Mind Map based Survey of Conventional and Recent Clustering Algorithms: Learning's for Development of Parallel and Distributed Clustering Algorithms

Rahul Joshi
Research Scholar at SIU
Symbiosis Institute of Technology (SIT), Pune
Affiliated to Symbiosis International University (SIU), Pune, India

Preeti Mulay, PhD
Guide at SIU
Symbiosis Institute of Technology (SIT), Pune
Affiliated to Symbiosis International University (SIU), Pune, India

ABSTRACT
Till date, different papers are available on survey of clustering algorithms. The novel approach used in this paper is use of Mind Maps to present key details about clustering algorithms in visual form. This paper spans from Mind Maps for basic clustering process, similarity and distance indices, evaluation indices, conventional clustering algorithms, recent clustering algorithms, recent parallel and distributed clustering algorithms and key learning's about development of parallel and distributed clustering algorithms.

Keywords
Mind Map; Clustering; Learning; Parallel; Distributed; Algorithm etc.

1. INTRODUCTION
Clustering is grouping of data into clusters. It's mostly of unsupervised type and has applications in the diverse fields but mainly into data mining and machine learning. With advent of different types of data sets, their volumes and as per the specificity of application, form, processing power and involved complexities of clustering algorithms gets changed from time to time. Different authors tried to put survey of clustering algorithms in different ways viz. for specific set of clustering algorithms, through implementation of specific algorithms, inputting particular or different datasets, for a particular paradigm like parallel processing algorithms and to name a few. The motto behind use of mind maps in this survey is to overcome burden of reading lengthier papers and difficulty in getting crisp details. Mind maps are created using coggle.it [79] interface. Used mind maps capture details about clustering algorithms as per category by considering four main dimensions viz. a) basic idea, b) types of algorithms under particular category, c) advantages and d) disadvantages. The broader objective of this paper is to make aware the reader about evolution of clustering algorithms, artefacts required for development of parallel and distributed data clustering algorithms aimed for large datasets. This paper surveys twenty six algorithms under nine categories of conventional algorithms, twenty seven algorithms under ten categories of recent algorithms and twenty algorithms of six type's algorithms implemented parallel or distributed for clustering of large data. So, in all survey of seventy seven clustering algorithms is carried out.

2. CLUSTERING ALGORITHMS
There are different definitions for clustering algorithm. The definition for clustering algorithm is complete in all sense when following factors are taken into account [1]:

1] Same cluster contains similar type of instances and different clusters have instances of different types;

2] Similarity and distance measurements must be apparent and realistic;

3] Evaluation indices must be appropriate.

The mind map in figure 1 shows clustering process. It comprises of six steps [2]. After inputting raw data set, the pre-clustering phase plays an important role in removal of anomalies from input. The clustering algorithm is selected as per the characteristics of undertaken problem. The formed clusters need to be validated to evaluate clustering results. The post clustering phase confirms suitability of the algorithm for considered problem. Lastly cluster needs to be stored either in shared or distributed database.

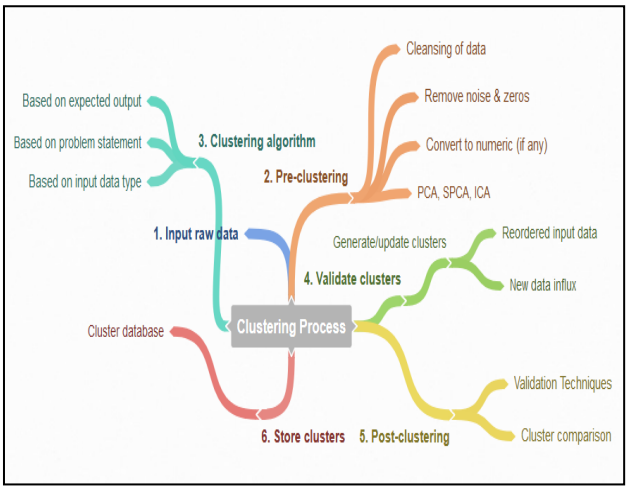

Fig. 1 Mind Map for Clustering Process

The outline of remaining part of the paper is section 3 elaborates common similarity and distance indices for evaluating data attributes, evaluation indices in section 4, comprehensive analysis of conventional and recent clustering algorithms in section 5 and 6 respectively, parallel and distributed data clustering algorithms for large data and key learning for design of a parallel and distributed algorithm in sections 7 and 8 respectively and the outlook in section 8.

3. SIMILARITY AND DISSIMILARITY INDICES
Similarity indices measures alikeness of qualitative attributes among two data objects. If data objects are nearly alike then their similarity value is 1 else it is 0. It is in another way called as pattern matching and it reflects relationship strength between two data objects. Pattern matching may not suffer from curse of dimensionality like distance measure and objects are not scaled out. Following table 1 [3] represents formulae for commonly used similarity indices. The key

aspects of considered similarity measures are presented in mind map 2.

Table 1: Similarity Indices

Similarity Indices	Jaccard Similarity	Hamming Similarity	For Mixed Data Type
Formulae	$J(A,B)=\frac{\lvert A\cap B\rvert}{\lvert A\cup B\rvert}$	$HammingSim(s,t)=\frac{\sum_{i=1}^{n} IdSim(s[i],t[i])}{n}$	$s_{ij}=\frac{1}{d}\sum_{l=1}^{d} s_{ijl}$ $s_{ij}=\left(\sum_{l=1}^{d} w_{ijl}s_{ijl}\right)/\left(\sum_{l=1}^{d} w_{ijl}\right)$

The selection of distance measure can influence clustering result and shape of clusters. It is useful for evaluation of numeric data attributes. Even though the method is same, distance measurement to present individuals dissimilarity may yield different result, for e.g. Euclidean vs. Squared Euclidean in hierarchical clustering. Table 2 [3] shows common dissimilarity indices and their aspects are captured in mind map 2 [3].

Table 2: Dissimilarity Indices

Dissimilarity Indices	Minkowski Distance	Standardized Euclidean Distance	Cosine Distance	Pearson Correlation Distance	Mahalanobis Distance
Formulae	$\left(\sum_{i=1}^{d}\lvert x_{ij}-x_{jl}\rvert^{n}\right)^{1/n}$	$\left(\sum_{i=1}^{d}\left\lvert\frac{x_{ij}-x_{jl}}{s_i}\right\rvert^{2}\right)^{1/2}$	$1-\cos\theta=\frac{x_i^T x_j}{\lVert x_i\rVert\lVert x_j\rVert}$	$1-\frac{Cov(x_i,x_j)}{\sqrt{D(x_i)}\sqrt{D(x_j)}}$	$\sqrt{(x_i-x_j)^T C^{-1}(x_i-x_j)}$

So, similarity indices consider union and intersection of data points within the cluster or identical similarity. On the other hand distance measures distance can be computed by using different available techniques n = 1 when it is City-block distance, Euclidean when n = 2 and Chebyshev when n = 3 in case of Minkowski distance.

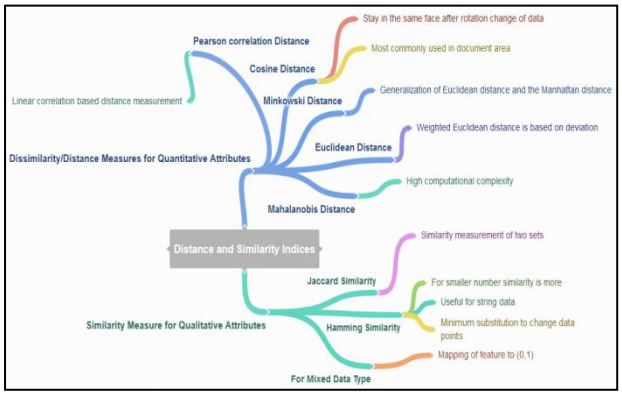

Fig. 2 Mind Map for Similarity and Dissimilarity Indices

4. EVALUATION INDICES

It is important to validate used clustering algorithm using evaluation indices based on internal or external data. It is difficult to gauge which clustering algorithm is better one when internal and external evaluation indices are different for them. The external evaluation is called as gold standard test method. Table 3 [4] and 4 [5] lists commonly used internal and external evaluation indices. Mind map 3 [4] [5] represents their key details.

Table 3: Internal Evaluation Indices

Internal Evaluation Indices	Davies-Bouldin index	Dunn index	Silhouette index
Formulae	$DB=\frac{1}{k}\sum_{i=1}^{k}\max_{i\neq j}\left(\frac{\sigma_i+\sigma_j}{d(c_i,c_j)}\right)$	$D=\min_{1\le i\le n}\left\{\min_{1\le j\le n, i\neq j}\left\{\frac{d(i,j)}{\max_{1\le k\le n} d'(k)}\right\}\right\}$	$s(i)=\frac{b(i)-a(i)}{\max\{a(i),b(i)\}}$

Table 4: External Evaluation Indices

External Evaluation Indices	Rand index	F index	Jaccard index	Fowlkes-Mallows index	Confusion matrix
Formulae	$RI=\frac{TP+TN}{TP+FP+FN+TN}$	$F_\beta=\frac{(\beta^2+1)\cdot P\cdot R}{\beta^2\cdot P+R}$	$J(A,B)=\frac{\lvert A\cap B\rvert}{\lvert A\cup B\rvert}=\frac{TP}{TP+FP+FN}$	$FM=\sqrt{\frac{TP}{TP+FP}\cdot\frac{TP}{TP+FN}}$	Matrix of FP, FN, TP and TN

Here, σ is the average distance between any data in cluster. $d(c_i,c_j)$ is the distance c_i and c_j. TP, TN, FP and FN are true positive, true negative, false positive and false negative respectively. a(i) is the average distance of i with all other data in the same cluster.

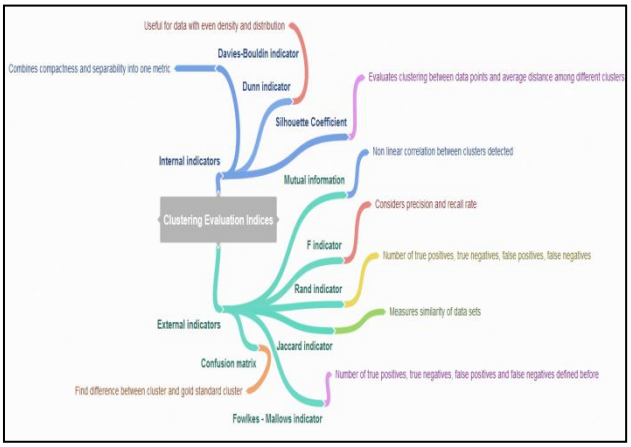

Fig. 3 Mind Map for Clustering Evaluation Indices

5. CONVENTIONAL CLUSTERING ALGORITHMS

Conventional clustering methods mostly deal with uncertain data. They usually create tight clusters. In most of the cases cluster numbers are required to defined before hand and usually suitable for convex dataset. They are fragile to outliers and noise. In this section first nine categories of conventional clustering algorithms are detailed out.

5.1 Partition Based

Popular algorithms are K-Means [6] and K-Medoids [7] which considers cluster center as center of data points. These methods shift data points between clusters per iteration. This shifting decreases criterion function for clustering until convergence. Changes to clustering criterion, makes clustering method insensitive to erroneous and missing data. Representative algorithms for partition method are Partition Around Medoids (PAM) [8], Clustering for Large Applications (CLARA) [9] and Clustering Large Applications based on RANdomized Search (CLARANS) [10]. K-Medoids, PAM and CLARA are moderately sensitive to sequence of inputting data.

5.2 Hierarchy Based

Hierarchical clustering or merging generates larger structures through continuous merging of smaller ones. Tree or dendogram is produced in top-down or bottom-up manner to show cluster hierarchies [11]. Three different strategies are supported by bottom-up hierarchical clustering viz., single-link, complete-link and average-link based on pair-wise distance between two clusters. Clustering algorithms of this type are Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) [12], Clustering Using Representatives (CURE) [13], and Robust Clustering Algorithm for Categorical Attributes (ROCK) [14] and Chameleon [15]. BIRCH through formation of cluster feature tree, CURE through random sampling to cluster sample, ROCK through similarity from data around the cluster and Chameleon through bottom up approach realizes clustering. They are moderately sensitive to sequence of inputting data.

5.3 Fuzzy Theory Based

Membership grades assignment to show degree to which data points belong to each cluster. Algorithms that belong to this family are Fuzzy C-Means (FCM) [16], Fuzzy C-Shells (FCS) [17] and Mountain Method (MM) [18]. FCM uses optimized object function for data point's membership. FCS uses hyper sphere based distance function to do clustering. MM devices cluster centres based on mountain function. All are moderately sensitive to sequence of inputting data.

5.4 Distribution Based

Clustering is based on distribution model. Though these clustering techniques are theoretically excellent but practically they suffer from over-fitting. They capture correlation and dependence between attributes. Algorithms of this category are Distribution Based Clustering of Large Spatial Databases (DBCLASD) [19] and Gaussian Mixture Model (GMM) [20]. DBCLASD considers nearest point belongs to cluster if they satisfy expected distance distribution generated from data of that cluster. GMM takes into account independent Gaussian distribution for cluster belongingness. DBCLASD is little and GMM is highly sensitive to sequence of inputting data.

5.5 Density Based

Reach and establish connection among dense data points is the working principle of density based clustering. Density-based spatial clustering of applications with noise (DBSCAN) [21], Ordering points to identify the clustering structure (OPTICS) [22] and Mean-Shift [23] are typical algorithms of this type. DBSCAN works on basic principle. OPTICS is insensitive to minimum points and radius of neighborhood. Mean-Shift iteratively calculates mean offset of current data points till convergence criteria is met.

5.6 Graph Based

Pair of elements is connected via nodes. The nodes can be further categorized as highly connected, important and unimportant nodes. CLuster Identification via Connectivity Kernels (CLICK) [24] and Minimum Spanning Tree (MST) [25] clustering are typical candidates of this type. CLICK considers iterative minimum weight division of graph for clustering. MST clustering as the name suggests generates graph of data in the form of MST for cluster analysis. Both are highly sensitive to sequence of inputting data.

5.7 Grid Based

Use of grid based data. Dependency on number of grids and not on data in data set. Ease in identification of neighbouring clusters. This clustering formulates multi-level granularity structure. STatistical INformation Grid (STING) [26] and CLustering In QUEst (CLIQUE) [27] are examples of grid based clustering. STING constructs hierarchical rectangular units to parallel cluster data at different levels. CLIQUE has grid density combo features and advantages. Both are little sensitive to sequence of inputting data.

5.8 Fractal Based

Fractal represents shape divided into different parts sharing common characters with the whole. Fractal Clustering (FC) [28] is the example. Change in data does not change fractal quality. FC has high sensitivity to sequence of inputting data.

5.9 Model Based

Optimal fit of data into model. Probability distribution based clusters. So, clustering method performs well when data conforms to model. COBWEB [29] generates tree for classification features by considering basic idea of model based clustering. Self Organizing Map (SOM) [30] establishes mapping between input and output by dimensionality reduction. Adaptive Response Theory (ART) [31] dynamically generates neurons to match pattern to cluster. COBWEB and ART are incremental clustering algorithms.

Mind maps 4A [7-31] and 4B [7-31] highlights key details about clustering algorithms discussed in this section under nine categories of conventional algorithms.

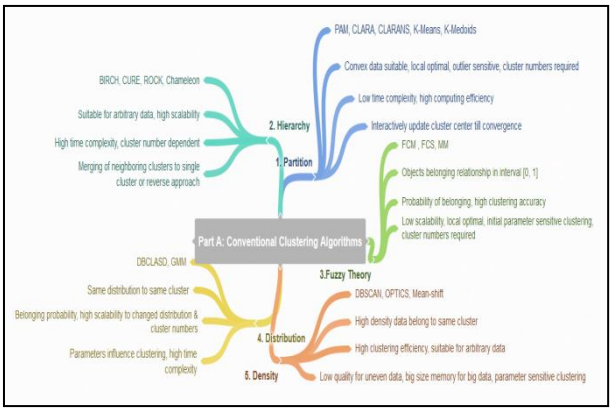

Fig. 4 Mind Map for Conventional Clustering Algorithms (Part A)

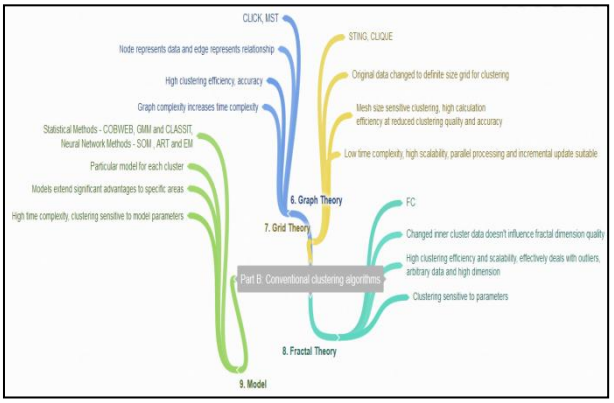

Fig. 4 Mind Map for Conventional Clustering Algorithms (Part B)

6. RECENT CLUSTERING ALGORITHMS

In this section survey of forty five clustering algorithms is presented. This section considers ten categories of recent clustering algorithms. These categories are in turn sub sections of this section.

6.1 Kernel Based

The core idea is nonlinear kernel function transform data into high dimensional feature space to carry out clustering. Kernel K-means [32], kernel SOM [33] and kernel FCM [34] works on this principle. Support Vector Clustering (SVC) [35] formulates minimum radius sphere based isoline including cluster data. Maximum Margin Clustering (MMC) [36] finds maximum hyper plane to cluster multi-label problem. Multiple Kernel Clustering (MKC) [37] finds best hyper plane based on multiple kernels to cluster. All these kernel based clustering algorithms are little sensitive to sequence of inputting data.

6.2 Ensemble Based

Nine categories of consensus functions viz. co-association, graph partition, relabeling, information theory, genetic algorithm, local adoption, kernel method and fuzzy theory for initial clustering. Final clustering result is summation of initial clustering results [38 - 40].

6.3 Swarm Intelligence Based

Random distribution of data on two dimensional grids, further selection of data is based on the simulated behavior of biological entities like ants [41], particle [42], frogs [43], bees [44] and the process is iterated till satisfactory results are achieved. Different aspects of biological entities are taken into account like speed, location, local search, global information interaction, duties performed by them etc. They are moderately sensitive to sequence of inputting data.

6.4 Quantum Theory Based

Distribution of data is based on potential energy. The object with minimum potential energy determined cluster center. Using defined distance function objects are put into clusters. Quantum clustering (QC) [45] and Dynamic quantum clustering (DQC) [46] are typical examples of this clustering. QC uses Schrodinger equation and DQC uses time-based Schrodinger equation with iterative gradient descent algorithm for getting potential energy of the object. Both are little sensitive to sequence of inputting data.

6.5 Spectral Graph Theory Based

The key idea here is object acts as a node and object similarity as weighted edge converting clustering problem into graph partition. SM [47] and NJW [48] make use of Eigen vector for clustering. SM uses minimized heuristic normalized cut for image segmentation and NJW considers K largest values of Laplacian matrix. Both are little sensitive to sequence of inputting data.

6.6 Affinity Propagation Based

Affinity propagation takes input as pair wise similarities among data points and clusters formed by maximizing total similarity between data points and their best representing data points as exemplar. The affinity sum of data point for other data points is higher; their probability to become cluster centre is higher. This message passing iterations considers two steps viz. a) responsibilities – original similarities and b) availabilities – calculated in previous iterations. The stopping criterion is when changes in values are below threshold or maximum iterations are done. Affinity Propagation (AP) [49] clustering is typical clustering of this kind. It is moderately sensitive to sequence of inputting data.

6.7 Distance and Density Based

Density of local and other points is calculated. Decision graph shows densities of local points and higher densities of other points. Cluster centers are determined from decision graph. Remaining points are kept in nearby clusters having high density local points at last. Crucial aspect is performance is subjective to decision graph. Density Distance (DD) [50] clustering is this type of clustering. DD is little sensitive to sequence of inputting data.

6.8 Spatial Data Based

Data has two dimensions time and space. Parallel processing for clustering of new data by applying transform on original data like Wavelet transform in Wavelet clustering [51]. CLARANS [10] improvement to CLARA [9] uses PAM [8] for clustering. DBSCAN [21] and STING [26] are other examples of spatial data clustering and they are discussed in earlier section of the paper.

6.9 Data Stream Based

STREAM [52] finds hierarchical clustering using divide and conquer successively on arriving data sequence. CluStream [53] deals with dynamic data to form micro-clusters online and offline as a timely response. HPStream [54] considers data attenuation over time and can handle high dimensions. DenStream [55] considers density of non convex data, effectual towards outliers.

6.10 Large Scale Data Based

Data which is large in dimension, rich in diversity, high drift and distinguishable shares characteristics in large data clustering. K-means [6], BIRCH [12], CLARA [9], CURE [13], DBSCAN [21], DENCLUE [56], Wavecluster [51] and FC [28] are clustering algorithms preferred for processing of large scale data.

Mind maps 5A, 5B and 5C represents gist about recent clustering algorithms covered under above discussed ten categories.

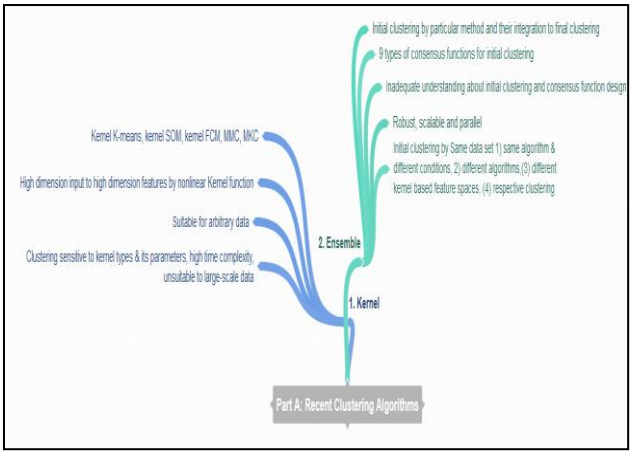

Fig. 5A Mind Map for Recent Clustering Algorithms (Part A)

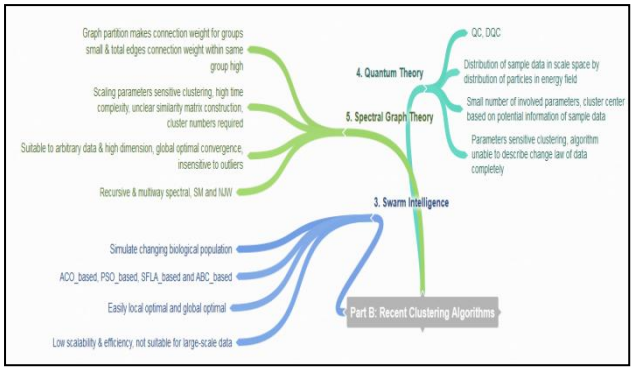

Fig. 5B Mind Map for Recent Clustering Algorithms (Part B)

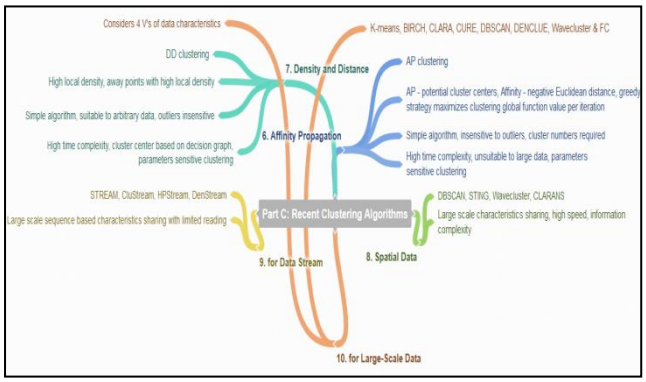

Fig. 5C Mind Map for Recent Clustering Algorithms (Part C)

7. PARALLEL AND DISTRIBUTED DATA CLUSTERING ALGORITHMS

As discussed in section 6.10, here survey of parallel and distributed clustering algorithms for large scale data is presented. Algorithms considered for survey here are K-Means, BIRCH, DBSCAN, CLARA, CURE and Wavecluster. Mind Maps in figure 6A and 6B present survey of twenty two papers which are implemented either parallel or distributed. These two mind maps present key aspects from implementation, performance and other critical issues for considered parallel or distributed clustering algorithms. Citations to referred papers are provided in mind maps itself.

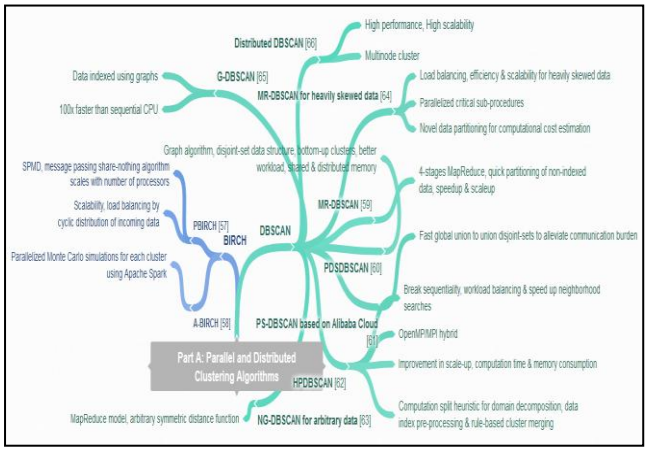

Fig. 6A Parallel and Distributed Clustering Algorithms(Part A)

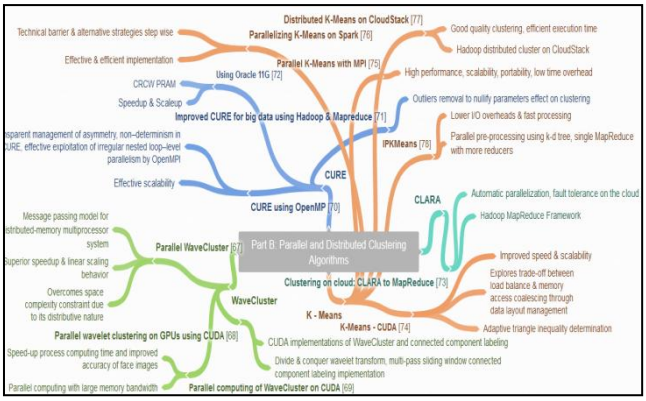

Fig. 6B Parallel and Distributed Clustering Algorithms(Part B)

8. LEARNING'S FROM MIND MAPS FOR PARALLEL AND DISTRIBUTED CLUSTERING ALGORITHMS

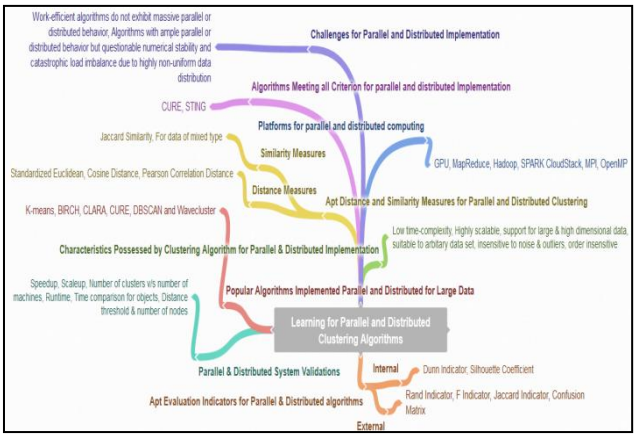

Fig. 7 Learning for Parallel and Distributed Clustering Algorithms

9. OUTLOOK

As discussed earlier in this paper, this paper uses Mind Maps to give idea about key points of each clustering algorithm whether it will be from conventional or recent category. It is difficult to survey all available clustering algorithms but these paper surveys 24 conventional, 20 recent and 22 parallel, distributed clustering algorithms. Section 7 presents survey of 22 parallel and distributed algorithms through mind maps in fig 6A and 6B. Each main branch is the particular clustering algorithm, its sub-branches shows recent parallel and distributed clustering algorithms related to it. Fig. 7 shows learning's from each mind map and for development of parallel and distributed algorithm. Table 5 shows comparative details about considered parallel and distributed clustering algorithms. From this table, it is clear that only algorithms are suitable for parallel and distributed implementation. Table 5 considers 1 to 10 parameters for comparison. They are 1) Category 2) Algorithm 3) Time Complexity 4) Scalability 5) Large Scale data handling 6) High Dimension 7) Data Shape 8) Order Insensitive 9) Noise/outlier sensitivity and 10) references.

Table 5: Comparative Analysis of Algorithms Suitable for Large Scale Data

1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**8**	**10**
Partition Based	K-Means	Low	Middle	Yes	No	Convex	High	High	[6]
Hierarchy Based	BIRCH	Low	High	Yes	No	Convex	Moderate	Little	[12]
Density Based	DBSCAN	Middle	Middle	Yes	No	Arbitrary	Moderate	Little	[21]
Partition Based	CLARA	Middle	High	Yes	No	Convex	Moderate	Little	[9]
Hierarchy Based	CURE	Low	High	Yes	Yes	Arbitrary	Moderate	Little	[13]
Grid Based	STING	Low	High	Yes	Yes	Arbitrary	Littl e	Little	[26]

From this table, it is clear CURE and STING are best suited for clustering of large scale data.

10. REFERENCES

[1] Jain A, Dubes R (1988) Algorithms for clustering data. Prentice-Hall, Inc, Upper Saddle River.

[2] Shinde, K., & Mulay, P. (2017, April). Cbica: Correlation based incremental clustering algorithm, a new approach. In *Convergence in Technology (I2CT), 2017 2nd International Conference for* (pp. 291-296). IEEE.

[3] Xu R, Wunsch D (2005) Survey of clustering algorithms. IEEE Trans Neural Netw 16:645–678.

[4] Estivill-Castro V (2002) Why so many clustering algorithms: a position paper. ACMSIGKDD Explor Newsl 4:65–75.

[5] Färber I, Günnemann S, Kriegel H, Kröger P, Müller E, Schubert E, Seidl T, Zimek A (2010) On using class-labels in evaluation of clusterings. In MultiClust: 1st international workshop on discovering, summarizing and using multiple clusterings held in conjunction with KDD, Washington, DC.

[6] MacQueen J (1967) Some methods for classification and analysis of multivariate observations. Proc Fifth Berkeley Symp Math Stat Probab 1:281–297.

[7] Park H, Jun C (2009) A simple and fast algorithm for K-medoids clustering. Expert Syst Appl 36:3336–3341.

[8] Kaufman L, Rousseeuw P (1990) Partitioning around medoids (program pam). Finding groups in data: an introduction to cluster analysis. Wiley, Hoboken.

[9] Kaufman L, Rousseeuw P (2008) Finding groups in data: an introduction to cluster analysis, vol 344. Wiley, Hoboken. doi:10.1002/9780470316801.

[10] Ng R, Han J (2002) Clarans: a method for clustering objects for spatial data mining. IEEE Trans Knowl Data Eng 14:1003–1016.

[11] Johnson S (1967) Hierarchical clustering schemes. Psychometrika 32:241–254.

[12] Zhang T, Ramakrishnan R, Livny M(1996) BIRCH: an efficient data clustering method for very large databases. ACM SIGMOD Rec 25:103–104.

[13] Guha S, Rastogi R, Shim K (1998) CURE: an efficient clustering algorithm for large databases. ACM SIGMOD Rec 27:73–84

[14] Guha S, Rastogi R, Shim K (1999) ROCK: a robust clustering algorithm for categorical attributes. In: Proceedings of the 15th international conference on data engineering, pp 512-521.

[15] Karypis G, Han E, Kumar V (1999) Chameleon: hierarchical clustering using dynamic modeling. Computer 32:68–75.

[16] Bezdek J, Ehrlich R, Full W (1984) FCM: the fuzzy c-means clustering algorithm. Comput Geosci 10:191–203.

[17] Dave R, Bhaswan K (1992) Adaptive fuzzy c-shells clustering and detection of ellipses. IEEE Trans Neural Netw 3:643–662.

[18] Yager R, Filev D (1994) Approximate clustering via the mountain method. IEEE Trans Syst Man Cybern 24:1279–1284.

[19] Xu X, Ester M, Kriegel H, Sander J (1998) A distribution-based clustering algorithm for mining in large spatial databases. In: Proceedings of the fourteenth international conference on data engineering, pp 324-331.

[20] Rasmussen C (1999) The infinite Gaussian mixture model. Adv Neural Inf Process Syst 12:554–560.

[21] Ester M, Kriegel H, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the second ACM SIGKDD international conference on knowledge discovery and data mining, pp 226–231.

[22] Ankerst M, Breunig M, Kriegel H, Sander J (1999) OPTICS: ordering points to identify the clustering structure. In: Proceedings on 1999 ACMSIGMOD international conference on management of data, vol 28, pp 49–60

[23] Comaniciu D, Meer P (2002) Mean shift: a robust approach toward feature space analysis. IEEE Trans Pattern Anal Mach Intell 24:603–619.

[24] Sharan R, Shamir R (2000) CLICK: a clustering algorithm with applications to gene expression analysis. In: Proc international conference intelligent systems molecular biolgy, pp 307–316.

[25] Jain A, Murty M, Flynn P (1999) Data clustering: a review. ACMComput Surv (CSUR) 31:264–323.

[26] Wang W, Yang J, Muntz R (1997) STING: a statistical information grid approach to spatial data mining. In VLDB, pp 186–195.

[27] Agrawal R, Gehrke J, Gunopulos D, Raghavan P (1998) Automatic subspace clustering of high dimensional data for data mining applications. In: Proceedings 1998 ACM sigmod international conference on management of data, vol 27, pp 94–105.

[28] Barbará D, Chen P (2000) Using the fractal dimension to cluster datasets. In: Proceedings of the sixth ACM SIGKDD international conference on knowledge discovery and data mining, pp 260–264.

[29] Fisher D(1987) Knowledge acquisition via incremental conceptual clustering. Mach Learn 2:139–172.

[30] KohonenKohonen T (1990) The self-organizing map. Proc IEEE 78:1464–1480.

[31] Carpenter G, Grossberg S (1988) The ART of adaptive pattern recognition by a self-organizing neural network. Computer 21:77–88.

[32] Schölkopf B, Smola A, Müller K (1998) Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput 10:1299–1319.

[33] MacDonald D, Fyfe C (2000) The kernel self-organising map. Proc Fourth Int Conf Knowl-Based Intell Eng Syst Allied Technol 1:317–320.

[34] Wu Z, Xie W,Yu J (2003) Fuzzy c-means clustering algorithm based on kernel method. In: Proceedings of the fifth ICCIMA, pp 49–54.

[35] Ben-Hur A, Horn D, Siegelmann H, Vapnik V (2002) Support vector clustering. J Mach Learn Res 2:125–137.

[36] Xu L, Neufeld J, Larson B, Schuurmans D (2004) Maximum margin clustering. In: Advances in neural information processing systems, pp 1537–1544.

[37] Zhao B, Kwok J, Zhang C (2009) Multiple kernel clustering. In SDM, pp 638–649.

[38] Strehl A, Ghosh J (2003) Cluster ensembles—a knowledge reuse framework for combining multiple partitions. J Mach Learn Res 3:583–617.

[39] Topchy A, Jain A, Punch W (2004) A mixture model for clustering ensembles. In: Proceedings of the SIAM international conference on data mining, pp 379.

[40] Topchy A, Jain A, Punch W (2005) Clustering ensembles: models of consensus and weak partitions. IEEE Trans Pattern Anal Mach Intell 27:1866–1881.

[41] Handl J, Meyer B (2007) Ant-based and swarm-based clustering. Swarm Intell 1:95–113.

[42] Van der Merwe D, Engelbrecht A (2003) Data clustering using particle swarm optimization. Congr Evol Comput 1:215–220.

[43] Amiri B, Fathian M, Maroosi A (2009) Application of shuffled frog-leaping algorithm on clustering. Int J Adv Manuf Technol 45:199–209.

[44] Karaboga D, Ozturk C (2011) A novel clustering approach: artificial bee colony (ABC) algorithm. Appl Soft Comput 11:652–657.

[45] Horn D, Gottlieb A (2001) The method of quantum clustering. In: Advances in neural information processing systems, pp 769–776.

[46] Weinstein M, Horn D (2009) Dynamic quantum clustering: a method for visual exploration of structures in data. Phys Rev E 80:066117.

[47] Shi J, Malik J (2000) Normalized cuts and image segmentation. IEEE Trans Pattern Anal Mach Intell 22:888–905.

[48] Ng A, Jordan M, Weiss Y (2002) On spectral clustering: analysis and an algorithm. Adv Neural Inf Process Syst 2:849–856.

[49] Frey BJ, Dueck D (2007) Clustering by passing messages between data points. Science 315(5814):972–976.

[50] Rodriguez A, Laio A (2014) Clustering by fast search and find of density peaks. Science 344:1492–1496.

[51] Sheikholeslami G, Chatterjee S, Zhang A(1998) Wavecluster: Amulti-resolution clustering approach for very large spatial databases. In: VLDB, pp 428–439.

[52] O'callaghan L, Meyerson A, Motwani R, Mishra N, Guha S (2002) Streaming-data algorithms for high-quality clustering. In: ICDE, p 0685.

[53] Aggarwal C, Han J, Wang J, Yu P (2003) A framework for clustering evolving data streams. In: VLDB, pp 81–92.

[54] Aggarwal C, Han J, Wang J, Yu P (2004) A framework for projected clustering of high dimensional data streams. In: VLDB, pp 852–863.

[55] Cao F, Ester M, Qian W, Zhou A (2006) Density-based clustering over an evolving data stream with noise. SDM 6:328–339.

[56] Hinneburg A, Keim D (1998) An efficient approach to clustering in large multimedia databases with noise. In Proceedings of the 4th ACM SIGKDD international conference on knowledge discovery and data mining 98: 58–65.

[57] Garg, A., Mangla, A., Gupta, N., & Bhatnagar, V. (2006, December). PBIRCH: A scalable parallel clustering algorithm for incremental data. In *Database Engineering and Applications Symposium, 2006. IDEAS'06. 10th International* (pp. 315-316). IEEE.

[58] Lorbeer, B., Kosareva, A., Deva, B., Softić, D., Ruppel, P., & Küpper, A. (2017). Variations on the Clustering Algorithm BIRCH. *Big Data Research*.

[59] He, Y., Tan, H., Luo, W., Mao, H., Ma, D., Feng, S., & Fan, J. (2011, December). Mr-dbscan: an efficient parallel density-based clustering algorithm using mapreduce. In Parallel and Distributed Systems (ICPADS), 2011 IEEE 17th International Conference on (pp. 473-480). IEEE.

[60] Patwary, M. A., Palsetia, D., Agrawal, A., Liao, W. K., Manne, F., & Choudhary, A. (2012, November). A new scalable parallel DBSCAN algorithm using the disjoint-set data structure. In Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis (p. 62). IEEE Computer Society Press.

[61] Hu, X., Huang, J., & Qiu, M. (2017, November). A Communication Efficient Parallel DBSCAN Algorithm based on Parameter Server. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 2107-2110). ACM.

[62] Götz, M., Bodenstein, C., & Riedel, M. (2015, November). HPDBSCAN: highly parallel DBSCAN. In *Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments* (p. 2). ACM.

[63] Lulli, A., Dell'Amico, M., Michiardi, P., & Ricci, L. (2016). NG-DBSCAN: scalable density-based clustering for arbitrary data. *Proceedings of the VLDB Endowment*, *10*(3), 157-168.

[64] He, Y., Tan, H., Luo, W., Feng, S., & Fan, J. (2014). MR-DBSCAN: a scalable MapReduce-based DBSCAN algorithm for heavily skewed data. *Frontiers of Computer Science*, *8*(1), 83-99.

[65] Andrade, G., Ramos, G., Madeira, D., Sachetto, R., Ferreira, R., & Rocha, L. (2013). G-dbscan: A gpu accelerated algorithm for density-based clustering. *Procedia Computer Science*, *18*, 369-378.

[66] Merk, A., Cal, P., & Woźniak, M. (2017, May). Distributed DBSCAN Algorithm–Concept and Experimental Evaluation. In International Conference on Computer Recognition Systems (pp. 472-480). Springer, Cham.

[67] Yıldırım, A. A., & Özdoğan, C. (2011). Parallel WaveCluster: A linear scaling parallel clustering algorithm implementation with application to very large datasets. Journal of Parallel and Distributed Computing, 71(7), 955-962.

[68] Yıldırım, A. A., & Özdoğan, C. (2011). Parallel wavelet-based clustering algorithm on GPUs using CUDA. Procedia Computer Science, 3, 396-400.

[69] Anggraini, E. L., Suciati, N., & Suadi, W. (2013, June). Parallel computing of WaveCluster algorithm for face recognition application. In QiR (Quality in Research), 2013 International Conference on (pp. 56-59). IEEE.

[70] Hadjidoukas, P. E., & Amsaleg, L. (2008). Parallelization of a hierarchical data clustering algorithm using openmp. In OpenMP Shared Memory Parallel Programming (pp. 289-299). Springer, Berlin, Heidelberg.

[71] Lathiya, P., & Rani, R. (2016, August). Improved CURE clustering for big data using Hadoop and Mapreduce. In *Inventive Computation Technologies (ICICT), International Conference on* (Vol. 3, pp. 1-5). IEEE.

[72] Maitrey, S., Jha, C. K., Gupta, R., & Singh, J. (2012). Enhancement of CURE clustering technique in data mining. International Journal of Computer Applications.

[73] Jakovits, P., & Srirama, S. N. (2013, September). Clustering on the cloud: Reducing clara to mapreduce. In *Proceedings of the Second Nordic Symposium on Cloud Computing & Internet Technologies* (pp. 64-71). ACM.

[74] Wu, J., & Hong, B. (2011, May). An efficient k-means algorithm on CUDA. In Parallel and Distributed Processing Workshops and Phd Forum (IPDPSW), 2011 IEEE International Symposium on (pp. 1740-1749). IEEE.

[75] Zhang, J., Wu, G., Hu, X., Li, S., & Hao, S. (2011, December). A parallel k-means clustering algorithm with mpi. In *Parallel Architectures, Algorithms and Programming (PAAP), 2011 Fourth International Symposium on* (pp. 60-64). IEEE.

[76] Wang, B., Yin, J., Hua, Q., Wu, Z., & Cao, J. (2016, August). Parallelizing k-means-based clustering on spark. In Advanced Cloud and Big Data (CBD), 2016 International Conference on (pp. 31-36). IEEE.

[77] Mao, Y., Xu, Z., Li, X., & Ping, P. (2015, August). An optimal distributed K-Means clustering algorithm based on cloudstack. In *Information and Automation, 2015 IEEE International Conference on* (pp. 3149-3156). IEEE.

[78] Jin, S., Cui, Y., & Yu, C. (2016). A New Parallelization Method for K-means. *arXiv preprint arXiv:1608.06347*.

[79] https://coggle.it/