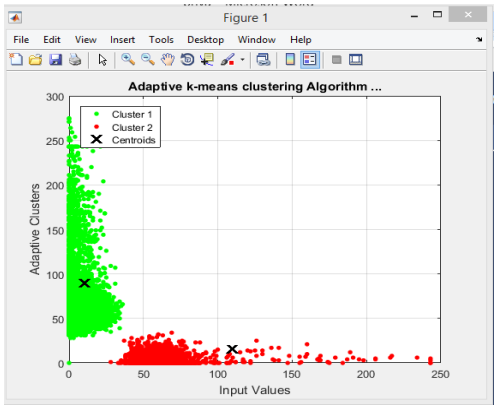

Figure 2: Adaptive Clustering Algorithm

Above figure defined that the clustering technique consequences. The clustering technique is used to handle the database from the efficient retrieval. One of the main problems in clustering is dependency on adaptive k-means algorithms like that the information is separated into k-clusters. In this method the several numbers of clusters is defined and the technique is huge dependent on the initial verification of data elements that define the cluster forms. The large area of research work is clustering has major focused on improving the clustering techniques and procedure like as a clusters dependent on verification data. Adaptive clustering technique that represents that the procedure the clusters without regards to initial selection of cluster defined. This method described that can identify the k-clusters in an input data stream in internet of thing by combining the previous clusters and the generating novel ones, while keeping several clusters constant. It method has been used to obtained a highly speed of exploring process when other efficient search techniques may not be available.

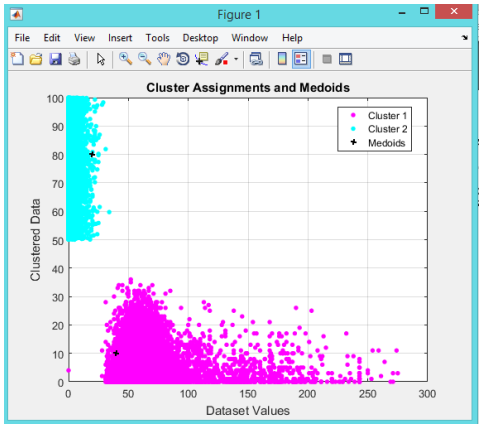

Figure 3: K-mediod clustering

In this figure shows that the more robust as computed t k-means as in K-mediod, it search as a representation object to minimize the sum of difference of Db, whereas the cluster is utilized avg_of_sq ED for dataset objects. Distance metrics reduce the interference and ERs.

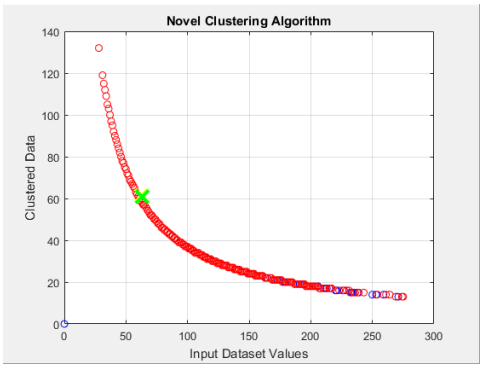

Figure 4: Novel Clustering Approach

Above figure shows that the Novel Approach by allocated membership to each data values corresponding to each cluster Centre on the normal distance among the cluster Centre and d value. Member-ship of individual data value shall be equal to 1.

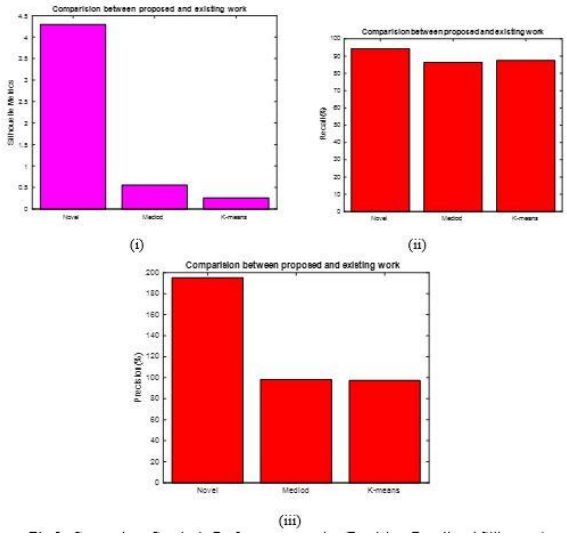

Figure 5: Comparison Graphs in Performance metrics (Precision, Recall and Silhouette)

Above fig 5 improves the performance parameters in Silhouette, Recall and Precision rate with K-mediod, Adaptive and novel clustering technique. In K-mediod and Novel clustering Technique offers the high value of the coefficient parameters.

Table 1: Comparison

Performance Metrics	Values (Proposed Work)	Values (Existing Work)
Silhouette Metrics	0.388	0.91
Recall	94	90
Precision	96	95

Above table 2 shows that the comparison in performance metrics in proposed and existing work.

VI. CONCLUSION AND FUTURE SCOPE

In this research article an effective technique for handling large IoT databases using clustering is represented. In this conclusion here has overviewed an adaptive and K-mediod clustering method that is developed for dynamic traffic data. The implemented model is also able to determine the amount of classifiers found searched inherently. In traffic databases depends on the data separation and cluster enhancement that is quality of clusters. To resolve the major issues that over load the system or centres in IoT which consequences the huge kind of data on internet. It evaluated a set of consequences experiments using token and manufacture data from traffic use case view where the traffic considerations from the city monitor. Enhance the performance metrics like as a Recall, Precision and Silhouette in traffic data streams.
In future scope, will implement an optimization and encryption technique to enhance the memory storage. It helps in eliminating replicated databases and effective using the storage capacity of data points.

REFERENCES

1. Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications–A decade review from 2000 to 2011. *Expert systems with applications*, *39*(12), 11303-11311.
2. Bharathi, A., & Deepankumar, E. (2014). Survey on classification techniques in data mining. *International Journal on Recent and Innovation Trends in Computing and Communication*, *2*(7), 1983-1986.
3. Da Xu, L., He, W., & Li, S. (2014). Internet of things in industries: A survey. *IEEE Transactions on industrial informatics*, *10*(4), 2233-2243.
4. Puschmann, D., Barnaghi, P., & Tafazolli, R. (2017). Adaptive clustering for dynamic IoT data streams. *IEEE Internet of Things Journal*, *4*(1), 64-74.
5. Bangui, H., Ge, M. and Buhnova, B.. Exploring Big Data Clustering Algorithms for Internet of Things, Applications, *In Proceedings of the 3rd International Conference on Internet of Things, Big Data and Security (IoTBDS 2018), pages 269-276.*

6. Praveen Kumar, Data Stream Clustering in Internet of Things, *SSRG International Journal of Computer Science and Enginee ring (SSRG-IJCSE) – volume 3 Issue 8–August 2016.*
7. Xu, Y., Qu, W., Li, Z., Min, G., Li, K., & Liu, Z. (2014). Efficient $ k $-Means++ Approximation with MapReduce. *IEEE Transactions on Parallel and Distributed Systems*, *25*(12), 3135-3144.
8. Tao, X., & Ji, C. (2014, November). Clustering massive small data for iot. In *The 2014 2nd International Conference on Systems and Informatics (ICSAI 2014)* (pp. 974-978). IEEE.
9. Atzori, L., Iera, A., & Morabito, G. (2010). The internet of things: A survey. *Computer networks*, *54*(15), 2787-2805.
10. Uckelmann, D., Harrison, M., & Michahelles, F. (2011). An architectural approach towards the future internet of things. In *architecting the internet of things* (pp. 1-24). Springer, Berlin, Heidelberg.
11. Nallaperuma, D., De Silva, D., Alahakoon, D., & Yu, X. (2017, October). A cognitive data stream mining technique for context-aware IoT systems. In *IECON 2017-43rd Annual Conference of the IEEE Industrial Electronics Society* (pp. 4777-4782). IEEE.
12. Khalilian, M., & Mustapha, N. (2010). Data stream clustering: Challenges and issues. *arXiv preprint arXiv:1006.5261*.
13. Xhafa, F., Naranjo, V., Barolli, L., & Takizawa, M. (2015, September). On streaming consistency of big data stream processing in heterogenous clutsers. In *Network-Based Information Systems (NBiS), 2015 18th International Conference on* (pp. 476-482). IEEE.
14. Li, Y., Orgerie, A. C., Rodero, I., Parashar, M., & Menaud, J. M. (2017, May). Leveraging renewable energy in edge clouds for data stream analysis in IoT. In *Cluster, Cloud and Grid Computing (CCGRID), 2017 17th IEEE/ACM International Symposium on* (pp. 186-195). IEEE.

AUTHORS PROFILE

Pursuing M.Tech in the Department of CSE, SRKR Engineering College. B.tech from BVRIT college of Engineering for women, Hyderabad.

Working as Assistant professor in the Department of CSE, SRKR Engineering College.

Working as Assistant professor in the Department of ECE, SRKR Engineering College.

Working as Assistant professor in the Department of ECE, SRKR Engineering College, my area of interest is biomedical image processing and IoT.