Journal of Informatics and Mathematical Sciences
Vol. 9, No. 3, pp. 509–520, 2017
ISSN 0975-5748 (online); 0974-875X (print)
Published by RGN Publications

http://www.rgnpublications.com

Proceedings of the Conference
Current Scenario in Pure and Applied Mathematics
December 22-23, 2016
Kongunadu Arts and Science College (Autonomous)
Coimbatore, Tamil Nadu, India

Research Article

Cluster Analysis using Rough Set Theory

Girish Kumar Singh and Shrabanti Mandal*

Department Computer Science & Applications, Dr Harisingh Gour University, Sagar, India
***Corresponding author:** shrabmandal@gmail.com

Abstract. Cluster is a group of objects of similar type and clustering is the process of finding clusters in dataset. Finding a set of clusters in a dataset is one fold of the data mining and it should be further analysis for knowledge. This paper present a method based on the concepts of Rough Set Theory to analysis the outcome of clustering process. The proposed method will able to explain the existence of clusters and why two clusters are different.

Keywords. Data mining; Clustering; Rough set theory

MSC. 54Hxx

Received: January 4, 2017 **Accepted:** March 16, 2017

1. Introduction

A useful technique in data mining known as Clustering is used to discover the data distribution and patterns in the underlying data. The process of making groups of a set of physical or abstract objects into classes of similar objects is called clustering. Discovering both the dense and the sparse regions in the data set is the main goal of clustering. A cluster is a collection

of data objects that are more similar to one another within the same cluster as compare to the objects in any of the other clusters. Data clustering is under dynamic developments and the contributing areas of research include data mining, statistics, machine learning, spatial database technology, biology and marketing. Clustering algorithms have several requirements as clustering has the potential for application in several fields. A typical set of requirements are: scalability, ability to deal with different types of attributes, discovery of clusters of arbitrary shape, minimal requirements for domain knowledge to determine input parameters, ability to deal with noisy data, insensitivity to order of input records, high dimensionality, constraint based clustering and interpretability and usability. The development and improvements of the main clustering techniques have been based either on the type of data or application domain. Therefore of the several good clustering algorithms none display a satisfactory tradeoff between several criteria.

Most of the clustering algorithm focuses on finding the best possible set of clusters. The objective focused in this paper is to introduce cluster analysis to explain why clusters are exists and how one cluster is different from other clusters. To explain these, concepts of rough set theory has been used.

Rest of the paper organized as follows: in Section 2, 3 and 4 basic concepts of clustering, rough set theory and reduct computation respectively has been given. Proposed method for cluster analysis is presented in Section 5. Section 6 gives the result of the experiment and analysis of the proposed method lastly Section7 concludes the paper.

2. Clustering

Clustering is a useful technique of data mining for the discovery of data distribution and patterns in the underlying data. The main focus of clustering techniques is to discover both the dense and the sparse regions in the data set. The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters [2]. The contributing areas of research include data mining, statistics, machine learning, spatial database technology, biology and marketing. Cluster analysis has been successfully applied in various disciplines which include pattern recognition, computer vision and data mining [10]. The clustering methods can be classified into partitioning, hierarchical, grid-based, density-based, model based and soft computing methods [15].

Density-based approach produce clusters of arbitrary shape cluster which are preferred for a number of application domains while the other approaches may unable to give the arbitrary shape cluster [5].

Partitioning algorithms partitions the dataset D of n objects in a d-dimensional space into k partition such that the cluster distribution is minimized where k the input parameter is the number of clusters. The deviation of a point can be computed differently in different algorithms and is more commonly called a similarity function. Partitioning algorithms use a two-step

procedure. Firstly, determine k representatives which minimize the objective function. Secondly, assign each object to the cluster with its representative "closest" to the considered object. The second step resulted in a partition which is equivalent to a Voronoi diagram and each cluster is contained in one of the Voronoi cells. Thus the shape of all clusters found by a partitioning algorithm is convex which is very restrictive in nature.

Hierarchical clustering algorithms create a hierarchical decomposition of the database D and this hierarchical decomposition is represented by a dendrogram. Dendrogram is a tree that iteratively splits dataset D into smaller subsets until each subset consists of only one object. In such a hierarchy, each node of the tree represents a cluster of D. The dendrogram can either be created from the leaves up to the root or from the root down to the leaves by merging or dividing clusters at each step. A termination condition has to be defined indicating when the merge or division process should be terminated. The complexity of hierarchical algorithms are $O(n^2)$. The one algorithm, Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) has the complexity $O(n)$, but this algorithm can handles only numerical data and moreover it is order-sensitive [21]. Also, BIRCH does not perform well when the clusters do not have uniform size and shape since it uses only the centroid of a cluster when redistributing the data points in the final phase. On the other hand, CURE employs a combination of random sampling and partitioning to handle large databases [6]. It identifies clusters having non-spherical shapes and wide variances in size by representing each cluster by multiple points. However, CURE is sensitive to some parameters. ROCK is a representative hierarchical clustering algorithm for categorical data [7]. It introduces a novel concept called "link" in order to measure the similarity/proximity between a pair of data points. Thus, the ROCK clustering method extends to non-metric similarity measures that are relevant to categorical data sets.

The grid-based clustering approach quantizes the space into a finite number of cells that form a grid structure on which all of the operations for clustering are performed. The main advantage of the approach is its fast processing time, which is typically independent of the number of data objects, yet dependent on only the number of cells in each dimension in the quantized space.

Density-based clustering approach based on the concept of continue growing a cluster as long as the density in the neighborhood exceeds some threshold; that is for each data point within a given cluster, the neighborhood of given radius has to contain at least a minimum number of points. Density based approach able to discover clusters of arbitrary shape. Such method may also be used to filter out noise (outlier) from the database. The two parameters for radius and minimum number of points have to be specified by the user. Also these two parameters are global and it may not reflect the local distribution of the data.

Model based methods attempt to fit given data to mathematical models. The fit between given data and mathematical should be optimized. A model-based method identifies groups of objects like conventional clustering method. Each group represents a concept or class and model-based clustering methods also find characteristic descriptions for each group.

In most of the clustering algorithm, to measure the similarity/dissimilarity between objects all the attributes are considered. In real life problem it may not be possible to give equal weight to the entire attribute i.e. all attributes are not of equal importance. For example, consider the admission process for science group at graduation level. There are various courses in science at graduation level viz. Physics (Hons.), Chemistry (Hons.), Mathematics (Hons.) etc. and each course has different weights for marks obtained in different paper in qualifying examination. In [4] weighted similarity/dissimilarity and other dissimilarity have been proposed.

Parameters used in various clustering algorithms are either local or global. A good clustering gives the local as well as global view and so parameters used in algorithms should be global as well as local. A good tradeoff between global and local view can be achieved in two ways: parameters should be local while cluster formation is global or parameter should be global and cluster formation is local to the data.

Another major problem in existing clustering algorithm is that they can handle same type of data either numerical or categorical, while a good clustering algorithm should handle the mixed data type. In view of these discussions, Attributed based Hierarchical Clustering (AHC) algorithm [16] has been presented and is given in Figure 1. This clustering algorithm has been used to analysis the result of the experiment. The basic idea of this algorithm is divisive and to measure the dissimilarity of objects at any time of division consider only one attribute. This clustering algorithm is also able to handle mixed data types.

Input:	Dataset D, Order of Attribute $A_1, A_2, \ldots, A_m$
Output:	Set of Cluster C
Step 1.	Initially $C \rightarrow \{D\}$
Step 2.	for each attribute A_i perform the following
Step 3.	Partition the each element of C with respect to the attribute A_i. Replace the elements of C with new set of partitions
Step 4.	End

Figure 1. The AHC Algorithm

3. Rough Set Theory

Rough Set theory is a technique for dealing with uncertainty and for identifying cause-effect relationship in databases as a form of data mining and database learning. It has also been used for improved information retrieval and for uncertainty management in relational databases. The Rough Set theory [16] was introduced by Pawlak in early 1980's.

An information system is a pair $S = (U, A)$, where U is a non-empty finite set of objects called the universe and A is a non-empty finite set of attributes. For each attribute there is a value set $V\alpha$, such that $\forall\ \alpha \in A$, $\alpha : U \rightarrow V\alpha$. Any information system S, of the form $S = (U, A \cup D)$ is called decision system where $A \cap D = \Phi$, D is set of decision attributes and A the set of conditional attributes.

For any subset of attribute B, i.e. $B \subseteq A$ there is associated an equivalence relation $\text{IND}_S(B)$ and is given by

$$\text{IND}_S(B) = \{(x, x') \in U \times U \mid \alpha \in B\alpha(x) = \alpha(x')\} \tag{3.1}$$

If $(x, x') \in \text{IND}_S(B)$, then objects x and x' are indiscernible from each other by attributes from B and $\text{IND}_S(B)$ is called the B-indiscernibility relation. The equivalence classes of the B-indiscernibility relation are represented by $[x]_B$. Indiscernibility relation partitions the domain into equivalence classes and is fundamental to rough sets.

For any subset of objects, $X \subseteq U$ a description of X can be obtained using only the information contained in some $B \subseteq A$ by constructing the B-lower approximation and B-upper approximation of X. B-lower approximation is given by

$$\underline{B}X = \{x : [x]_B \subseteq X\}$$

and B-upper approximation is given by

$$\bar{B}X = \{x : [x]_B \cap X \neq \phi\}.$$

The concept of Lower and upper approximation regions helps to distinguish between certain and possible (or partial) inclusions of objects from the universe U, respectively. The approximation regions can be defined using the indiscernibility relation as equivalence relation, which are equivalence classes.

The objects in $\underline{B}X$ with certainty are classified as members of X on the basis of the knowledge in B, while the objects in $\overline{B}X$ are classified as possible members of X on the basis of knowledge in B.

The set $BN_B = \overline{B}X - \underline{B}X$ is called the boundary region of X, and thus consists of those objects that cannot decisively be classified into X on the basis of knowledge in B.

The set $U - \overline{B}X$ is called the B-outside region of X and consist of those objects, which are with certainty classified as not belonging to X on the basis of knowledge in B. A set is said to be rough set if the boundary region is non-empty and if it is empty then crisp set.

A degree of attributes dependency and their significance can be determined using Rough Set Theory. For a given set of conditional attributes B, we can define the B-positive region $\text{POS}_B(D)$ in the relation $\text{IND}(D)$ as

$$\text{POS}_B(D) = \cup\{\underline{B}X : X \in D^*\}, \tag{3.2}$$

where D^* is the partition obtained by the equivalence class relation $\text{IND}(D)$.

All the objects in positive region $\text{POS}_B(D)$ contains those objects of U which can be classified without any error into distinct classes defined by $\text{IND}(D)$, based only on information in the relation $\text{IND}(B)$. An attribute or a set of attributes is said to be highly significant if cardinality of $\text{POS}_B(D)$ is high.

The Rough membership function used to express how strongly an object x belongs to the rough set X in view of the information about the object expressed by the set of attributes B. The Rough membership function can measure the significance of an attribute and is defined by,

$$\mu_X^B(x) = \frac{\text{card}(X \cap [x]_{\text{IND}(B)})}{\text{card}([x]_{\text{IND}(B)})}. \tag{3.3}$$

4. Reduct Computation

One of the most important concepts in application of rough set theory into data mining is Reduct. A minimal set of attributes preserving classification accuracy of original dataset is known as reduct. To find reducts is similar to the problem of feature selection. All reducts of a dataset are obtained by constructing a discernibility function from the dataset [13]. Finding minimal reduct or all reducts are both NP-hard problems. Therefore, efficient methods to solve this NP-hard problem play an important role in the development of rough set-based data mining. To find reduct some heuristics algorithms have been proposed as well as some algorithms using genetic algorithm are also proposed. Starzyk use strong equivalence to simplify discernibility function [17]. However, there are no universal solutions. This is still an open problem in rough set theory.

There are many algorithms that have been developed for reduct computation. The conventional reduct computational algorithms can be categorized into two categories: namely reduction algorithms based on heuristic information and the reduction algorithms based on random strategies. However, all of these algorithms cannot guarantee to find a complete set of reducts for the dataset.

Heuristic Algorithm. Johnson's strategy [12] is based on Johnson approximation algorithm for computing minimal prime implicant of any Boolean function in *conjunctive normal form* (CNF) formula. The main idea behind this algorithm is to find an attribute discerning the largest number of pairs of objects, i.e., attribute most often occurring in the entries of discernibility matrix and continue this process until a reduct set is found. The time complexity of this algorithm is $O(|A|^2|U|^2)$ and the space complexity of this algorithm is $O(|A||U|^2)$, where A is the set of attributes and U is database.

An attributes reduction algorithm based on significance of attributes in discernibility matrix to find reduct has been proposed by Jue Wang [20]. Significance of attributes, defined as the attributes frequency in discernibility matrix uses in this algorithm and so this algorithm regards the number of occurrences of each attribute as the significance of each attribute. This algorithm first selects an attribute with the largest frequency, and deletes the elements involved with the selected attribute in discernibility matrix. Then count the frequency of other attributes and select again. This algorithm proceeds until a reduct set is found. The time complexity of this algorithm is also $O(|A|^2|U|^2)$ and the space complexity of this algorithm is also $O(|A||U|^2)$.

Keyun Hu [9] has developed a feature ranking mechanism using attribute frequency information in discernibility matrix. Using feature ranking as heuristics reduct computation algorithm has been proposed. The time complexity of this algorithm is $O((|A|+\log|U|)|U|^2)$ and the space complexity of this algorithm is $O(|A||U|^2)$.

X. Hu *et al*. [8] proposed a new rough sets model and redefined the core attributes and reducts based on relational algebra which is very efficient in set-oriented database operations. Based on this proposed new model and new definitions, they presented two new algorithms. One to calculate core attributes and second to reducts for feature selections. However, the time complexity of their reduct algorithm is at least $O(|A|^2|U|)$ even if the database systems provide both hashing and indexing mechanism.

Random Reduct algorithm. Vinterbo [19] has formulated the problem of rough set based attribute reduction as 'minimal hitting set' problem. In this formulation he defined an r-approximate hitting set. r-approximate hitting set is a set that intersects with at least a fraction r of given sets. Approximations of reducts from rough set theory are defined by means of minimal r-approximate hitting sets. A genetic algorithm is devised to compute r-approximate hitting sets. The time complexity of this algorithm is $O(|A|^2|U|\log|U|)$ and the space complexity of this algorithm is $O(|U|)$. Problem with this algorithm is that, the reducts obtained by this algorithm are not guaranteed to be complete.

Bazan [1] analysis the above methods and found that Vinterbo not taking into account the fact that part of reduct set is chaotic i.e. it is not stable in randomly chosen samples of a given decision table. Based on his observation he has proposed a method for selection of feature (attribute) sets relevant for extracting rules corresponding to the data and these sets of attributes are called dynamic reducts. Dynamic reducts are in some sense the most stable reducts of given decision table, i.e., they are the most frequently appearing reducts in sub tables created by randomly sampling of a given decision table. Reduct computing of variable size dynamically can be extremely computationally intensive, even for decision tables of moderately size. This algorithm is quite stable in most cases, but it is still not a complete reduct computation algorithm.

QuickReduct algorithm [3], [11] attempts to calculate a minimal reduct without exhaustively generating all possible subsets. It starts with an empty set and adds in turn those attributes that result in the greatest increase in $\gamma_p(Q)$, Q is decision attribute, until this produces its maximum possible value for the dataset (usually 1), where

$$\gamma_p(Q) = \frac{|\mathrm{POS}_p(Q)|}{|U|} \tag{4.1}$$

and

$$\mathrm{POS}_p(Q) = \bigcup\{\underline{p}X : X \in [x]_Q\}. \tag{4.2}$$

However, it has been proved that this method does not always generate a minimal reduct. It does result in a close-to-minimal reduct.

5. Cluster Analysis

After clustering the outcome of the clustering process needs to be analyzed to answer the questions such as the following ones:

- What do these clusters represent?
- Why do these clusters exist?
- Why is one cluster different from other clusters?

By answering these questions, analyst will able to propose suggestions to improve the certain clusters so that objects of those clusters have the same properties like the properties of objects of other cluster having good properties. For example, consider the case of migration of population from one place to another. Cluster analysis on this data may give different clusters. From these cluster one should be able to give the answer to following questions:

- Why the rate of migration is more in particular places than the others?
- What is the reason of migration of people?

To answer above questions the cluster analysis should produce the characteristics of the clusters such as: what do the clusters represent, where they are and why they are? By answering these questions the authority may be able to take some appropriate action to control the migration of the people.

In this section a method to answer the following two questions is proposed based on Rough Set Theory:

- Why does each cluster exist?
- What distinguishes one cluster from other clusters?

By the definition of the cluster, cluster is the collection of the similar objects and so one can assume the cluster id as the class of the data. By assuming the cluster id as class label, rough set theory can be applied on the dataset. A reduct is the minimal attribute set preserving classification accuracy of all attribute of original dataset. Reducts can be used as a tool for analysis for the existence of a cluster. In next subsections method to discriminate between clusters and to characterize of clusters has been discussed.

5.1 Difference between Clusters

Reduct by definition is the minimal set of attribute which preserve the classification accuracy of the original data. This property of the reduct has been used to identify and describe the difference between clusters.

To determine the difference between any two clusters say C_1 and C_2 with cluster id 1 and 2 find the reduct of the objects belonging to the two clusters, considering cluster id as the

class label. There may be two possibilities, first reducts exist and second there is no reduct for the object belonging to these two clusters. If reduct not exist then find the set of significant attributes. The significance of an attribute a_1 with respect to decision attribute D is given by $\gamma_{ai}(D)$, as in equation (4.1). The set of attributes with significance higher than some threshold value θ, $0 < \theta \leq 1$, may be considered as significant attribute, a suitable value for θ is 0.5.

As the clusters have different set of attribute values corresponding the reduct or set of significant attributes, the values of these attributes help to discriminate the clusters or to explain to differentiation between clusters.

5.2 Existence of Cluster

The attribute values characterizing a cluster corresponding to the reduct can be used to comprehend the reason behind the existence of the cluster. Find the reduct for the given dataset considering cluster id as the class data is to be computed. As reducts preserve the classification accuracy of the original data so the different clusters exist due to different values set of these reduct. In this case if reducts do not exist the set of significant attributes are proposed to be used for this purpose. As the clusters have different attribute values of attributes from the reduct or set of significant attribute, the attribute values of reduct or significant attribute for each cluster is proposed to be used to characterize the clusters i.e. the existence of the clusters

6. Result and Analysis

Proposed scheme for cluster analysis has been implemented in C language. The experiments were carried out on the labeled Iris Plants dataset (iris) obtained from the UC Irvine ML repository [18]. This database has 150 instances in number with all four attributes having continuous values. Detailed information about the attribute has been presented in the Table 1.

Table 1. Description of Attribute of the Iris Database

Sl. No.	Attribute	Short name
1	Sepal Length	*SL*
2	Sepal Width	*SW*
3	Petal Length	*PL*
4	Petal Width	*PW*

The proposed cluster analysis method has been applied on the results of two clustering algorithms AHC and DBSCAN. To apply these two clustering algorithm class label has not been considered. AHC has been applied on the iris data two times with different order of attributes consider for clustering. Results are presented in the Table 2 corresponding to two different ordering of attributes for clustering, number of clusters obtained and details of clusters.

Table 2. Description of clusters obtained by AHC of the Iris Database

Order of attribute used for clustering	Number of clusters	Details of object in clusters
SL, *SW*, *PL*, *PW*	5	{1:17, 2:11}, {2:23, 3:10}, {1:33, 3:1}, {2:15,3:31}, {2:1,3:8}
PL, *PW*, *SL*, *SW*	4	{1:50}, {2:11}, {2:38, 3:5}, {2:1, 3:45}

Detail of cluster shows type of object in the cluster with respective count. Object type and respective count is presented in the form $\{o_1 : n_1, o_2 : n_2\}$ where o_1 and o_2 represents object type i.e. class label while n_1 and n_2 represents the count of the object of type o_1 and o_2 respectively in this cluster. DBSCAN has been used to partition the attribute values with MinPts = 3 and Eps has been calculated.

The result of cluster analysis method is presented in the Table 3. Result indicates that the number of clusters obtained by different clustering algorithm viz. DBSCAN and AHC. The set of reduct of the dataset has been computed considering the resultant cluster id as class label. This table also includes the significance of all the attributes with respect to cluster id of various clustering approaches. Significance of the attribute is calculated by the rough set theory concept positive region POS$p(Q)$, where p is the attribute for which significance has to be measure and Q is the class label i.e. cluster id.

Table 3. Existence of clusters obtained by DBSCAN and AHC

Clustering Algorithm	Number of Cluster	Reduct if any	Significance of attributes
DBSCAN	4	{*SL*,*SW*,*PL*}, {*SL*,*SW*,*PW*}, {*SW*,*PL*,*PW*}	$SL = 0.13$, $SW = 0.17$, $PL = 0.86$, $PW = 0.69$
AHC	4	{*PL*,*PW*}	$SL = 0.19$, $SW = 0.21$, $PL = 0.8$, $PW = 0.78$

Table 4. Difference of clusters obtained by DBSCAN

Clusters consider for analysis		Reduct if any	Significance of attributes
First cluster	Second Cluster		
1	2	{*PL*}, {*PW*}	$SL = 0.75$, $SW = 0.47$, $PL = 1$, $PW = 1$
1	3	{*SW*}, {*PL*}, {*PW*}	$SL = 0.57$, $SW = 1$, $PL = 1$, $PW = 1$
1	4	{*PL*}, {*PW*}	$SL = 0.89$, $SW = 0.41$, $PL = 1$, $PW = 1$
2	3	{*SL*}	$SL = 1$, $SW = 0.8$, $PL = 0.96$, $PW = 0.8$
2	4	{*SL*,*SW*,*PL*}, {*SL*,*SW*,*PW*}, {*SW*,*PL*,*PW*}	$SL = 0.18$, $SW = 0.07$, $PL = 0.78$, $PW = 0.62$
3	4	{*SL*}, {*SW*}, {*PL*}, {*PW*}	$SL = 1$, $SW = 1$, $PL = 1$, $PW = 1$

To see the result of pair wise cluster analysis using the proposed method to differentiate between the resultant clusters of DBSCAN and AHC is displayed in Table 4 and 5, respectively.

Table 5. Difference of clusters obtained by AHC

Clusters consider for analysis		Reduct if any	Significance of attributes
First cluster	Second Cluster		
1	2	{*PL*}, {*PW*}	*SL*=0.73, *SW*=0.44, *PL*=1, *PW*=1
1	3	{*PL*}, {*PW*}	*SL*=0.39, *SW*=0.93, *PL*=1, *PW*=1
1	4	{*PL*}, {*PW*}	*SL*=0.9, *SW*=0.32, *PL*=1, *PW*=1
2	3	{*PL*}	*SL*=0.63, *SW*=0.6, *PL*=1, *PW*=0.39
2	4	{*PW*}	*SL*=0.24, *SW*=0.09, *PL*=0.66, *PW*=1
3	4	{*PL*}, {*PW*}	*SL*=0.84, *SW*=0.72, *PL*=1, *PW*=1

From Table 4 and 5, it is clear that to differentiate any two clusters obtained by AHC only one attribute is sufficient while in DBSCAN for some cluster 2 or 3 attributes are required for same.

7. Conclusion

This paper presents a method for cluster analysis based on rough set theoretic concepts. This analysis can give the reason behind the existence of the cluster and why one cluster is different from other clusters. By this analysis of clusters, analyst will able to explain why such type of grouping exists in the data. Moreover knowing the reasons for existence of different clusters, bad characteristics cluster may be improve for good one.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

All the authors contributed significantly in writing this article. The authors read and approved the final manuscript.

References

[1] J.G. Bazan, A comparison of dynamic and non-dynamic rough set methods for extracting laws from decision tables, in *Rough Sets in Knowledge Discovery 1: Methodology and Applications*, Polkowski and Skowron (editors), Physica-Verlag, Heidelberg, Germany, Chapter 17, pp. 321–365 (1998).

[2] P. Berkhin, A survey of clustering data mining techniques, in *Grouping Multidimensional Data*, Springer, pp. 25–71 (2002).

[3] A. Chouchoulas and Q. Shen, Rough set-aided keyword reduction for text categorisation, *Applied Artificial Intelligence* **15** (9) (2001), 843–873.

[4] P.A. Devijver and J. Kittler (eds.), *Pattern Recognition Theory and Application*s, Springer-Verlag (1987).

[5] M. Ester, H.P. Kriegel, J. Sander and X. Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining* (KDD'96), Portland: Oregon, pp. 226–231.

[6] S. Guha, R. Rastogi and K. Shim, Cure: an efficient clustering algorithm for large databases, in *ACM SIGMOD International Conference on the Management of Data*, (Seattle, WA, USA), pp. 73–84 (1998).

[7] S. Guha, R. Rastogi and K. Shim, ROCK: A robust clustering algorithm for categorical attributes, *Information Systems* **25** (5) (2000), 345–366.

[8] X. Hu, T.Y. Lin and J. Han, A new rough set model based on database systems, *Journal of Fundamental Informatics* **59** (2004), 135–152.

[9] K. Hu, Y. Lu and C. Shi, Feature ranking in rough sets, *AI Communications, Special issue on Artificial intelligence advances in China* **16** (1) (2003), 41–50.

[10] A.K. Jain and R.C. Dubes, *Algorithms for Clustering Data*, Prentice Hall, Upper Saddle River, New Jersey (1988).

[11] R. Jensen and Q. Shen, A rough set-aided system for sorting WWW bookmarks, in *Proceedings of the First Asia-Pacific Conference on Web Intelligence: Research and Development* (WI'2001), 2001, pp. 95–105 (2001).

[12] D.S. Johnson, Approximation algorithms for combinatorial problems, *Journal of Computer and System Sciences* **9** (1974), 256–278.

[13] S.K. Pal and A. Skowron, *Rough Fuzzy Hybridization – A New Trend in Decision Making*, Springer (1999).

[14] Z. Pawlak, Rough sets, *International Journal of Computer and Information Sciences*, **11** (1982), 341-356.

[15] L. Rokach and O. Maimon, Clustering methods, in *Data Mining and Knowledge Discovery Handbook*, Springer US, pp. 321–352 (2005).

[16] G.K. Singh and S. Minz, Attribute based hierarchical clustering algorithm, *National Conference on Trends in Advance Computing* (NCTAC'07), pp.177–183 (2007).

[17] J. Starzyk, D.E. Nelson and K. Sturtz, Reduct generation in information systems, *Bulletin of International Rough Set Society* **3** (1998), 19–22.

[18] UCI Repository of Machine Learning, `https://archive.ics.uci.edu/ml/datasets/Iris`.

[19] S. Vinterbo and A. Ohrn, Minimal approximate hitting sets and rule templates, *International Journal of Approximate Reasoning* **25** (2) (2000), 123–143.

[20] J. Wang and J. Wang, Reduction algorithms based on discernibility matrix: the ordered attributes method, *Journal of Computer Science & Technology* **16** (6) (2001), 489–504.

[21] T. Zhang, R. Ramakrishnan and M. Livny, BIRCH: An ecient data clustering method for very large databases, in *Proceedings of ACM SIGMOD*, Montreal Canada, pp. 103–114 (1996).