International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 02, 2020
ISSN: 1475-7192

# Clustering Algorithms using Splitting Attribute Approach

[1]Arun K R, [2]Tripathy B K

***ABSTRACT*** **-**Computer field is growing rapidly in many areas like Artificial intelligence, data mining, cloud computing and more. Out of those data mining keenly seeks the attraction by many people due to its applicability. Out of the steps in data mining, clustering is one of the most important steps that face lot of problems due to lack of data. Factors like imprecision, uncertainty, missing value and more affects the data directly or indirectly. In order to find a solution for the above said many algorithms are evolving day by day according to the need of the problems. Fuzzy set and rough set are the older techniques which are most often used by many researchers. In the modern era, algorithms like MMR, MMeR, MMeMeR, SSDR are highly attracted by recent researchers which gives the better results compared to rough and fuzzy. In this paper, the detailed comparison of those algorithms with each other and few applications related to attribute clustering is briefly explained.

***keywords–*** *Rough Set, Fuzzy dominance, MMR, MMeR, SDR.*

## I. INTRODUCTION

Data mining is the main area which has enormous of growth for research and also paves a way for building an application in real time scenarios. Out of the seven steps in data mining, clustering plays a vital role. Taking an element and placing it into the correct cluster is the major task in clustering. To cluster the data into the correct cluster sufficient of data about the element is needed. This is the main thing that is missing in the recent modern databases. The data are incomplete and is of no use for anyone. Imprecise and uncertainty data are not useful for applications or research because of its lack of accuracy and exactness respectively. Missing value are also not helpful to create a powerful application which leads to bugs. The various approaches like fuzzy set[15] and rough set[7] gives an approximation results for all type of incomplete data. These approaches somehow solve the incompleteness of data. Followed by both rough set and fuzzy set, several algorithms are developed namely TR[8], MMR[6], MMeR[9], SDR[12], SSDR[13], MMeMeR[14] etc. and many applications are build based on these algorithms.

Both fuzzy set and rough set are lacking in parameterization. Hence a new approach called soft set[4] was introduced which solves the parameterization problem. Because of its parameterization in nature, soft set has been used in various algorithms and applications and got attracted by various researches.

Section 2 comprises of the comparative study about the various algorithms related to rough set. Section 3 comprises of the comparative study about the various algorithms related to fuzzy set. Section 4 comprises of the comparative study about the various algorithms related to soft set. Also, few corrections that are to be made in the applications are also given.

[1] *School of Computer Science and Engineering, VIT University, Tamilnadu, Vellore 632014, INDIA, arun.kr@vit.ac.in*
[2] *School of Computer Science and Engineering, VIT University, Tamilnadu, Vellore 632014, INDIA, tripathybk@vit.ac.in*

**Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020** 

# II. ROUGH SET BASED PARTITIONING

Rough set theory was the first approximation approach that is used for clustering the data. Out of the two methods namely supervised and unsupervised, supervised mining methods can deal with basic domain knowledge whereas the unsupervised mining has only the heuristic approach. This paper mainly deals with the partitioning of attributes where cohesion is more and coupling is less. Different approaches for partitioning attribute are arbitrary, imbalanced and balanced method. Arbitrary method chooses an attribute with minimum dissonance irrespective of size of the partition. Imbalanced method partitions the attribute based on the minimum disorder which gives the most imbalanced partitions. Balanced partitions is the quite opposite of imbalanced one.

During both balance and imbalanced partition, two major problems will occur. First is, if there are two attributes deals with balanced partition then there will be the two candidate partition attributes. Second is for forming binary partitions two attributes are needed. If there is no two valued attribute then there is no way of multiple valued attributes. Here rough set plays a vital role where both lower and upper approximations are used. Using lower and upper approximations boundary region and total roughness can be calculated.

*Partitioning multiple values:*

Three steps are used for partitioning the multiple values:

1) Choosing the multiple valued attribute
2) Find the cluster center by calculating the total roughness
3) Clustering other values with the clustering center.

### *A. Min Min Roughness (MMR)*

In modern database, categorical data plays a vital role. All the dataset that is taken from all the resources contains categorical data in it. Out of those, the main task is to separate the data according to their type. Here clustering comes into the picture. Fuzzy set, rough set and soft set are the various new approaches that is used to solve uncertain data. Out of those fuzzy c means is the first algorithm that is used to cluster the data according to their nature. After a while, MMR is the algorithm which was developed by Parmar [6] along with rough set theory paves a new way for clustering categorical data.

*Drawback in earlier methods:*

Rough set theory is the approach that is been followed to classify the attributes for clustering. Using lower and upper approximation the total roughness is calculated for finding the value for the partition. Binary valued attributes are used to find the roughness for multi valued attributes is the great backlog when the value of the multi value is lower.

***Methodology used:***

MMR overcomes the above said problem and introduces the new algorithm to find the data similarity between the objects.

***Algorithm:***

Step1: Indiscernibility relation and equivalence class is defined

Step2: Lower and upper approximation is defined.

**Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020** 

Step3: Roughness and Mean roughness is calculated.

Step4: MR is calculated with respect to all attributes which takes the minimum from the n-roughness calculated. Then the same way MR is calculated for all the other attributes.

Step5: From the n-MR calculated, Minimum value is chosen, which is MMR.

Finally as per our need of cluster k, the algorithm is repeatedly executed.

*B. MMeR*

MMeR algorithm was developed by Tripathy et al. which was the extended work of MMR. MMeR has a few better enhancement compared to the older techniques like handling uncertain data, large data sets and also in the efficiency of the algorithm. Also, MMeR has an ability to work on heterogeneous data.

***Algorithm:***

Following MMR, MMeR has the basic steps same:

Step1: Calculating the indiscernibility relation

Step2: Calculating approximations (lower and upper)

Step3: finding the roughness which is used to find the accuracy of estimation.

Step4: finding the relative roughness with approximations.

Step5: finding the Mean of roughness of all the attributes (MeR)

Step6: taking the minimum out of all MeR's obtained. (MMeR).

***Enhancement done:***

- Finding the distance of relevance (DR) between the two objects is the new approach that has been introduced in this paper.
- Relative roughness is enhanced by considering the least roughness of attribute rather than taking the least mean with equivalence class is changed.
- Cluster selection uses DR for finding the average distance between the objects which is been derived from the hamming distance.

***Formula:***

Taken objects A and D from the categorical data with n attributes, DR is defined as

Here $a_i$ and $d_i$ are values from the objects A and D under ith attribute.

Case 1: If the attributes values are equal, zero and one is added. Also, the cluster with least average distance is selected. If the values are found to be same, then the number of objects in which cluster is more will be chosen. If a tie occurs, the random selection will be made.

***Advantage over MMR:***

When there are more elements in the cluster is said to be same or equal, the splitting attribute selection is not clear where, MMeR has given a solution for it.

### *C. Min Mean Mean Roughness: (MMeMeR)*

Compared to the older algorithms like MMR and MMeR, MMeMeR has a better accuracy which is calculated by the purity ratio. The results were tested with the three datasets soybean, zoo and mushroom where the data is taken from UCI repository.

***Algorithm:***

Step1: Calculating the indiscernibility relation and equivalence class

Step2: finding the roughness after calculating lower and upper approximations

Step3: calculating the relative roughness

Step4: finding the mean roughness MeR

Step5: finding the mean for mean roughness MeMeR

Step6: Taking the minimum of from step5.

Thus MMeMeR is done.

As mentioned in MMeR, the accuracy is measured in two terminologies.

1. Distance of relevance
2. Overall purity

*Formula for purity ratio:*

$$purity = \frac{\begin{matrix}The\,no.of\ data\,occuring\,in\,the\,ith\,cluster \\ and\,its\,corresponding\,class\end{matrix}}{The\,number\,of\ data\,in\,the\,dataset}$$

$$overall\ purity = \frac{\sum_{i=1}^{\#ofclusters} Purity(i)}{\#ofclusters}$$

*Analysis of MMeMeR using the datasets:*

***1)*** Soybean ***data set***

The data set is taken from UCI benchmark which contains 47 entities, 35 attributes and 4 classes. The measures of efficacy with respect to purity ratio with various algorithms are tabulated below:

TABLE I

| K Modes | Fuzzy K Modes | Fuzzy Centroids | MMR | MMeR | SDR | MMeMeR |
|---|---|---|---|---|---|---|
| 0.69 | 0.77 | 0.97 | 0.83 | 0.83 | 0.93 | 0.9425 |

Out of the various algorithms, fuzzy centroid gives the better ratio compared to all other. MMeMeR stands in second which shows that it is more accurate compared to MMR, MMeR and SDR.

***2) Zoo data set***

The data set is taken from UCI benchmark which contains 101 entities, 18 categorical attributes and 7 decision classes. The measures of efficacy with respect to purity ratio with various algorithms are tabulated below:

TABLE II

| K Modes | Fuzzy K Modes | Fuzzy Centroids | MMR | MMeR | SDR | MMeMeR |
|---|---|---|---|---|---|---|
| 0.60 | 0.64 | 0.75 | 0.787 | 0.902 | 0.907 | 0.902 |

Out of the various algorithms, SDR gives the better ratio compared to all other. MMeMeR stands in second which shows that it is more accurate compared to others.

***3) Mushroom data set***

The data set is taken from UCI benchmark which is the very large data set that contains 8124 entities, 22 categorical attributes and 20 decision classes. The measures of efficacy with respect to purity ratio with various algorithms are tabulated below:

TABLE III

| MMR | MMeR | SDR | MMeMeR |
|---|---|---|---|
| 0.84 | 0.9641 | 0.9723 | 0.973 |

Out of the various algorithms, MMeMeR gives the better ratio compared to all.

*D. SSDR*

The heuristic algorithm that are in use are not have much efficacy and also not able to deal with categorical data and uncertainty. This helps to increase the research in the algorithm that deals with the above said. MMR, MMeR and SDR are some of the algorithms which have a pros and cons of its own. In this paper, the standard deviation of standard deviation roughness (SSDR) is introduced. This algorithm works for numerical data and also works fine for the categorical data. The efficiency of this algorithm is also shows the better results compared to the earlier one.

***Deficiency:***

The algorithms that are working good are deals only with the numerical data as it is easy to work on it. But categorical data algorithms are not good enough in purity because of multivalued in nature. Few algorithms runs better way but have major issues like instability in multiple runs of algorithm and fails to deal with uncertainty.

***Comparison with other algorithms****:*

- Expectation-Maximization :

cons: locally optimal solution

- Association rule hyper graph:

Pros: Deals with Binary transactional data

Cons: Assumes clusters are disjoint and no overlap

- K-modes and K-means:

K means introduces new dissimilarity measure for the categorical data and K mode extends the clusters into modes.

Cons: produces local optimal solution, stability issues

- Cactus: Clustering categorical data using summaries

Methodology used: Summarization based algorithm.

Rock: Robust clustering using links

Methodology used: Measures the similarity or proximity values between pair of objects in the clusters.

- Stirr: Sieving through iterated relational reinforcement

Methodology used: Iterative approach, Converts categorical data into non-linear dynamic systems

- Cons: convergence issue

After all the algorithms defined above, a new algorithm for uncertainty was developed as fuzzy k means which deals with membership values. It is been further enhanced with various hybrid methods and found that rough set theory suits the need of all.

The SSDR approach also follows the rough set concept along with standard deviation.

***Algorithm:***

Step1: Indiscernibility relation and equivalence class

Step2: lower and upper approximation

Step3: roughness and relative roughness

Step4: mean roughness

Step5: standard deviation for all attributes

Step6: Minimum of all SD

***Experimental analysis:***

The zoo data set with 18 attributes where out of that 15 are Boolean values which has only true or false and 2 attributes deals with numbers and one attribute is the name of the animal. In total it has 101 objects in it. It is been divided into 7 classes and so cluster value is also 7.

TABLE IV

| K-modes | Fuzzy K modes | Fuzzy centroids | MMR | MMeR | SDR | SSDR |
|---|---|---|---|---|---|---|
| 0.6 | 0.64 | 0.75 | 0.787 | 0.902 | 0.907 | 0.907 |

The zoo data set is been taken and the algorithm is verified with distance of relevance and purity ratio is calculated which gives the better results compared to other algorithms.

***Advantages of SSDR:***

1) Uses uncertain data and provides stable results

2) The algorithm is developed in such a way it deals with both numerical and categorical data.

3) Distance of relevance and purity ratio are better compared to the previous algorithms and is verified using the data set.

## III. A REVIEW ON FUZZY DOMINANCE MATRIX[11]

Dominance matrix using fuzzy set concept is used for multiple attribute decision making. When there are n-experts, the decision making is a major task. The dominance over the other experts is calculated to find the better decision. Usually dominance is given in the form of linguistic variables. This paper also follows fuzzy dominance matrix (FDM) with an algorithm. The main part in multiple attribute decision is to find the better output result than compared to the alternatives decisions.

Methodology ***used:***

1) Fuzzy decision matrix is used to store the individual expert value. In this, the values are stored in the form of fuzzy where the value ranges from 0 to 1.

2) Fuzzy dominance matrix is used to find the degree of dominance of an individual expert compared to others.

3) It is achieved with operations like addition, subtraction and product of dominance matrix.

***Application:***

An application is built with three experts and four candidates with respect to four parameters and the algorithm is explained. Initially the opinion of the individual expert is noted in the form of matrix with fuzzy values. Then the difference between the expert 1 and 2, expert 1 and 3 and expert 2 and 3 are calculated and stored in the matrix form. The maximum from the above difference is chosen and the sum of all gives the final choice value.

## IV. SOME COMMON ATTRIBUTE A REVIEW ON A NOVEL SOFT SET APPROACH IN SELECTING CLUSTERING ATTRIBUTE [5]

***Introduction:***

One of major task in data mining is data clustering. In data clustering, finding the clustering attribute is the major task. In 1999, Molodtsov introduced a new approach for dealing with uncertainty called as soft set. In this paper, the rough set approximations are combined with soft set to select the clustering attribute from others. Also, the algorithm that is defined in this paper is used to solve categorical data. The experimental analysis shows that the proposed algorithm novel soft set approach (NSS) is better than maximum dependency attribute (MDA) and has been tested with 15 benchmark datasets.

Advantage: scalability.

Compared to total roughness, MMR and MDA, NSS performance is better and also the efficiency is verified with the large data sets.

***NSS methodology:***

Both soft set and rough set approaches are joined together. A new model for soft set for information system is built. Rather than considering the single object, equivalence classes for rough set theory is generated. Using the

soft set model, the equivalence classes and approximations for the rough set are created. Then an algorithm is built based on that for finding the clustering attribute for categorical data.

Basic definitions:

***Rough set theory-Definition:***

Rough set theory was initiated by Pawlak (1982). Let U be a universe and R be an equivalence relation over U. This equivalence relation decomposes U into disjoint equivalence classes. We denote the equivalence class of an element x with respect to R by $[x]_R$ , which is defined as $[x]_R = \{y \mid yRx\}$ .

Then for any $X \subseteq U$ , we associate two crisp sets $\underline{R}X$ and $\bar{R}X$ called the lower and upper approximations of X with respect to R respectively and are defined as, $\underline{R}X = \{x \in U : [x]_R \subseteq X\}$

$\bar{R}X = \{x \in U : [x]_R \cap X \neq \phi\}$ Also, we define the boundary region of X with respect to R by $BN_R(X)$ and it is defined as $BN_R(X) = \bar{R}X - \underline{R}X$ . X is said to be rough with respect to R if and only if $\bar{R}X \neq \underline{R}X$ or equivalently, $BN_R(X) \neq \phi$ . Otherwise, X is said to be R-definable.

Any element *x* of the lower approximation of X is said to be certainly the element *x* belong to X. Similarly, any element *x* of the upper approximation of X is said to be the element *x* belong to or may not belong to X. The boundary region is the region of uncertainty of X with respect to R. This convention follows the idea of uncertainty introduced by Gotlab Frege, the father of modern logic.

***Soft set theory-Definition:***

It was observed by Molodtsov in 1999 that fuzzy sets lack parametrization. In order to handle this he introduced the notion of soft set. It is worth noting that Molodtsov has shown that all fuzzy sets are soft sets. The definition of soft set is as follows:

Let U be a universal set and E be a set of parameters. Then the pair (U, E) is called a soft universe. Members of the universe and the parameter set are generally denoted by x and e respectively. A soft set over the soft universe (U, E) is denoted by (F, A), where $F : A \rightarrow P(U)$ , where A is a subset of E and P (U) is the power set of U which comprises of all the subsets of U.

MDA algorithm:

**Algorithm 1: MDA**

**Input:** Data set without a clustering attribute
**Begin**
Step 1: Compute the equivalence classes using the indiscernible relation on each attribute;
Step 2: Determine the dependency degree of attribute $a_i$ with respect to all $a_j$ , where $i \neq j$ ;
Step 3: Select the maximum dependency degree of each attribute;
Step 4: Select a clustering attribute based on the maximum degree of the dependency attributes.
**Output:** A clustering attribute
**End**

NSS algorithm:

| **Algorithm 2: A novel soft set (NSS) approach in selecting clustering attribute** |
|---|
| **Input:** Data set without a clustering attribute<br>**Begin**<br>Step 1: Compute the equivalence classes using the indiscernible relation on each attribute;<br>Step 2: Construct soft set $(F_1, E)$;<br>Step 3: Construct the tabular representation of the soft set $(F_1, E)$;<br>Step 4: Compute the cardinality of lower approximation of an attribute with respect to other attributes in terms of Equation (8);<br>Step 5: Compute the dependency of an attribute with respect to other attributes in terms of Equation (14);<br>Step 6: Select the attribute with the highest dependency as the clustering attribute.<br>**Output:** A clustering attribute<br>**End** |

***Corrections to be made:***

There are three major mistakes found in this paper.

1. According to the table given as object appearance, the equivalence classes are defined. Out of those, the equivalence class for shape is not given correctly.

Given: U / shape ={{1,2},{3,4},{5}}

Corrected: U / shape ={{1,2,5},{3,4}}

2. The example $K_{Color}$(Area) is not correct.

Given:

$$k_{Color}(\text{Area}) = \frac{\sum_{X \in U/Color=} \left|\underline{Color}(\text{X})\right|}{|U|}$$

$$= \frac{|\{1,2\}| + |\{3,5\}| + |\{4\}|}{|1,2,3,4,5|} = 1$$

Corrected:

$$k_{Color}(\text{Area}) = \frac{\sum_{X \in U/Color=} \left|\underline{Color}(\text{X})\right|}{|U|}$$

$$= \frac{|\{\phi\}| + |\{3,5\}| + |\{4\}|}{|1,2,3,4,5|} = \frac{3}{5}$$

3. The entries in the soft set $F_1$ equivalence classes are also not properly tabulated.

Given:

$$F_1(\{1,2,5\}) = \{\{1,2,5\},\{1,2\},\{1\}\},$$
$$F_1(\{3,4\}) = \{\{3,4\},\{4\}\},$$
$$F_1(\{1,2\}) = \{\{1,2\},\{1\}\},$$
$$F_1(\{3,5\}) = \{\{3,5\}\}$$
$$F_1(\{4\}) = \{\{4\}\},$$
$$F_1(\{1\}) = \{\{1\}\},$$
$$F_1(\{2,3,4,5\}) = \{\{3,4\},\{3,5\},\{4\},\{2,3,4,5\}\}.$$

**Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020** 

Table 6
The tabular representation of the soft set $(F_2, E)$.

| $U/E$ | $e_1$[1, 2, 5] | $e_2$[3, 4] | $e_3$[1, 2] | $e_4$[3, 5] | $e_5$[4] | $e_6$[1] | $e_7$[2, 3, 4, 5] |
|---|---|---|---|---|---|---|---|
| $x_1$[1, 2, 5] | 1 | 0 | 1 | 1 | 0 | 1 | 1 |
| $x_2$[3, 4] | 0 | 1 | 0 | 1 | 1 | 0 | 1 |
| $x_3$[1, 2] | 1 | 0 | 1 | 0 | 0 | 1 | 1 |
| $x_4$[3, 5] | 1 | 1 | 0 | 1 | 0 | 0 | 1 |
| $x_5$[4] | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| $x_6$[1] | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| $x_7$[2, 3, 4, 5] | 1 | 1 | 1 | 1 | 1 | 0 | 1 |

Corrected one:

TABLE V

| $U$ \| E | $e_1$[1,2,5] | $e_2$[3,4] | $e_3$[1,2] | $e_4$[3,5] | $e_5$[4] | $e_6$[1] | $e_7$[2,3,4,5] |
|---|---|---|---|---|---|---|---|
| $X_1$[1,2,5] | 1 | 0 | 1 | 1 | 0 | 1 | 1 |
| $X_2$[3,4] | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| $X_3$[1,2] | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| $X_4$[3,5] | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| $X_5$[4] | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| $X_6$[1] | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| $X_7$[2,3,4,5] | 0 | 1 | 0 | 1 | 1 | 0 | 1 |

***Experimental analysis:***

The new NSS algorithm has been verified and tested using benchmark datasets and also it gives better results compared to MDA. The execution time is improved in all the datasets verified and also number of instances and the number of attributes are also given.

A review on "A soft set approach for clustering student assessment datasets"

The approach of soft set theory in clustering the data is the methodology that is used in this paper. The concept of Maximum degree of domination in soft set theory (MDDS) is implemented in the student assessment datasets. In data clustering, the clustering attribute selection is the major task. This paper focuses on choosing the attribute selection using soft set theory. This paper is built with the application perspective.

***Methodology used:***

Multi valued attribute is taken as the input for processing and is been tabulated using soft set. Since the attribute has a multiple values in it, the soft set will be the multi-soft set. After the multi soft set is tabulated, the decomposition of soft set is made. Domination for the decomposed soft set is assigned and the degree of domination is calculated. From the degree of domination the maximum value degree is chosen as a best clustering attribute.

***Application:***

Using soft set, the approach is tested with data taken from various departments in Yogyakarta, Indonesia with respect to the various parameters like student name, race, age and the attendance. The domination value is separately tabulated for all the individual parameters and the maximum domination degree is calculated form the individual tables.

***Drawback:***

1) The individual table generation and the domination degree calculation are not detaily explained.
2) Parameter reduction (clustering attribute selection) is also not explained with an application directly.

## V. CONCLUSION

This paper combines all the different approaches of rough set based algorithms with its advantages and disadvantages along with fuzzy dominance and splitting attribute selection. Also, it is helpful for fresh researchers to start with fundamentals.

## REFERENCES

1. Hartama D., Yanto I.T.R. and M. Zarlis: "A soft set approach for fast clustering attribute selection", ICIC, 2016.
2. Mamat R., Herewan T. and Mat Deris M., "Maximum Attribute Relative of soft set for clustering attribute selection", Knowledge Based Systems, vol.52, 2013, pp.11-20.
3. Mazlack Lj. and Zhu Y.: "A rough set approach in choosing partitioning attributes", proceeding of ICSA 13th International conference, CAINE, 2000, pp.1-6.
4. Molodtsov D: "Soft Set Theory - First Results", Computers and Mathematics with Applications, vol.37, 1999, pp.19-31
5. Qin H., Ma X., Mohamad Zain J. and Herewan T., "A novel soft set in selecting clustering attribute", Knowledge-Based Systems, vol.36, 2012, pp.139-145.
6. Parmar D., Wu T. and Blackhurst J., "MMR: An algorithm for clustering categorical data using Rough Set Theory", Data and Knowledge Engineering, vol.63, 2007, pp.879-893.
7. Pawlak.Z: "On Rough sets", Bulletin of the European association for theoretical computer science, vol.24, 1984, pp. 94-109.
8. Pawlak.Z: "Rough classification", International journal of man machine studies, vol.20, 1984, pp. 469-483.
9. Prakash Kumar and Tripathy B.K., "MMeR: An algorithm for clustering heterogeneous data using rough set theory", Int. J. Rapid Manufacturing, 1, 2, 2009, pp.189-207.
10. Suhirman, Herewan T., Yanto I.T.R., Mohamad Zain J., Qin H. and Abdullah Z., "A soft set approach for clustering student assessment datasets", Journal of Computational and Theoretical Nanoscience, vol.12, 2015, pp.5928-5939. Vol.4, issue1, 2014, pp.2231-2307.
11. Sulekha G. and Sujit D.: "Fuzzy dominance matrix and its application in decision making problems", IJSCE,
12. Tripathy B.K and Ghosh A: "SDR: An algorithm for clustering categorical data using rough set theory", RAICS, IEEE, 2011, pp. 867-872.
13. Tripathy B.K and Ghosh A: SSDR: An algorithm for clustering categorical data using rough set theory, Advances in applied science research, 2(3), (2011), 314-326.
14. Tripathy B.K., Akarsh Goyal, Rahul Chowdhury and Patra anupam Sourav, MMeMeR: "An Algorithm for clustering Heterogeneous Data using Rough Set Theory", IJISA, vol.8, 2017, pp. 25-33.

15. Zadeh .L.A: “Fuzzy sets”, Information and Control, vol. 8, 1965, pp. 338-353

**Received: 02 Dec 2019 | Revised: 24 Jan 2020 | Accepted: 16 Feb 2020** 