ISSN 2278-3091
Volume 10, No.2, March - April 2021
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse261022021.pdf
https://doi.org/10.30534/ijatcse/2021/261022021

A Comparative Review of Incremental Clustering Methods for Large Dataset

Arun Pratap Singh Kushwah[a], Shailesh Jaloree[b], Ramjeevan Singh Thakur[c],
[a]Dept. of Computer Science,, BU,Bhopal, Madhya Pradesh, India, apsingh.kvdamoh@gmail.com
[b] Dept. of Appl. Maths & Computer Applications, SATI, Vidisha, Madhya pradesh, India, shailesh_jaloree@rediffmail.com
[c] Dept. of Computer Applications, MANIT, Bhopal, Madhya pradesh, India, ramthakur2000@yahoo.com

ABSTRACT

Several algorithms have developed for analyzing large incremental datasets. Incremental algorithms are relatively efficient in dynamic evolving environment to seek out small clusters in large datasets. Many algorithms have devised for limiting the search space, building, and updating arbitrary shaped clusters in large incremented datasets. Within the real time visualization of real time data, when data in motion and growing dynamically, new data points arrive that generates instant cluster labels. In this paper, the comparative review of Incremental clustering methods for large dataset has done.

Key words: DBSCAN, dynamic data, Incremental clustering, K-means.

1.INTRODUCTION

The emergence of IOT technology gives the increasing use of the web connected IOT devices causes an impetus to the quantity and disparity of knowledge. The quantity keeps increasing with every information exchange over the web or maybe the minuscule IoT objects we use [1][2].Today, most of the devices are connected online.

Now, if a corporation wants to gather information online, it needs a rigorous process because the info generated are going to be massive also there could also be disparity within the format of knowledge . This adds up to the complexity in most sort of data viz. Structured, semi-structured, or unstructured data [3]. Handling efficiently such amounts of knowledge may be a big challenge and traditional methods and tools aren't found suitable to affect very large dynamic high dimensional datasets [6].

The characteristics of huge data, including high volume and dynamically evolving [3], created a new challenge for the researchers because this type of data cannot be analyzed using classic data mining techniques. Rather it requires a clustering method that can rapidly partition continuously arriving datasets and effectively update the cluster structures.

2. INCREMENTAL CLUSTERING

In the situation where data is coming dynamically, it is impossible to cluster all those data at once. In non-incremental clustering, we need re-cluster the data collected newly or updated in database periodically. The incremental clustering technique uses previous resulted clusters to classify new data points. There are many challenges also in the way of Incremental clustering: (1) accurate prior clusters in the absence of enough datasets, (2) time taken for cluster updating, (3) prone to outliers etc. Seeking these challenges a lot of work still to do.

The Incremental clustering process is given in figure.1. we can see that, in step (1) the dataset is preprocessed for clustering, step (2) clustering techniques applied on dataset, step (3) new data point arrived in the database, step (4) distance is calculated from previous clusters resulted using similarity measure techniques, (5) assign new data point to the appropriate cluster or create new cluster.

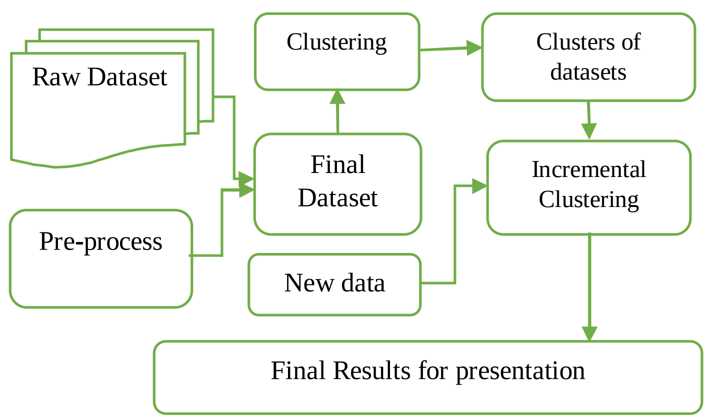

Figure1: Incremental Clustering Process.

Some popular Incremental Clustering Techniques are:

- **Incremental K-means [1],[48].**
- **Incremental DBSCAN [13].[36],[55].**
- **Incremental Fuzzy Clustering [21,[25],[37],42].**
- **Incremental Genetic Algorithm [30],[50],[51].**

There are many incremental clustering methods are evolving these days.

3.REVIEW OF INCREMENTAL CLUSTERING TECHNIQUES

Incremental clustering or dynamic pattern recognition aims to adapt the training models according to the arriving data without forgetting the acquired knowledge. Incremental

Clustering transfers the acquired knowledge from different batches to classify the test data. It helps to grow the pattern recognition capacity by overcoming the problem of catastrophic interference i.e., successive training of each batch of newly arrived data causes the clusters to forget the previously acquired knowledge partially or completely. This type of learning can be applied on the data set of incremental nature to provide accurate results addressing the issues of limited computational resources such as memory and time [58].

The example of Incremental clustering is given in the figure.2. In the figure, A and B are the clusters generated from database and C is the new dataset, case (i), (ii) and (iii) represents the cluster assignment for new data points.

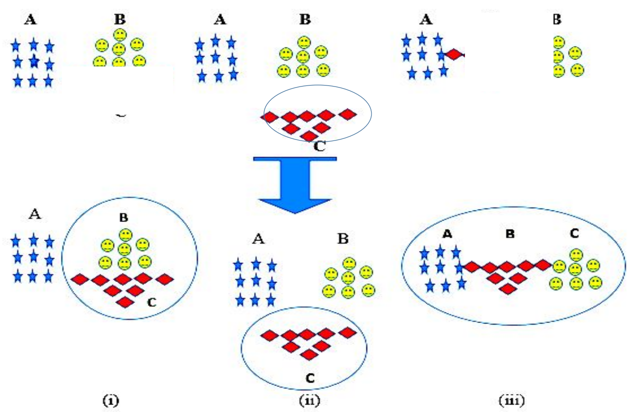

Figure2: Incremental Clustering example

Incremental clustering algorithms work by processing data objects one at a time, incrementally assigning data objects to their respective clusters while they progress [3][4].

An automated process requires helping and managing this growing large data and information. So the clustering system is being investigated that can cluster data from disparate sources into event-centric clusters. In the past decade, many incremental clustering methods have proposed to fulfill requirements for clustering of large incremental data.

A comparative review of Incremental Clustering Algorithms for large dataset has done in this paper, based on the different parameters such as clustering methods used, complexity, dimensionality of data used, outlier effect on the result, type of data domain, the volume of data supported, nature of algorithm, dynamic updation and year published with corresponding authors, given in Tabular form for the shake of convenience in Table 1.

Table 1: A comparative Review of Incremental Clustering Techniques

Sl No.	Techniques Applied	Nature of Algorithm	Complexity	High Dimensional	Outlier Effect	Dataset	Suitable for Large Dataset	Incremental	Dynamic Update (Y/N)	Noise Reduction (Y/N)	Ref./ Year
1	Incremental K-means	Cluster Centre Jumping Operation	O(K2 * n * no. of iterations)	NO	NO	Real & Imaginary	YES	YES	N	Y	[1] 2004
2	Spectral Clustering / Stochastic Block Model (SBM)	Filtering random signals on the graph	O(kn log n)	NO	-	real-world dataset: RCV1 (Reuters Corpus Volume I)	YES	YES	Y	-	[11] 2020
3	Fully-Unsupervised PCM (FU-PCM)	Pearson Correlation Coefficient for describing the global structure of samples:		NO	NO	Real & Artificial Data sets	YES	YES	Y	-	[12] 2020
4	Inc Any DBC	Work-efficient parallel method. Parallel dynamic clustering. unique anytime work-efficient technique	O(n2)	YES	YES	Very large real, dynamic and synthetic datasets	YES	YES	Y	Y	[13] 2019
5	Graph Convolutional Network (GCN)	an incremental face clustering method	O(n)	YES	YES	Very large real, dynamic and synthetic datasets	YES	YES	Y	Y	[14] 2021

6	SVM	incremental spam images filtering (ISIF) approach ISIF approach with a SVM filter & higher-order local autocorrelation approach	$O(n^2)$	YES	YES	Personal Spam Image Dataset, PSID , real dataset for spam image	YES	YES	NO	N	[15] 2011
7	Graph-based incremental clustering	Probabilistic encoding	Bloom filters and counting Bloom filters	NO	NO	Multiple large real datasets	YES	YES	NO	YES	[17] 2020
8	Nearest Neighborhood Assignments & K-Medoids	Efficient Incremental Clustering by Fast Search driven Improved K-Medoids (EICFS-IKM)	$n \log_2 n$	YES	YES	large real Dynamic datasets	YES	YES	NO	YES	[20] 2019
9	stochastic average gradient algorithm with less memory	Non-replacement sampling strategy	Non-replacement sampling strategy sampling strategy	YES	YES	large real Dynamic datasets	YES	YES	YES	YES	[22] 2018
10	Incremental Conic Functions (ICF) algorithm (k-means based)	polyhedral conic functions Generation	polyhedral conic functions Generation	YES	YES	large real Dynamic datasets	YES	YES	NO	YES	[23] 2018
11	uCLUST and ELM++. MEHOD-ELM	CUIL (Classification of Unstructured data using Incremental Learning) framework	a framework CUIL which adopts Extreme Learning Machine (ELM) neural networks	NO	NO	Binary and multi-class datasets	YES	YES	NO	YES	[24] 2018
12	Incremental fuzzy clustering	Single-Pass FCMdd based on DTW (spFDTW) and Online FCMdd based on DTW (oFDTW)	O(n)	NO	NO	Binary and multi-class datasets	YES	YES	NO	YES	[25] 2018
13	UPFICC Unsupervised Parameter-Free Incremental MULTI-MODAL CO-CLUSTERING	Multi-modal Cyber-Physical-Social Systems (CPSS)	O(N)	YES	NO	multi-modal datasets collected from CPSS	YES	YES	NO	NO	[26] 2017
14	fast finding and searching of density peaks based on kmediods (ICFSKM)	Improvised fast finding and searching of density peaks (CFS)	k-mediods is employed to modify the clustering centers	YES	YES	large real Dynamic datasets	YES	YES	YES	YES	[27] 2017
15	An incremental clustering method based on the CFS.	CFS clustering by fast search and find of density peak. An enhanced cluster adjustment strategy. multiple representatives based partitioning one-time cluster splitting-merging strategy	Maximum min-distance mechanism, based on convex hull theory.	NO	NO	Real-world image datasets	YES	YES	YES	NO	[28] 2017

16	Evolving Fractal-based Clustering of Data Streams (eFCDS)	framework e FCDS - Evolving Fractal-based Clustering of Data Streams using fractal Dimension designed to analyze evolving data streams	Cluster multivariate data streams based on their evolving behavior over time.	YES	YES	multivariate data stream Real-world dynamic datasets	YES	YES	YES	YES	[29] 2016
17	improvement of the BFR algorithm, with data streams composed by data points of "medium" dimension	local distance approach, based on t he computation of the Mahalanobis distance	Local Metric, Based On The Mahalanobis Distance. A Technique Based On Hierarchical Clustering	YES	YES	multivariate data stream Real-world dynamic datasets	YES	YES	YES	YES	[30] 2015
18	Incremental concept ual hierarchical text clustering approach using CFu-tree (ICHTC-CF) representation.	A novel incremental conceptual hierarchical text clustering method using CFu-tree	Term-Based Feature Extraction	YES	YES	large real Dynamic datasets	YES	YES	YES	YES	[32] 2015
19	An enhanced version of the incremental DBSCAN algorithm	Incremental version of DBSCAN and incremental similarity-histogram based clustering algorithm (specific for documents clustering)	Incremental clustering process by limiting the search space	YES	YES	large real Dynamic datasets	YES	YES	YES	YES	[36] 2015
20	IGrid	Partition the grid space by dimensional radius in a dynamic and incremental manner	Partition the grid space by dimensional radius in a dynamic and incremental manner	YES	YES	Real datasets and synthetic datasets	YES	YES	YES	YES	[44] 2011
21	Boundary-vector-based boundary point detection (BV-BPD) algorithm.	A boundary-profile-based incremental clustering (BPIC) method to find arbitrarily shaped clusters with dynamically growing datasets.	$O(n^2)$	YES	YES	Real static and dynamic datasets	YES	YES	YES	YES	[45] 2018
22	NJW Spectral Clustering Algorithm. incremental spectral clustering algorithm have been introduced with k-Eigen vector	Incidence vector used to update the change of data in the form of Eigen-system to keep a newest set of representative point	Employed ARPACK (a variant of Lanczos method) to compute the spectrum of D−1L	YES	NO	Real evolutional data sets	YES	YES	NO	NO	[47] 2011

24	ICFS with multiple representatives (ICFSMR) and the enhanced ICFSMR (E_ICFSMR)	Multiple representatives-based ICFS clustering approach to rapidly partition new arrivals into current clusters.	ICFSMR based on convex hull theory to modify the representatives identified for each cluster.	YES	NO	Large real Dynamic datasets	YES	YES	YES	NO	[49] 2010
25	ICGA is density-based in nature. Currently works only for insertion operation	Incremental Clustering in Data Mining using Genetic Algorithm	Genetic algorithm Based on the formal definition of clusters for rmetric database	YES	NO	Any database containing data from a metric space,	YES	YES	NO	NO	[50] 2010
26	Incremental spatial clustering algorithm, based on density clusters ISC(GA-RTree).	The density based algorithm is applied to find a cluster for a new spatial object. Euclidian distance has been applied.	It is based on GA and R-tree structure to solve a clustering task in spatial data mining.	NO	NO	Shape datasets (Spatial Data Set)	YES	YES	NO	NO	[51] 2012
27	new Incremental Induction of Multivariate DT algorithm for large datasets (IIMDT),	IIMDT processes the training objects one by one, in this way only the processed object must be kept in main memory at each step.	It is an algorithm for inducing decision trees for large numerical datasets	YES	NO	large real Dynamic datasets	YES	YES	NO	NO	[52] 2008
28	Algorithm employed GA-based clustering methods and Two techniques, one with fixed number of clusters and another with a variable number of fuzzy clusters	FVGA-clustering technique can be attributed to the use of both genetic search and the optimizing Criterion.	The fuzzy c-means (FCM) algorithm is a fuzzy counterpart of the K-means technique	YES	NO	Large real Dynamic datasets , (Numerical as well as image) data	YES	YES	NO	NO	[53] 2011
29	The partial/merge k-means as a set of data stream operators, employed k-means in a highly scalable way in order to cluster massive data set efficiently.	Data stream based Approach to clustering, the partial/merge k-means algorithm based on based on the data Stream paradigm.	clustering technique that produces the multivariate histograms	YES	NO	Large real Dynamic datasets , (Numerical , Stream as well as image) data	YES	YES	YES	NO	[54] 2003
30	Sorting Genetic Algorithm II (NSGA-II)& parallel incremental DBSCAN	Density-Based Clustering and Classification	Suitable fitness functions for both labeled and unlabeled datasets.	YES	YES	Large real Dynamic datasets , (Numerical , Stream as well as image) data	YES	YES	YES	YES	[55] 2021

4.FINDINGS

We have reviewed many literatures available in different sources; a year wise graph is given in the figure 3, in which it is clear that most of the work done in the area of incremental clustering has grown in recent years.

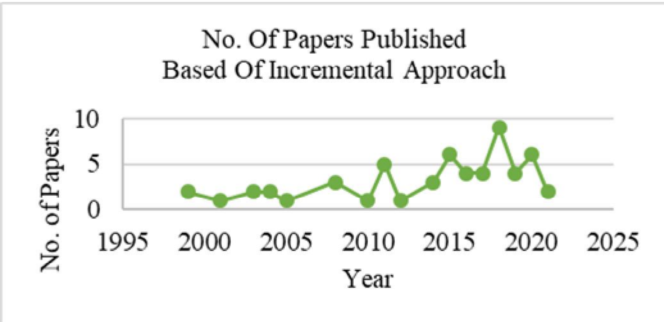

Figure 3: Year wise graph of literature published on Incremental clustering.

In Figure 4 the effectiveness on outliers is shown w.r.t. the No. of papers.

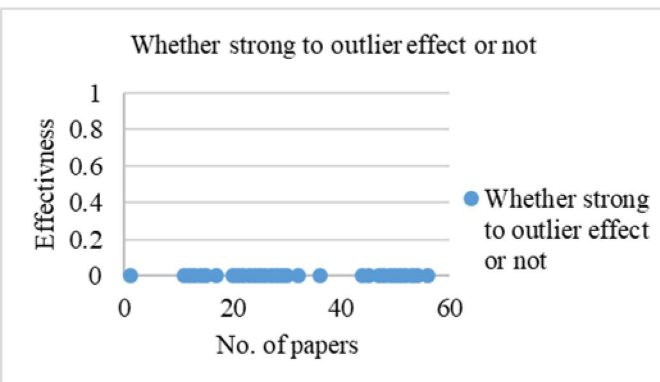

Figure 4: Effective on Outliers.

In figure 5, shown the percentage of papers that can handle the high dimensional datasets.

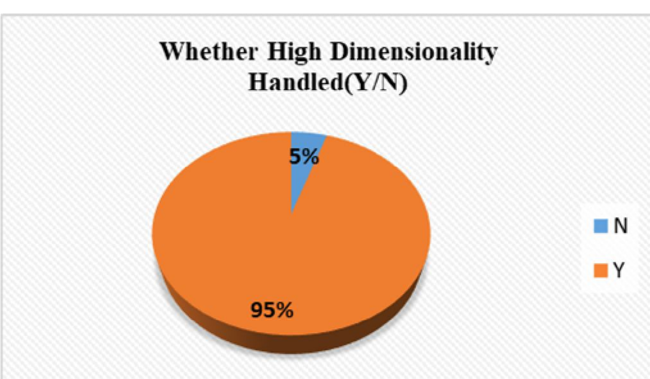

Figure 5: Percentage of paper Handles High Dimensionality

5.CONCLUSION

Authors in different publications for incremental clustering presents numerous techniques. Some of them conversed with reasonable accuracy and some conversed with compromised accuracies. The requirement of an efficient clustering technique for incremental large dataset is still required due to the advancement in dynamic data generation with optimum speed. Further, we are seeking to present a framework for incremental data clustering to enhance the acceptability and accuracy of predictions. By considering this study as base further also try to propose a method for the incremental clustering.

REFERENCES

1 D. T. Pham, S.S. Dimov and C. Nguyen. **An Incremental K-means algorithm**, In Proc. of the Institution of Mechanical Engineers In: Part C: Journal of Mechanical Engineering Science, vol.218, no.7, pp. 783–795, 2004.

2 C. Dobre and F. Xhafa. **Intelligent services for Big Data science**, Future Generation Computer Systems, vol.37, pp. 267, 2014.

3 A. Zappone, M. Di-Renzo and M. Debbah. **Wireless Networks Design in the Era of Deep Learning: Model-Based, AI-Based, or Both?,** IEEE Trans. on Communications, vol.67, no.10, pp. 7331- 7376, 2019.

4 C. Sreedhar, N. Kasiviswanath and P.C. Reddy. **A survey on big data management and job scheduling,** International Journal of Computer Applications, vol.130, no.13, pp. 41–49, 2015.

5 Pang-Ning Tan, M. Steinbach, A. Karpatne and V. Kumar. **Introduction to data mining**, 2nd Edition. Published by Pearson Education Limited, 2019.

6 C. C. Aggarwal and C. K. Reddy, **Data Clustering: Algorithms And Applications**,Chapman and Hall/CRC,2018.

7 C. C. Aggarwal, A. Hinneburg, and D. Keim. **On the surprising behavior of distance metrics in high dimensional space**, International Conference on Database Theory (ICDT) -2001, Springer, Berlin, Heidelberg. Lecture Notes in Computer Science(LNCS), vol.1973, pp. 420-434, 2001.

8 K. Beyer and J. Goldstein, U. Shaft and R. Ramakrishnan. **Nearest Neighbors Can Be Found Efficiently If the Dimension Is Small Relative to the Input Size**, International Conference on Database Theory (ICDT) -2003, Springer, Berlin, Heidelberg., Lecture Notes in Computer Science (LNCS), Vol 2572, pp. 440-454, 2003.

9 S. Muthukrishnan**. Data streams: algorithms and applications**, Foundations and Trends in Theoretical Computer Science, vol.1, no.2, pp. 117–236, 2005.

10 A. K. Jain and N. N. Murty and P.J. Flynn. **Data clustering: a review**, ACM Computing Surveys, vol.31, no.3, pp.264–323, 1999.

11 G. Shu-Juan. **Fast Incremental Spectral Clustering in Titanate Application via Graph Fourier**

Transform, *IEEE Access*, vol.8, pp. 57252-57259, 2020.

12 J. Zhang, T. Chen, and Y. Zhang. **Incremental Clustering With Hard Centers**, IEEE MultiMedia, vol.27, no. 4, pp. 102-111, 2020.

13 S.T. Mai , J. Jacobsen, S. Amer-Yahia, I. Spence, N.P. Tran, I. Assent and Q.V.H. Nguyen**. Incremental Density-based Clustering on Multicore Processors**, IEEE Trans. on Pattern Analysis and Machine Intelligence, pp. 1-1,2019.

14 X. Zhao, Z. Wang, L. Gao, Y. Li, and S. Wang. **Incremental face clustering with optimal summary learning via graph conVolutional network**, *Tsinghua Science and Technology*, vol.26, no.4, pp.536-547, 2021.

15 Y. He ,W. Man and H. Haibo, **Incremental clustering-based spam image filtering using representative images,** *International Conference on System science, Engineering design and Manufacturing informatization*, Guiyang, 2011, pp. 323-327,2011.

16 Aarti, G. Sikka and R. Dhir. **Grey Relational Feature Extraction Algorithm for Software Fault-Proneness Using BBO (B-GRA)**, *Arabian Journal for Science and Engineering, vol.*45, pp. 2645–2662, 2020.

17 D. Vatsalan, P. Christen and E. Rahm. **Incremental clustering techniques for multi-party Privacy-Preserving Record Linkage**, Data & Knowledge Engineering, vol.128, p.e.101809, 2020.

18 M. Z. Hossain, M.N. Akhtar,R. Ahmad and M. Rahman. **A dynamic K-means clustering for data mining**, Indonesian Journal of Electrical Engineering and Computer Science, vol.13, issue.2, pp. 521-526, 2019.

19 A. Prahara, D.P. Ismi and A. Azhari. **Parallelization of Partitioning Around Medoids (PAM) in K-Medoids Clustering on GPU**, Knowledge Engineering and Data Science, vol.3, issue.1, pp. 40-49, 2020.

20 S. Balakrishna, M, Thirumaranet, R. Padmanaban and V. Solanki. **An efficient incremental clustering based improved K-Medoids for IoT multivariate data cluster analysis**, Peer-to-Peer Networking and Applications, vol.13, issue.4, pp. 1152-1175, 2019.

21 L.A. Rasyid and S. Andayani. **Review on Clustering Algorithms Based on Data Type: Towards the Method for Data Combined of Numeric-Fuzzy Linguistics**, IOP Conference Series: Journal of Physics. 5th International Conference on Research, Implementation, & Education of Mathematics and Sciences vol. 1097,pe.012082, 2018.

22 J. Huang and Z. Zhou. **Practical Incremental Gradient Method for Large-Scale Problems**, *TENCON IEEE Region 10 Conference*, Jeju, Korea (South), pp. 1845-1848, 2018.

23 E. Cimen, G. Ozturk and O. Gerek. **Incremental conic functions algorithm for large scale classification problems**, Digital Signal Processing, vol.77, pp. 187-194, 2018.

24 S, Madhusudhanan, S. Jaganathan, and L.S. Jayashree. **Incremental Learning for Classification of Unstructured Data Using Extreme Learning Machine**, Algorithms vol.11, pp. 158, 2018.

25 Y. Liu, J. Chen, S. Wu, Z. Liu, and H. Chao. **Incremental fuzzy C medoids clustering of time series data using dynamic time warping distance**, PLOS ONE, vol.13, no. 5,2018

26 L. Zhao, Z. Chen and Y. Yang. **Parameter-Free Incremental Co-Clustering for Multi-Modal Data in Cyber-Physical-Social Systems**, *IEEE Access*, vol.5, pp. 21852-21861, 2017.

27 Q. Zhang, C. Zhu, L.T. Yang, Z. Chen, L. Zhao, and P. Li. **An Incremental CFS Algorithm for Clustering Large Data in Industrial Internet of Things**, *IEEE Trans. on Industrial Informatics*, vol.13, no. 3, pp. 1193-1201, 2017.

28 L. Zhao, Z. Chen and Y. Yang. **Incremental CFS clustering on large data** , *IEEE Global Conference on Signal and Information Processing (Global SIP)*, Montreal, QC, 2017, pp. 687-690, 2017.

29 C.C.Bones, L.A.S. Romani and E.P.M. deSousa. **Improving Multivariate Data Streams Clustering**, Procedia Computer Science, vol.80, pp. 461–471, 2016.

30 G. Aletti and A. Michelett. **A clustering algorithm for multivariate data streams with correlated components**, Journal of Big Data, vol.4, no.1, pp. 48,2017.

31 M. Baydoun, M. Dawi and H. Ghaziri. **Enhanced parallel implementation of the K-Means clustering algorithm**. 3rd International Conference on Advances in Computational Tools for Engineering Applications (ACTEA), Beirut, pp. 7-11, 2016.

32 T. Pengand and L. Liu. **A novel incremental conceptual hierarchical text clustering method using CFu-tree**, Applied Soft Computing, vol.27, C, pp. 269-278, 2015.

33 A. Kaneriya and M. Shukla. **A novel approach for clustering data streams using granularity technique**, *International Conference on Advances in Computer Engineering and Applications*, Ghaziabad, pp. 586-590, 2015.

34 D. Xu and Y. Tian. **A Comprehensive Survey of Clustering Algorithms**, Annals of Data Science (AODS), vol.2 issue.2, pp. 165–193, 2015.

35 M.F. Khan, K. Yau, R.M. Noor and M. Imran. **Survey and taxonomy of clustering algorithms in 5G**", Journal of Network and Computer Applications, vol.154, pp. 102539, 2020.

36 A. M. Bakr, N.M. Ghanem and M.A. Ismail. **Efficient incremental density-based algorithm for clustering large datasets**, Alexandria Engineering Journal, vol.54, issue.4, pp. 1147–1154, 2015.

37 C. Zhang, L. Hao and L. Fan. **Optimization and improvement of data mining algorithm based on efficient incremental kernel fuzzy clustering for large data**, Cluster Computing, vol. 22, no.2, pp. 3001-3010, 2018.

38 I. García, R. Casado, V. García and A. Bouchachia. **An Incremental Approach for Real-Time Big Data Visual Analytics**, 4th IEEE International Conference

on Future Internet of Things and loud Workshops (FiCloudW),2016

39 B. Sarada, M. M Vinayaka and U.V. Rani. **An approach to achieve high efficiency for large Volume data processing using multiple clustering algorithms**, International Journal of Engineering & Technology, vol.7, no.4.5, pp. 689-692. 2018.

40 J. Khalfallah and J. Ben-Hadj Slama. **A Comparative Study of the Various Clustering Algorithms in E-Learning Systems Using Weka Tools**, JCCO Joint International Conference on ICT in Education and Training, International Conference on Computing in Arabic, and International Conference on GeoComputing (JCCO: TICET-ICCA-GECO), Tunisia / Hammamet, Tunisia, 2018, pp. 1-7, 2018.

41 A. K. Toor and A. Singh. **An Advanced Clustering Algorithm (ACA) for Clustering Large Data Set to Achieve High Dimensionality**, International Journal of Applied Information Systems, vol.7,no.2, pp. 5-9,2014.

42 D. Vanisri. **A Novel Fuzzy Clustering Algorithm Based on K-Means Algorithm**, International Review on Computers and Software (IRECOS), vol.9 (10), pp. 1731, 2014.

43 N. A. Barbosa, L. Trave-Massuyeset and V.H. Grisales. **A Novel Algorithm for Dynamic Clustering: Properties and Performance**, *15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Anaheim, CA, 2016, pp. 565-570, *2016.*

44 G. Lei, X. Yu,X. Yang and S. Chen. **An incremental clustering algorithm based on grid**, 8th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), Shanghai, vol.2 PP. 1099-1103, 2011.

45 J. Bao, W. Wang ,T. Yang and G. Wu. **An incremental clustering method based on the boundary profile**, PLOS ONE, vol.13, no.4, 2018.

46 M. Ackerman and S. Dasgupta. **Incremental clustering: The case for extra clusters**, Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'14). MIT Press, Cambridge, MA, USA., pp. 307-315,2014.

47 T. Kong, Y. Tian and S. Hong. **A Fast Incremental Spectral Clustering for Large Data Sets**, *12th International Conference on Parallel and Distributed Computing, Applications and Technologies*, Gwangju, 2011, pp. 1-5, 2011.

48 S. Chakraborty and N.K. Nagwani. **Analysis and Study of Incremental K-Means Clustering Algorithm** In: A. Mantri, S. Nandi, G. Kumar and S. Kumar (eds) High Performance Architecture and Grid Computing. HPAGC 2011. Communications in Computer and Information Science, Springer, Berlin, Heidelberg, vol.169.PP. 338-341, 2011.

49 L. Zhao, Z. Chen, Y.Yang, L.Zou and Z.J. Wang. **ICFS Clustering With Multiple Representatives for Large Data**, *IEEE Trans. on Neural Networks and Learning Systems*, vol.30, no.3, pp. 728-738, 2019.

50 A. Kamble. **Incremental Clustering in Data Mining using Genetic Algorithm**, International Journal of Computer Theory and Engineering vol. 2, no. 3, pp. 326-328, 2010.

51 N. N.Vinh and B. Le. **Incremental Spatial Clustering in Data Mining Using Genetic Algorithm and R-Tree**, In: L.T. Bui, Y.S. Ong, N.X. Hoai, H. Ishibuchi, P.N. Suganthan (eds) Simulated Evolution and Learning(SEAL). Lecture Notes in Computer Science, vol.7673. Springer, Berlin, Heidelberg, 2012.

52 A. Franco-Arcega, J.A. Carrasco-Ochoa, G. Sánchez-Díaz and J.F. Martínez-Trinidad. **New Incremental Algorithm for Induction of Multivariate Decision Trees for Large Datasets**, Springer Berlin, Heidelberg, pp. 282-289, 2008.

53 S. Bandyopadhyay. **Genetic algorithms for clustering and fuzzy clustering**, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol.1, issue.6, pp. 524–531, 2011.

54 S. Nittel and K.T. Leung. **Scaling clustering algorithms for massive data sets using data stream**, 19th international conference on data engineering, sponsored by IEEE Computer Society, Bangalore, India, 2003.

55 E. Azhir, N. J. Navimipour, M. Hosseinzadeh, A. Sharifi, and A. Darwesh. **An efficient automated incremental density-based algorithm for clustering and classification**, Future Generation Computer Systems, 114, pp. 665-678, 2021.