World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:9, No:10, 2015

Hierarchical Clustering Algorithms in Data Mining

Z. Abdullah, A. R. Hamdan

***Abstract*—**Clustering is a process of grouping objects and data into groups of clusters to ensure that data objects from the same cluster are identical to each other. Clustering algorithms in one of the area in data mining and it can be classified into partition, hierarchical, density based and grid based. Therefore, in this paper we do survey and review four major hierarchical clustering algorithms called CURE, ROCK, CHAMELEON and BIRCH. The obtained state of the art of these algorithms will help in eliminating the current problems as well as deriving more robust and scalable algorithms for clustering.

***Keywords*—**Clustering, method, algorithm, hierarchical, survey.

I. Introduction

DATA mining is a method of mining and extracting useful information from large data repositories. It involves with the process of analyzing data and finds some valuable information. There are several methods in data mining such as classification, clustering, regression, association, and sequential pattern matching [1]. Clustering basically tries to assemble the set of data items into clusters of the similar identity. Clustering is an example of unsupervised learning because there are no predefined classes. The quality of the cluster can be measure by high intra-cluster similarity and low inter-cluster similarity.

Nowadays, clustering becomes one of the important topics and has been applied in various fields like biology, psychology and statistic [2]. There are many types of clustering and the most influence ones can be divided into partitioning, hierarchical, density-based, grid-based and model-based [3]. Partitioning method classifies the data based on the distance between objects. Hierarchical method creates a hierarchical decomposition of the given set of data objects. Density-based methods categorized the data based on density or based on an explicitly constructed density function. Grid-based methods organize the object space in a form of grid structure. Model-based methods arranged the object that the best fit of the given model.

In a hierarchal method, separate clusters are finally joined into one cluster. The density of the data points is employed to determine the relevant clusters. The main advantage is it uses less computation costs in term of combinatorial number of data points. However, it is very rigid and unable to reverse back once it performed the merging or splitting process. As a result, any decision that prior the earlier mistakes are not able to be rectified.

Z. Abdullah is with the Universiti Malaysia Terengganu, 21030, Kuala Terengganu, Malaysia (corresponding author; phone: 603-6683536; fax: 603-669-4100; e-mail: zailania@umt.edu.my).

A.R. Hamdan is with the Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia (e-mail: arh@ftsm.ukm.my).

Generally, hierarchical clustering algorithms can be divided into two categories: Divisive and Agglomerative. Agglomerative clustering performs the bottom-up strategy, in which it initially considers each data point as a singleton cluster. After that, it continues by merging all those clusters until all points are combined into a single cluster. A dendogram or tree graph is used to represent the output. Then the algorithm splits back the single cluster in gradually manner until the required number of clusters is obtained. To be more specific, two major steps are involved. First is to choose a suitable number of clusters to split. Second is to determine the best approach on how to split the selected clusters into two new clusters [4]. In hierarchical clustering algorithms, many algorithms have been proposed and the widely studied are ROCK [2], BIRCH [5], CURE [6], and CHAMELEON [7].

In this paper, we review four (4) major algorithms in hierarchical clustering called ROCK, BIRCH, CURE and CHEMELEON. The review was carried out against related articles from the year 1998 until 2015.

The remaining part of the paper is organized as follows. Sections II presents the related works for hierarchical clustering algorithms. Section III reviews several prominent hierarchical clustering algorithms. Section IV highlights the some developments of the algorithm. Section V emphasizes on a few major issues that related to the algorithm. Section VI describes the challenges and limitations of the algorithm. Section VII includes a number of suggestions to improve the algorithm. Finally, Section VIII gives conclusion of reviewing the algorithm based on the survey.

II. Related Works

The main methods of data mining involve with classification and prediction, clustering, sequence analysis, outlier detection, association rules time series analysis and text mining. Among these methods, clustering is considered as among the widely and intensively studied by many data mining researchers. Richard and Hard [8] elaborated the unsupervised learning and clustering in pattern recognition. Ng and Han [9] discussed on partitioned or centroid-based hierarchical clustering algorithm by partitioning first the database and then iteratively optimizing an objective function. The limitation is that it is not suitable for categorical attributes [2].

Zhao and Karypis [10] suggested improvement of clustering algorithms and demonstrated both partition and agglomerative algorithms that use different criterion functions and merging schemes. On top of that, a new class of clustering algorithms called constrained agglomerative algorithms is proposed by combining features from both algorithms. The algorithm reduces the errors of classical agglomerative algorithms and thus improved the overall quality of hierarchical clustering algorithms.

Salvador and Chan [11] researched on determining the right number of clusters when using hierarchical clustering algorithms. L method that finds the "knee" in a number of clusters against clustering evaluation metric' graph is proposed. The challenge is most of the major clustering algorithms need to re-run many times in order to find the best potential number of clusters. As a result, it is very time consuming and quality of obtained clusters is still unknown and questionable.

Koga et al. [12] introduced fast agglomerative hierarchical clustering algorithm using Locality-Sensitive Hashing. The main advantage is that the time complexity is getting reduced by O(nB), where B is practically a constant factor and n represents the quantity of information points. However, it only relies on vector data and limited to a single linkage. Moreover, it is also not practical for a large knowledge.

Murthy et al. [13] investigated on content based image retrieval using hierarchical and k-Means clustering algorithms. In this algorithm, the images are filtered and then applied with k-Means, to get a high quality of image results. After determining the cluster centroid, the given query images go to the respective clusters centers. The clusters are graded based to their resemblance with the query. The advantage is the algorithm produce more tightly clusters than classical hierarchical clustering. The disadvantages are that it is very difficult to determine k-values and didn't work well with global cluster.

Hong et al. [14] associates SVM-based intrusion detection system with a hierarchical clustering algorithm. For this integration, all non-continuous attributes are converted into continuous attributes. On top of that, the entire datasets are balanced to ensure all feature values can have their own interval. Even though this approach reduces the training time, it requires several key parameters that need to be set correctly to achieve the best clustering results.

Balcan et al. [15] introduced a robust hierarchical clustering algorithm to examine a new robust algorithm for bottom-up agglomerative clustering. The algorithm is quite simple, quicker, and mostly valid in returning the clusters. In addition, the algorithm precisely clusters the data according to their natural characteristics in which the traditional agglomerative algorithms fail to do so.

Szilágyi and Szilágyi [16] studied on fast hierarchical clustering algorithms for large-scale protein sequence data sets. An altered sparse matrix structure is presented to overcome the most processes at the main loop. A fast matrix squaring formula is introduced to speed up the process. The proposed solution improves performance by two orders of magnitude against protein sequence databases.

Ng and Han [17] suggested CLARANS and used a medoids to represent cluster. Medoids represent objects of a data set or a cluster with a data set whose average dissimilarity to all the objects in the cluster is very minimal. The algorithm draws sample of neighbor dynamically in which that no nodes with corresponding to particular objects are completely eliminated. In addition, it requires a small number of searches and higher quality of clustering. CLARANS suffers from some disadvantage as it has issues with I/O efficiency. It also could not find a local minimum because of searching is controlled by maximum neighbor.

Huang [18] proposed K-prototypes algorithm based on K-means algorithm but it eliminates numeric data restrictions while conserving its effectiveness. The algorithm works similar to the K-means algorithm by clustering objects with numeric and categorical attributes. Square Euclidean distance is employed as the similarity measure on numeric attributes. The number of divergences between objects and the cluster samples is the similarity measure on the categorical attributes. The limitation of Means algorithm is that it is not suitable for categorical attributes due to constraints of similarity measure. A better clustering algorithm known as ROCK by [2] was proposed to handle the drawback of traditional clustering algorithms that uses distance measure to cluster data.

III. Hierarchical Clustering Algorithms

Hierarchical clustering is a method of cluster analysis to present clusters in hierarchy manner. Most of the typical methods are not able to make clusters rearrangement or adjustment after merging or splitting process. As a result, if the merging processes of objects have problems, it might produce the low quality of clusters. One of the solutions is by integrating the cluster with multiple clusters using a few alternative methods.

A. *Clustering Using Representatives Algorithm*

Guha et al. [6] proposed Clustering Using Representatives (CURE) algorithm that utilizes multiple representative points for each cluster. CURE is a kind of class-conscious bunch algorithmic rule that requires dataset to be partitioned. A mixture of sampling and partitioning is applied as a strategy to deal with vast information. A random sample from the dataset is partitioned to be part of the clusters. CURE first partitions the random sample and then partially clusters the data points according to the partition. After removing all outliers, the pre clustered data in each partition is then clustered again to produce the final clusters. The clustering algorithm can recognize arbitrarily shaped clusters. The algorithm is robust to the detect the outliers, and the algorithm uses space that is linear in the input size n and has a worst-case time complexity of $O(n^2 \log n)$. The clusters produced by CURE are also better than the other algorithms [6]. Fig. 1 presents the overview of the CURE algorithms in graphical manner.

B. *Robust Clustering Using Links Algorithm*

Guha et al. [2] suggested hierarchical agglomerative clustering algorithm called Robust Clustering Using Links (ROCK). It uses the concept of links to clusters data points and Jaccard coefficient [19] measure to obtain the similarity among the data. Boolean and categorical are two types of attributes that are most suited in this algorithm. The similarity of the objects in the respective clusters is determined by the number of points from different clusters that have the common neighbors.

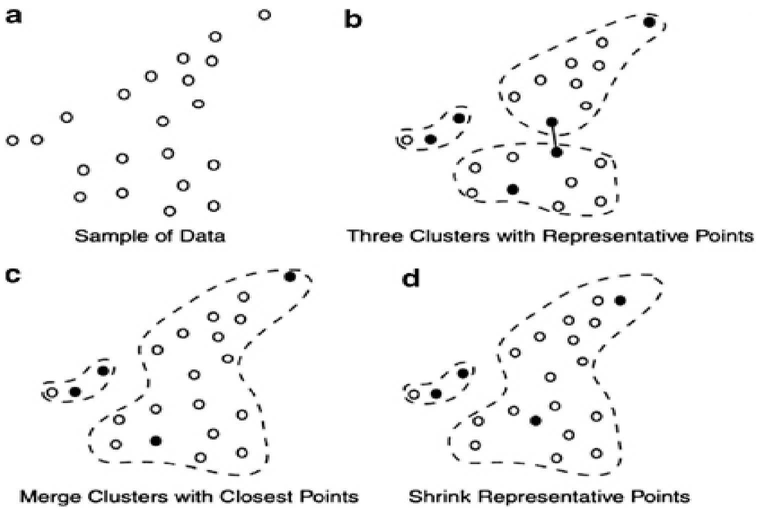

Fig. 1 Overview of CURE

The steps involved in ROCK algorithm are drawing random sample from datasets, cluster random links and label data on disk. After drawing random sample from the database, the links are applied into the sample points. Finally, only the sampled points are used to assign the remaining data points to the appropriate clusters. The process of merging the single clusters is continuous until it reaches the threshold of desired clusters or until there is number of common links between clusters becomes zero. ROCK is not only generates better quality cluster than traditional algorithms, but it also exhibits the good scalability property. Fig. 2 presents the overview of the ROCK algorithm in graphical manner.

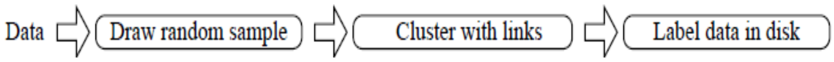

Fig. 2 Overview of ROCK

C. CHAMELEON Algorithm

Karypis et al. [7] introduced CHAMELEON that considering dynamic model of the clusters. CHAMELON discovers natural clusters of different shapes and sizes by dynamically adapting the merging decision based on different clustering model characteristics. Two phases are involved. First, partition the data points into sub-clusters using a graph partitioning. Second, repeatedly merging these sub-clusters until its find the valid clusters.

The key feature in CHAMELEON is that it determines the pair of the most similar sub-clusters by two considering relative inter-connectivity and the relative closeness of the clusters. The relative interconnectivity between pair of clusters is the absolute inter-connectivity between two normalized clusters with respect to the internal inter-connectivity of them. The relative closeness between pair of clusters is the absolute clones between two clusters normalized with respect to the internal closeness of them.

D. Balanced Iterative Reducing and Clustering Using Hierarchies Algorithm

Zhang et al. [20] proposed a collective hierarchal clustering algorithm called Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH). It is designed to minimize the quantity of I/O operations. BIRCH clusters incoming multi-dimensional metric information points in incrementally and dynamically manner. The clusters are formed by single scanning of the data and their quality will be improved by multiple scanning. The noise in the database is also taken into consideration by the algorithm.

Four phases are involved in producing the refined clusters. First, scan all data and build an initial in-memory Clustering Features (CF) Tree. CF Tree represents the clustering information of dataset within a memory limitation. Second, is an optional phase and only applicable if relevant. The data is compress into desirable range by building a smaller CF Tree. All outliers are removed at this phase. Third, perform the global or semi-global algorithm as remedy to cluster all leaf in CF Tree. Fourth, is an optional phase to correct the inaccuracy and further refinement of the clusters. It can be used to disregard the outliers.

IV. Development of Hierarchical Clustering Algorithms

TABLE I
VARIOUS EXAMPLES OF HIERARCHICAL CLUSTERING ALGORITHMS ARTICLES WITH AUTHORS

Author	Article Title
Zhao and Karypis (2002) [10]	Evaluation of hierarchical clustering algorithms for document datasets
Salvador and Chan (2004)[11]	Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms.
Laan and Pollard. (2003) [26]	A new algorithm for hybrid hierarchical clustering with visualization and the bootstrap.
Zhao et al. (2005) [27]	Hierarchical clustering algorithms for document datasets.
Mingoti and Lima (2006) [28]	Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms.
Shepitsen et al. (2008) [29]	Personalized recommendation in social tagging systems using hierarchical clustering.
Koga et al.(2007) [30]	Fast agglomerative hierarchical clustering algorithm using Locality-Sensitive Hashing.
Abbas. (2008) [31]	Comparisons Between Data Clustering Algorithms
Xin et al. (2008) [32]	EEHCA: An energy-efficient hierarchical clustering algorithm for wireless sensor networks
Jain (2010) [33]	Data clustering: 50 years beyond K-means.
Murthy et al. (2010) [34]	Content based image retrieval using Hierarchical and K-means clustering techniques.
Cai and Sun (2011) [35]	ESPRIT-Tree: hierarchical clustering analysis of millions of 16S rRNA pyrosequences in quasilinear computational time.
Horng et al. (2011) [36]	A novel intrusion detection system based on hierarchical clustering and support vector machines.
Kou and Lou (2012) [37]	Multiple factor hierarchical clustering algorithm for large scale web page and search engine clickstream data.
Krishnamurthy et al. (2012) [38]	Efficient active algorithms for hierarchical clustering.
Langfelder and Horvath (2012) [39]	Fast R functions for robust correlations and hierarchical clustering
Malitsky et al. (2013) [40]	Algorithm portfolios based on cost-sensitive hierarchical clustering.
Meila and Heckerman, (2013) [41]	An experimental comparison of several clustering and initialization methods.
Müllner (2013) [42]	Fast cluster: Fast hierarchical, agglomerative clustering routines for R and Python.
Balcan et al. (2014) [43]	Robust hierarchical clustering
Murtagh and Legendre. (2014) [44]	Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?
Szilágyi and Szilágyi (2014) [45]	A fast hierarchical clustering algorithm for large-scale protein sequence data sets.
Rashedi et al. (2015) [46]	An information theoretic approach to hierarchical clustering combination.
Ding et al. (2015) [47]	Sparse hierarchical clustering for VHR image change detection.

TABLE II
MAJOR HIERARCHICAL CLUSTERING ALGORITHMS FOCUSED IN THIS PAPER WITH AUTHORS

Author	Article Title
BIRCH	Zhang et al. (1999) [5]
ROCK	Guha et al. (1999) [2]
CURE	Shim et al. (1999) [6]
CHAMELEON	Karypis et al. (1999) [7]

TABLE III
COMPARISON OF VARIOUS HIERARCHICAL CLUSTERING ALGORITHMS

Algorithms	Large Datasets Suitability	Noise sensitivity
CURE	YES	Less
ROCK	YES	NO
CHAMELON	YES	NO
BIRCH	YES	Less

TABLE IV
COMPARISON OF ADVANTAGES AND DISADVANTAGES OF VARIOUS HIERARCHAL CLUSTERING ALGORITHMS

Clustering Algorithms	Advantages	Disadvantages
CURE	• Suitable to handle large data sets. • Outliers can be detected easily.	• Overlooks the information about the cumulative inter-connectivity of items in two clusters.
ROCK	• Suitable to handle large data sets. • Uses concept of links not distance for clustering thus improves quality of clusters of categorical data.	• The choice of a threshold function that is used to get a cluster quality is a difficult task for average users.
CHAMELON	• Fits for the applications where the size of the accessible data is big.	• Time complexity is high in dimension.
BIRCH	• Discover a good clustering with a single scan and increases the quality of clusters with further scans.	• Not applicable for categorical attributes where it handles only numerical data • Sensitive to the direction of the data record.

V. Current Issues

There are still many issues for hierarchical clustering algorithms and techniques. One of them is to search for the representatives of arbitrary shaped clusters. Mining arbitrary shaped clusters in large data sets is quite an open challenge [21]. Thus, there is no such well-established method to explain about the structure of arbitrary shaped clusters as defined by an algorithm. It is very crucial to find the appropriate representation of the clusters to describe their shape because clustering is a major mechanism for data reduction. As a result, explanation may effectively derive from the underlying data of the clustering results. In addition, hierarchical clustering algorithms need to be enhanced to be more scalable in dealing with the various shapes of clusters that stored in large datasets [22]. Another issue in incremental clustering, in which is the clusters in a dataset that may change due to insertion or update or deletion. Thus, it needs to reevaluate the clustering schemes that have been previously defined to cater for a dynamic dataset in a timely manner [23]. However, it is important to exploit the information hidden in the earlier clustering schemes so as to update them in an incremental way. Hierarchical clustering algorithms must be able to provide with similar efficiencies when dealing with a huge datasets.

Hierarchical clustering algorithms need to include the constraint-based clustering. Different application domains may consist of different clustering aspects to ensure their level of significant. Thus, some of the aspects might be stressed up and simply ignored which is relied on the requirements of the applications. In couple of years ago, Meng et al. [24] highlights that there is a trend so that cluster analysis is designed by providing less parameters but increasing more constraints. These constrains may potentially exist in data space or in users' queries. Therefore, clustering process must be able to consider these constraints and also define the inherent clusters that can fit a dataset.

VI. Challenges and Limitations

In many hierarchical clustering algorithms, once a decision is made to combine two clusters, it is impossible to reversed back [25]. As a result, the clustering process must be repeated several times in order to obtain the desired output. Hierarchical clustering algorithms also are very sensitive to noises and outliers like in CURE and BIRCH algorithms. Therefore, noises and outliers must be removed at early stages of clustering to ensure that the valid data points shouldn't fall into the wrong clusters. Another limitation is that it is difficult to deal with different sized clusters and convex shapes. At the moment, only CHAMELEON can produce the clusters in various shapes and sizes. The consequence is it may lead into the problems for building the final clusters quality where the shapes and cluster sizes is always the major concern. Besides that, breaking down the large clusters into smaller one also is against the principles of hierarchical clustering algorithms. Even though, some hierarchal algorithms can break the large clusters and merge back them, but the computational performance is still a major concern..

VII. Suggestions

There are some suggestions to be highlighted in order to improve the quality of the hierarchical clustering algorithms. Clustering process and the algorithms must work efficiently to derive for producing good quality of clusters. Determining a suitable algorithm for clustering the datasets that fit into the respective application is very important in ensuring a high quality of clusters. The algorithms should be able to produce random shapes of clusters rather than to some particular shapes as such as elliptical shapes as preferred by CHAMELEON algorithm. Due to the emerging of big data, the algorithms must be very robust in handling vast volume of data and high-dimensional structures with timely manner.

The algorithms should be incorporated with a feature that can accurately identify and finally eliminate all the possible outliers and noises as a strategy to reduce low quality of final clusters. The requirement of users-dependent parameters should be reduce because the users might uncertain in term of number of suitable clusters to be obtained and others things. Wrongly specifying the parameters might affect the overall computational performance as well as the quality of the clusters. Finally, algorithms should be more scalable in dealing with not only the categorical attributes, but also numerical and combination of both types of attributes.

VIII. Conclusion

Clustering is the process of grouping objects and data into groups of clusters to ensure that data objects from the same cluster are identical to each other. Generally, clustering can be divided into four categories and one of them is hierarchical. Hierarchical clustering is a method of cluster analysis aims at obtaining a hierarchy of clusters. Nowadays, it is still one of the most active research areas in data mining. In this paper, we do a survey on hierarchical clustering algorithms by highlighting in brief state of the art, current issues, challenge and limitations and some suggestions. It is expected that, the state of the art of hierarchical clustering algorithms will help the interested researchers to put forward in proposing more robust and scalable algorithms in the near future.

Acknowledgment

This work is supported by the research grant from Research Acceleration Center Excellence (RACE) of Universiti Kebangsaan Malaysia.

References

[1] M. Brown, "Data mining techniques" Retrieved from http://www.ibm.com/developerworks/library/ba-data-mining-techniques/

[2] S. Guha, R. Rastogi, and K. Shim, "ROCK: A robust clustering algorithm for categorical attributes" *Proceeding of 15th International Conference on Data Engineering* – ACM SIGKDD, pp. 512-521, 1999.

[3] M. Dutta, A.K. Mahanta, and A.K. Pujari, "QROCK: A quick version of the ROCK algorithm for clustering of categorical data," *Pattern Recognition Letters*, 26 (15), pp. 2364-2373, 2005.

[4] L. Feng, M-H. Qiu, Y-X. Wang, Q-L. Xiang and K. Liu, "A fast divisive clustering algorithm using an improved discrete particle swarm optimizer, *Pattern Recognition Letters*, 31, pp. 1216-1225, 2010

[5] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: An efficient data clustering method for very large databases," *NewsLetter – ACM-SIGMOD*, 25 (2), pp. 103-114, 1996.

[6] S. Guha, R. Rastogi, and K. Shim, "CURE: An efficient clustering algorithm for large databases," *News Letter – ACM-SIGMOD*, 7(2), pp. 73-84, 1998.

[7] G. Karypis, E-H Han, and V. Kumar, "CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling," *IEEE Computer*, 32 (8), 68-75, 1999.

[8] R.O. Duda and P.E. Hart, (1973). Pattern Classification and Scene Analysis. *A Wiley-Interscience Publication*, New York.

[9] R.T. Ng and J. Han, "Efficient and effective clustering methods for spartial data mining," *Proceeding of the VLDB Conference*, pp. 144-155, 1994.

[10] Y. Zhao and G. Karypis, "Evaluation of hierarchical clustering algorithms for document datasets," *Proceedings of the 11th International Conference on Information and Knowledge Management – ACM*, pp. 515-524, 2002.

[11] S. Salvador and P. Chan. "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms," *Tools with Artificial Intelligence - IEEE*, pp. 576-584, 2004.

[12] H. Koga, T. Ishibashi, and T. Watanabe. "Fast agglomerative hierarchical clustering algorithm using Locality-Sensitive Hashing," *Knowledge and Information Systems*, 12 (1), pp. 25-53, 2007.

[13] V.S. Murthy, E, Vamsidhar, J.S. Kumar, and P.S Rao, "Content based image retrieval using Hierarchical and K-means clustering techniques," *International Journal of Engineering Science and Technology*, 2 (3), pp. 209-212, 2010.

[14] S.J. Horng, M.Y. Su, Y.H. Chen, T.W. Kao, R.J. Chen, J.L. Lai, and C.D. Perkasa, "A novel intrusion detection system based on hierarchical clustering and support vector machines," *Expert Systems with Applications*, 38 (1), pp. 306-313, 2011.

[15] M.F. Balcan, Y. Liang, and P. Gupta, "Robust hierarchical clustering," *Journal of Machine Learning Research*, 15, pp. 3831-3871, 2014.

[16] S.M. Szilágyi, and L. Szilágyi, "A fast hierarchical clustering algorithm for large-scale protein sequence data sets," *Computers in Biology and Medicine*, 48, pp. 94-101, 2014.

[17] R.T. Ng, and J. Han, "CLARANS: A Method for Clustering Objects for Spatial Data Mining," *IEEE Transactions on Knowledge and Data Engineering*, 14 (5), pp. 1003-1016, 2005.

[18] Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data Mining and Knowledge Discovery,* 2 (3), pp. 283-304, 1998.

[19] H. Huang, Y. Gao, K. Chiew, L. Chen, and Q. He, "Towards effective and efficient mining of arbitrary shaped clusters," *Proceeding of 30th International Conference on Data Engineering – IEEE*, pp. 28-39, 2008

[20] T. Zhang, R. Ramakrishnan, and M. Livny, "BIRCH: An Efficient Data Clustering Method for Very Large Databases," *Proceedings of the 1996 ACM SIGMOD international conference on Management of data - SIGMOD '96*. pp. 103-114, 1996.

[21] H. Huang, Y. Gao, K. Chiew, K, L. Chen and Q. He, "Towards Effective and Efficient Mining of Arbitrary Shaped Clusters," IEEE 30th ICDE Conference, pp. 28-39, 2014.

[22] P. Berkhin, "A survey of clustering data mining techniques," *Grouping Multidimensional Data* – Springer, pp. 25-71, 2006.

[23] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "On clustering validation techniques," *Journal of Intelligent Information Systems*, 17 (2-3), pp. 107-145, 2001.

[24] J. Meng, S-J. Gao, and Y. Huang, "Enrichment constrained time-dependent clustering analysis for finding meaningful temporal transcription modules," *Bioinformatics*, 25 (12), pp. 1521–1527, 2009.

[25] A.T. Ernst and M. Krishnamoorthy, "Solution algorithms for the capacitated single allocation hub location problem," *Annals of Operations Research*, 86, pp. 141-159, 1999.

[26] M. Laan, and K. Pollard, "A new algorithm for hybrid hierarchical clustering with visualization and the bootstrap," *Journal of Statistical Planning and Inference*, 117 (2), p.275-303, Dec 2002.

[27] Y. Zhao, G. Karypis, and U. Fayyad, "Hierarchical Clustering Algorithms for Document Datasets," *Journal Data Mining and Knowledge Discovery archive,* 10 (2), pp. 141-168, March 2005

[28] S.A. Mingoti, and J.O. Lima, "Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms," *European Journal of Operational Research - Science Direct*. 174 (3), pp. 1742–17591, November 2006.

[29] A. Shepitsen, J. Gemmell, B. Mobasher, and R. Burke, "Personalized recommendation in social tagging systems using hierarchical clustering," Proceedings of the 2008 ACM conference on Recommender systems, pp. 259-266 (2008).

[30] H. Koga, T. Ishibashi, and T, Watanabe, "Fast agglomerative hierarchical clustering algorithm using Locality-Sensitive Hashing," *Knowledge and Information Systems,*12 (1), pp. 25-53, May 2007

[31] O.A. Abbas, Comparisons between Data Clustering Algorithms, *The International Arab Journal of Information Technology*, 5 (3), pp.320 – 325, 2008.

[32] G. Xin, W.H. Yang, and B. DeGang, "EEHCA: An energy-efficient hierarchical clustering algorithm for wireless sensor networks," *Information Technology Journal,* 7 (2), pp. 245-252, 2008.

[33] A.K. Jain,, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters - Science Direct,* 31 (8), pp. 651–666, June 2010

[34] V.S. Murthy, E. Vamsidhar, J.S. Kumar, and P.S. Rao, "Content based image retrieval using Hierarchical and K-means clustering techniques," *International Journal of Engineering Science and Technology*, 2 (3), pp. 209-212, 2010.

[35] Y. Cai, and Y. Sun, "ESPRIT-Tree: hierarchical clustering analysis of millions of 16S rRNA pyrosequences in quasilinear computational time". Nucleic Acids Res, 2011.

[36] S.J. Horng, M.Y. Su, Y.H. Chen, T.W. Kao, R.J. Chen, J.L. Lai, and C.D. Perkasa, " A novel intrusion detection system based on hierarchical clustering and support vector machines," *Exp. Sys. W. Appl*., 38, pp. 306-313, 2011.

[37] G. Kou, and C. Lou, "Multiple factor hierarchical clustering algorithm for large scale web page and search engine click stream data," *Annals of Operations Research,* 197 (1), pp. 123-134, August 2012 .

[38] A. Krishnamurthy, S. Balakrishnan, M. Xu, and A. Singh, "Efficient active algorithms for hierarchical clustering," *Proceedings of the 29th International Conference on Machine Learning*, pp. 887-894, 2012.

[39] P. Langfelder, and S. Horvath, "Fast R functions for robust correlations and hierarchical clustering," *J Stat Softw*., 46 (11), pp. 1-17, March 2012.

[40] Y., Malitsky, A. Sabharwal, H. Samulowitz, and M. Sellmann, "Algorithm portfolios based on cost-sensitive hierarchical clustering," *Proceedings of the 23rd international joint conference on Artificial Intelligence*, pp. 608-614, 2013.

[41] M. Meila, and D. Heckerman, "An experimental comparison of several clustering and initialization methods," *Proceedings of the 14th conference on Uncertainty in artificial intelligence*, pp. 386-395, 1998

[42] D. Müllner, "Fastcluster: Fast hierarchical, agglomerative clustering routines for R and Python," *Journal of Statistical Software*, 53 (9), pp. 1-18, 2013.

[43] M.F. Balcan, Y. Liang, and P. Gupta, "Robust hierarchical clustering" arXiv preprint arXiv:1401.0247, 2014.

[44] F. Murtagh, and P. Legendre, " Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion?", *Journal of Classification Archive,* 31 (3), pp. 274.295, October 2014.

[45] S.M. Szilágyi, and L. Szilágyi, "A fast hierarchical clustering algorithm for large-scale protein sequence data sets," *Comput. Biol. Med*., 48, pp. 94–101 (2014).

[46] E. Rashedi, A. Mirzaei, and M. Rahmati, "An information theoretic approach to hierarchical clustering combination," *Neurocomputing*, 148, pp. 487-497, 2015.

[47] K. Ding, C. Huo, Y. Xu, Z. Zhong, and C. Pan, " Sparse hierarchal clustering for VHR image change detection," *Geoscience and Remote Sensing Letters, IEEE,* 12 (3), pp. 577 – 581, 2015.