ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Covering Based Refined Rough K-Means Algorithm.

P Prabhavathy*, BK Tripathy, and Venkatesan Meenakshi Sundaram.

School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India

ABSTRACT

Today large majority of researchers have focused on the research on accurate pattern mining methods. Clustering techniques are most preferable methods for pattern mining even in case of huge and growing data. Most of the real-life data applicable to pattern mining are non-crisp by nature. This means data can't be clearly separable into disjoint clusters. Thus uncertainty based models like rough sets are most preferable for representation of data under such circumstances. With rough set approach, every cluster is a rough set and hence is represented by upper and lower approximations. The data objects belong to the upper approximation may or may not belong to clusters and data objects which belong to lower approximation definitely belong to clusters. Refined rough K-Means algorithm was proposed by Peters in-order to improve the rough K-Means clustering algorithm proposed by Lingras. In this paper, we propose an extension of the refined rough K-Means algorithm of Peters to the context of covers. Experimental analysis shows that the proposed algorithm is more efficient than the refined rough K-means algorithm.

Keywords: Clustering, Rough Sets, Covers, K-Means.

Corresponding author

INTRODUCTION

Today data that computer systems process can be in any format such as text, numbers, images, audio, video, facts etc. The huge amount of data is getting accumulated from different organizations, which keeps on growing as time passes. This data can be operational or transactional data, non- operational data, and metadata. Thus information we retrieve from this data is nothing but relationship or association or patterns among all these data. Information is transformed into knowledge which is used for benefit of organization. For this purpose data mining techniques are used. Data mining is a computer based procedure of examining volumes of data and obtaining information from them. Data mining tools are used to drive this process in order to discover different useful as well as meaningful patterns and rules. Data mining involves several steps such as access and preparation of data, data mining and data analysis. Data mining consists of 6 important tasks [1] such as Clustering process, Anomaly detection, Association rule learning, Classification process, Regression process and Summarization process. Here, in this paper we are going to focus only on the Clustering process. Clustering process [1] is referred as a method of grouping set of data objects into different groups. It is used when there is no class information available for any data object in advance which is the most common in real-life situations. There are many Clustering methods available such as Partition Method, Hierarchical Method, Density-based Method, Grid-Based Method, Model-Based Method, and Constraint-based Method. We are going to discuss the algorithm which is extension of Partitioned based method. The Classical K-Means Clustering algorithm is an existing algorithm based upon the Partition Based algorithm which forms a base layer for the algorithm which is proposed in this paper. The Classical K-Means Clustering algorithm [8] is perhaps the most popular algorithm and has huge no of real-life applications. Some of the major advantages of K-Means are simplicity and computational as well as memory efficiency. K-Means clustering algorithm is widely applicable as well as most efficient on large data sets [19]. K-Means is basically an approach of partitioned based clustering technique [19]. In partition based clustering partitions are subsets of original data set and represents clusters. So, data set is divided into partitions which never overlap. Thus there exists no element common in any two or more partitions. Many researchers have extended this Classical K-Means in different directions and they have come with different results [14-17] [20] [40].

In real-life scenarios, we have to deal with imperfect knowledge. The imperfect knowledge is a big problem as well as now a day it is a vital issue for researchers in the field of Artificial Intelligence. For the purpose of understanding as well as manipulating imperfect knowledge, there exist many approaches. Fuzzy set based techniques form one of these ways. In 1965 this concept was introduced by Lotfi A. Zadeh and Dieter Kalua as an extended classical set theory. In classical set theory the membership of every element belonging to any set is measured in binary values where as in fuzzy set theory it is measured in real unit interval [0, 1]. It is considered as one of the most successful theories. Later, Pawlak proposed a new model to handle imperfect knowledge, known as Rough Sets [37], which is based upon the mathematical notion of equivalence relation. Now a days to handle indiscernibility of objects, the Rough Set Concept is considered as an excellent tool. Among different extensions of Classical Rough Sets, we have is the Covering Based Rough Sets. In 2006, Zhu and Wang [4-6] [12] had introduced four different kinds of covering based rough sets with different properties analyzed. Since then covering based rough sets are used in many directions like; topological properties [30] approximate equalities [28][29][31][32], database extension [21] [36], Classifications and multigranular rough sets [22-26] [33- 35].

In 2004, Lingras[4] proposed rough K-Means algorithm based on classic K-Means as well as Rough set notion. In contrast to classical K-Means where clusters are non-overlapping, Rough K-Means algorithm works with overlapping clusters. In 2006 George Peters proposed refinements in Lingras Rough K-Means Algorithm. This refined algorithm is known as Refined Rough K-Means algorithm [5], aimed to tackle with limitations of Rough K-Means algorithm. In this paper, we proposean extension of Refined Rough K-Means algorithm from the partition based rough sets to covering based rough sets.

The further organization of the paper is as follows. In section 2, the background of the work is unfolded, the literature survey in this direction of research is discussed in section 3 and section 4 contains the architecture of the proposed module. In section 5, we present the proposed algorithm. The experimental results are presented in section 6 along with their analysis. Finally, we state the conclusion on our work followed by reference of the materials consulted during the preparation of this piece of work.

BACKGROUND

Rough Set Theory:

The Classical K-Means algorithm gives non-overlapping or crisp clusters, but in many scenarios we require overlapping clusters. The rough sets are used to get overlapping clusters. The rough set theory is an extension of classical set theory. In fuzzy set theory, membership degree decides whether objects are similar or distinct, but in rough set theory objects are either definite or possible members of any cluster [9]. In rough sets, the membership [10] of any object is decided by other objects, but it never happens in case of fuzzy sets. In fuzzy sets, main problem is to interpret the membership degree of objects [10]. The Rough Clustering deals with some properties of rough set. But we require upper as well as lower approximations which follow some important rough set properties [9] as given below:

- An object w can be contained by at most one lower approximation which ensures that any two lower approximations never contain common objects.
- Every object w which is part of lower approximation of any set is also part of upper approximation of that set.
- If any object w is not contained in any lower approximation then it must be part of at least two upper approximations.

Covering Based Rough Sets:

Covering Based rough sets are based on covers and given as (U,C) where U is universe and C is covering set. The rough sets differs from covering based rough sets, where rough sets use partitions or equivalence classes instead of covers. The sets in the partition used by the rough sets are disjoint in nature where as it is not necessary to be always true for covering. In case of covering the data items can be part of multiple subsets. The covering based subsets generates upper and lower approximations by applying covering which exposes more rough, thus covering based rough sets are considered more powerful than rough sets [22].

Let V be universal data set, C be the covering for V, Y be a feature vector and Md(y) is the minimal description of data object y, then we can define different covering types as below:

First Type [12]:

$$\mathrm{CLA(Y)} = \bigcup\{L \in C \mid \mathrm{L\ is\ a\ subset\ of\ Y}\}$$
$$\mathrm{FH(Y) = CLA(Y)} \bigcup \{\mathrm{Md(y)} \mid \mathrm{y} \in \mathrm{Y \backslash CLA(Y)}\} \tag{1}$$

Second Type [12]:

$$\mathrm{CLA(Y)} = \bigcup\{L \in C \mid \mathrm{L\ is\ a\ subset\ of\ Y}\}$$
$$SH(Y) = \{L \mid L \in C \wedge L \bigcap Y \neq \phi\} \tag{2}$$

Refined Second Type [13]:

$$\mathrm{CLA(Y)} = \{\mathrm{y} \in \mathrm{V} \mid \bigcap \mathrm{Md(y)} \in \mathrm{Y}\}$$
$$\mathrm{CUA(Y)} = \{\mathrm{y} \in \mathrm{V} \mid (\bigcap \mathrm{Md(y)}) \bigcap \mathrm{Y} \neq \phi\} \tag{3}$$

Third Type [12]:

$$\mathrm{CLA(Y)} = \bigcup\{L \in C \mid \mathrm{L\ is\ a\ subset\ of\ Y}\}$$
$$\mathrm{TH(Y)} = \bigcup\{\mathrm{Md(y)} \mid \mathrm{y} \in \mathrm{Y}\} \tag{4}$$

LITERATURE SURVEY

In 1967, James Mac Queen introduced the concept of “K-means” [14] where as in 1957 Stuart Lyod proposed the basic K-Means algorithm. There are many benefits [13] of K-Means algorithm including flexibility, simplicity, easy to understand and implement, whereas the selection of initial centroid affects the performance of algorithm.

In 2008, Tripathy and Tripathy B.K focused on extraction of knowledge from huge databases [29] and also elaborated on some applications of rough set in large databases. In 2004, Bhatia [15] presented Adaptive K-Means Clustering Algorithm, which is based on cluster rearrangement in order to get better partition when each time new element comes into picture. In 2010, Madhu and Shrinivas proposed Enhanced K-Means clustering algorithm with improved initial centre algorithm [16] which is a new method based on exploring better initial centroids as well as time efficient way of allocating data points to clusters. Rauf and Sheeba, proposed in 2012 an Enhanced K-Mean Clustering Algorithm [15] which is based on calculation of initial centroid in order to reduce number of iterations and time complexity. In 2013, Zhang and Fang introduced an Improved K-Means clustering algorithm [17] based on improved initial focal point and K value. In order to scale with big data sets Esteves, Hacker and Rong introduced Competitive K-Means clustering algorithm [20], which has improved cluster analysis accuracy as well as has decreased the variance. In 2014, Zhang, Chen and Ma came with an algorithm based on rough K-Means clustering with variable weighted distance measure [17] between data objects and cluster centres in order to improve performance. Bhoomi proposed an Enhanced K-Means clustering algorithm [16][38][39] based on the changes in existing K-Means algorithm in order to decrease time complexity for numeric data. In 2016, Zhang and Ma have introduced an improved rough k-means clustering algorithm which has used weighted distance measure with Gaussian function.

Tripathy and Mitra [30] discussed some topological properties of covering based rough sets. And later in 2009, Tripathy and Tripathy [32] he established Covering Based Rough Equivalence of Sets and Comparison of Knowledge. In 2015 Tripathy and Govindarajulu [22] [35]explained about Multigranular computing based on covering based rough sets and their properties are studied. Prabhavathy and Tripathy [23][24][25] studied covering based approach in sequential data. In their work they applied it in hierarchical method of clustering. We got our motivation from the above works to work on covering based K means.

In this paper we have demonstrated that, how covers can improve the performance of the refined rough K-Means algorithm, so that we can get better results than the original algorithm

OVERALL ARCHITECTURE

The overall structure of the proposed algorithm can be presented in the following five steps.

Step 1: Input dataset is taken for the processing. This input dataset must be numerical i.e. non-categorical data. If categorical attribute is present in input dataset then the corresponding attribute values can be transformed into numerical representations.
Step 2: In this stage, data pre-processing is performed on input data. This pre-processing includes data cleaning, binning strategies, normalization and transformation. This pre-processing makes our input data into intended format needed for processing.
Step 3: This pre-processed data is given as an input to feature extraction stage. At the point when the information is too huge to ever be processed and it is suspected to be repetitive, then it can be converted into a reduced feature set. The feature extractions techniques include PCA, kernel PCA, Non-linear dimensionality reduction etc.
Step 4: Finally we apply Covering based refined rough k-means algorithm over this input data or extracted features. Depends on the input parameter algorithm divides data into overlapping sets called “clusters”.
Step 5: The final output of algorithm is clusters which represents knowledge.

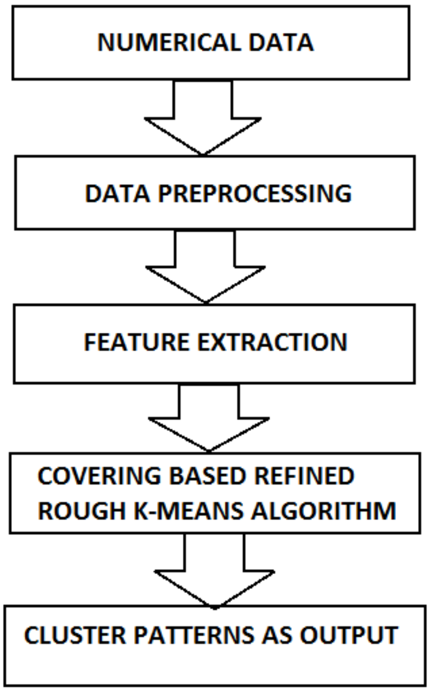

Fig 1: Proposed Architecture of Clustering

ALGORITHM

The proposed algorithm is based on refinements suggested by Peters [8], in Rough K-Means clustering algorithms introduced by Lingras as well as rough covering.

In classical set theory, elements of the set uniquely determines set itself. Hence it follows crisp notion. Ex: set of odd numbers. In real-time scenarios, we generally come across vague notions, which mean we can't partition objects into two classes. Ex: Classifying paintings on basis of beauty. Thus Rough Set Theory was proposed. If we want to classify elements as set X w.r.t. any indiscernibility relation R then we will get:

1. Lower Approximation of X w.r.t. R
2. Upper Approximation of X w.r.t. R
3. Boundary Region of set X.

Pawlak [8] has defined several properties for rough sets. Some of them apply to clustering.

1. Any data element can belong to at most one lower approximation.
2. If any data element belongs to a certain lower approximation then it also belongs to corresponding upper approximation also.
3. If data element is not part of any lower approximation then it is part of two or more upper approximations.

Following is the proposed algorithm for covering based refined rough K-Means:

Step 1: Let W be the universal data set. Initialize with data objects.
Step 2: Let V be the covering set i.e. nonempty subset of W. Initialize the covering set V. Here,$\cup\, v_i = W$, where $v_i \in V$.
Step 3: Let M be number of clusters, C_i be i^{th} cluster and Yi be subset of W where Yi corresponds to cluster ci. Initialize each Yi where $1 \le i \le M$.
Step 4: For y ϵ W, calculate minimum description for y as below:

ISSN: 0975-8585

$$Md(y) = \left\{ L \in V | y \in L \wedge \begin{pmatrix} \square S \in V \wedge y \in S \wedge S \subseteq L \\ \Rightarrow \\ S = L \end{pmatrix} \right\} \quad (5)$$

Step 5: Let CLA_i be covering lower approximation and CUA_i be covering upper approximation for cluster C_i. Calculate CLA_i and CUA_i for each cluster C_i. There exist four different types of rough clustering which are first type, second type, third type and enhance second type clustering. Our observation suggests to employ refined second type or third type covering along with refined rough K-Means clustering algorithm. So here we can calculate CLA_i and CUA_i using refined second covering type as given below:

$$\mathrm{CLA(Y)} = \{y \in V | \bigcap Md(y) \in Y\}$$
$$CUA(Y) = \{y \in V \,|\, (\bigcap Md(y)) \bigcap Y \neq \phi\} \quad (6)$$

[1] Let m_i be mean for cluster C_i. Calculate new means using following equation:

$$\overrightarrow{m_k} = Q_L * \sum_{\overrightarrow{y_n} \in CLA_k} \frac{\overrightarrow{y_n}}{|CLA_k|} + Q_U * \sum_{\overrightarrow{y_n} \in CUA_k} \frac{\overrightarrow{y_n}}{|CUA_k|} \quad (7)$$

Where, $Q_L + Q_U = 1$

[2] For every cluster C_i assign a data element to it which best represents that cluster.

a) Assign this data element to lower and upper approximation of this cluster.

I. Evaluate the least distance between every cluster C_i to every data element $\overrightarrow{y_L}$. Assign dataelement $\overrightarrow{y_L}$ to lower as well as upper approximation of the cluster C_H.

$$dist(\overrightarrow{y_L}, \overrightarrow{m_H}) = \min_{n,k} dist(\overrightarrow{y_n}, \overrightarrow{m_k}) = \overrightarrow{y_L} \in \overrightarrow{CLA_k} \text{ and } \overrightarrow{y_L} \in \overrightarrow{CUA_k} \quad (8)$$

II. Ignore $\overrightarrow{y_L}$ and $\overrightarrow{m_H}$. Repeat step[I] for those clusters left for processing in step[I]. If there exists no clusters to be processed in step [I] then go to step (b).

b) Identify the nearest mean $\overrightarrow{m_H}$ for each of data point $\overrightarrow{y'_m}$ which is not selected in step (a).

$$d_{m,h}^{min} = dist\left(\overrightarrow{y'_m}, \overrightarrow{m_H}\right) = \min_{k=1 \dots K} dist\left(\overrightarrow{y'_m}, \overrightarrow{m_k}\right) \quad (9)$$

Allocate $\overrightarrow{y'_m}$ to CUA_H.

c) Find out the means $\overrightarrow{m_t}$ which are too nearer to $\overrightarrow{y'_m}$.

$$F' = \left\{ t : \frac{dist(\overrightarrow{y'_m}, \overrightarrow{m_k})}{dist(\overrightarrow{y'_m}, \overrightarrow{m_H})} \leq \delta \text{ and } h \neq k \right\} \quad (10)$$

Where δ is a given relative threshold.

If $F' \neq \emptyset$ then allocate $\overrightarrow{y'_m}$ to CUA_t , else allocate $\overrightarrow{y'_m}$ to CLA_t.

Check convergence of algorithm**.** If algorithm converges then stop else go to step (6).

RESULTS

To endorse the effectiveness of the covering based refined rough K-Means clustering algorithm, this paper uses a Total 15-64 labour to population (%) data set, Employment Time Series data set and Household Income Disparity data set for experiment. The total 65+ labour to population data describes the age group above 65, percentage of all labour to total population. The Household Income Disparity represents different measures of disparity for 1990 and 2000.The variables for 1990 are based on data values of 1990, but aligned to the boundaries of U.S. Counties in 2000. Employment Time Series dataset represents key labour force measured for male and female employees with respect to several different measures. All attributes of each of the datasets are numerical in nature. The dataset name, number of attributes and number of data objects are given below in table 1.

Dataset Name	No of Attributes	No of Data Objects
Total 65+ labour to population (%)	29	189
Employee Time Series	24	90
Household Income Disparity	16	3110

We have measured the performance of our clustering algorithm by a well-known performance metric called Davies Bound Index. The Davies Bound index is the one of the most popular metrics to measure performance of clustering algorithms. Davies Bound Index represents the mean of resemblance between every cluster to respective most identical cluster. The Optimal clustering techniques tries to minimize Davies Bound index value. Suppose C_j represents cluster of data objects, Z_j represents size of the cluster 'j' i.e. no of elements that forms cluster 'j', X_i denotes the data object 'i' with N dimensions and E_j represents centre of cluster 'j'. Now we can calculate the value S_j denoting scatteredness of the data objects inside the cluster j as below:

$$S_j = \frac{1}{z_j}\sum_{i=1}^{z_j}\left\|X_i - E_j\right\|_q \qquad (11)$$

We have considered the value for 'q' as 2 in order to evaluate Euclidian distance. Let us consider value of $H_{a,b}$indicates how cluster C_a and cluster C_b are separated which should be as large as possible and it can be expressed as below:

$$H_{a,b} = \left\|E_a - E_b\right\|_q$$

$$H_{a,b} = \left(\sum_{i=1}^{n}\left|e_{i,a} - e_{i,b}\right|^q\right)^{\frac{1}{q}} \qquad (12)$$

Here data element $e_{i,a}$ indicates i^{th} element of clusterC_a. Suppose $R_{a,b}$ be a value indicates wellness of the clustering technique used. The value of $R_{a,b}$ can be calculated as below:

$$R_{a,b} = \frac{S_a + S_b}{H_{a,b}}$$

The value of $R_{a,b}$defines a new value D_a as below:

$$G_a \equiv \max_{a \neq b} R_{a,b}$$

Consider we have F no of clusters with our dataset. With help of this, we can define the Davis Bound Index value as below:

$$DB = \frac{1}{F}\sum_{k=1}^{F} G_k$$

The value of DB must be as small as possible.

We have tested refined rough K-Means clustering with all four covering types and came with several observations. The experiments we performed has shown that the application of refined second type covering as well as application of third type covering when applied with refined rough K-Means have provided lower Davies Bound index values as compared to the Davies Bound index values obtained from original refined rough K-Means algorithm. We have tested each of above data sets with different cover sets as well as different values ofQ_L,δ, and Q_U with each of covering techniques. We found that most of the times refined rough K-Means with refined second type covering or third type covering gives better clustering performance. We observed that the application of refined second type covering as well as third type covering provides better allocation of data elements to clusters which results into performance improvements. The performance results of refined second type clustering and third type clustering with refined rough k-means algorithm are almost similar in each test. We have considered Q_L value to be 0.45.

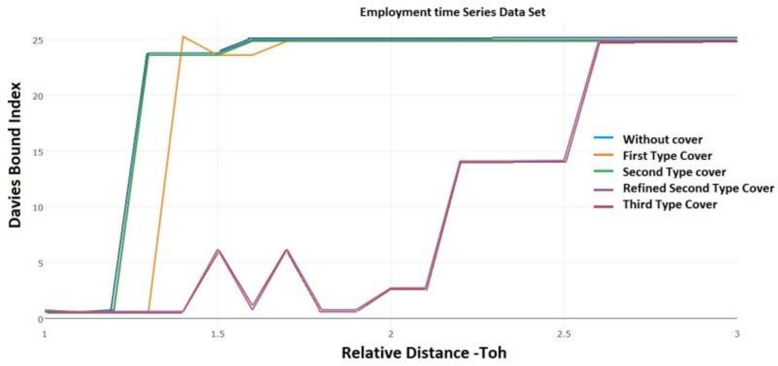

Figure-(a)

The graph in figure (a) shows comparison among normal refined rough K-Means and covering based refined rough K-Means algorithm with respect to employee time series data set. The x axis represents relative distance δ and Y axis represents Davies Bound index. It shows clearly that refined second type as well as third type has lower values of Davies Bound Index. Thus performance has improved.

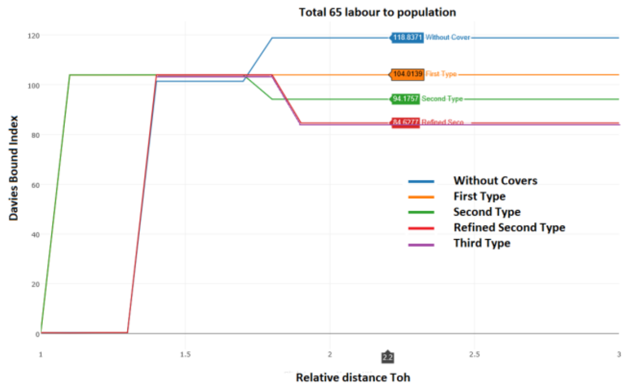

Figure-(b)

The graph in figure (b) shows comparison among normal refined rough K-Means and covering based refined rough K-Means algorithm with respect to total 65 labours to population dataset. It shows how refined rough K-Means gives better performance with refined second type covering.

The figure (c) indicates performance for experiments performed over Household Income Disparity dataset. It shows performance of refined second type covering or third type covering with refined rough K-Means algorithm is most of the times better than normal refined rough K-Means as well as Refined rough K-Means with firsts type or second type covering when relative distance (δ) less than 25.

Figure-(c)

Every time we have randomly generated the cover, feature vectors. The number of clusters as well as the relative distance δ needs to be manually defined. We have considered Q_L value to be 0.45. This random arrangement helps to eliminate the worries regarding the consideration of the values of cover and feature set.

CONCLUSION

The covering technique can be used as a choice for selecting initial allocation of data points to clusters. This initial allocation of data points to cluster in many cases improves the performance of clustering as we have shown in our results. The performance of covering based rough clustering heavily depends on the selection of covers and values of Xi. So, generating and exploring optimal cover as well as value of X_i is a crucial task. The value of Q_L, Q_U and δ also affects the performance. So we need to carefully select these values.

REFERENCES

[1] Kameshwaran K, Malarvizhi K. International Journal of computer Science and Information Technology 2014; Vol. 5(2): 2272-2276.

[2] Laurence M, Sylvain C. Tutorials in Quantitative Methods for Psychology 2013; Vol. 9(1): 15-24.

[3] Mohanty D. International Journal of Computer Science and Research 2010; Vol. 1 Issue 1: 43-49.

[4] Zhu W, Wang F Y. Sixth IEEE International Conference on Data-Mining Workshop, (ICDMW'06), March – 2006;

[5] Zhu W. IEEE Int'l Conf. Web Intelligence (WI '06) 2006; 494-497.

[6] Zhu W. International Conference on Hybrid Intelligent Systems (HIS'06) December 2006; 13-15: 43.

[7] Lingras P, West C. Intelligent Information Systems 2004; 23:5–16.

[8] Peters G. Pattern Recognition 2006; 39: 1481 – 1491.

[9] Lingras P, Peters G. Advanced Information and Knowledge Processing, Springer-Verlag London Limited 2012; 23-37.

[10] Yao Y Y. Information Sciences 1998; Vol. 109, No. 1-4:227-242.

[11] Shi Z Gong Z. Information Sciences 2010; 180:.3745–3763.

[12] Zhu W, Wang F Y.IEEE Transactions on Knowledge And Data Engineering 2007;Vol.19 no. 8:1131-1144.

[13] Kanungo T,Nathan S,Netanyahu Angela. IEEE Transactions On Pattern Analysis And Machine Intelligence July 2002;Vol. 24, No. 7:881-892.

[14] Zhang C, Fang Z. Journal of Information & Computational Science 2013; 10: 1: 193–199.

[15] Rauf A, Sheeba Saeed M ,Shad K, Huma, J. Middle-East Journal of Scientific Research 2012;12 (7):.959-963.

[16] Bangoria B M. International Journal of Computer Science and Information Technologies 2014; Vol. 5 (1): 876-879.

[17] Zhang T, Long C. F. M. International Journal of Database Theory and Application 2014; Vol.7,No.6:163-174.

[18] Sai Reddy S.Y, Manoj C and Lakkshmanan, K. A. International journal Computer Science and Software Engineering 2013;3(7):743-746.

[19] Qiao J, Zhang Y. Proceedings of the IEEE Conference 2015;CAC:51-54.

[20] Esteves R M, Hacker T, Rong, C.IEEE International Conference on Cloud Computing Technology and Science 2013; 17-23.

[21] Tripathy B.K, Patro V.M.1st IIMA International Conference on Advanced Data Analysis Business Analytics 2009; 37-38.

[22] Tripathy B.K, Govindarajulu K. Information system design and applications Advances in Intelligent Systems and Computing Springer 2015;Vol.339:555–564.

[23] Prabhavathy P, Tripathy B.K. International Journal of Intelligent Information Technologies (IJIIT) 2016; Vol.12, No.2:

[24] Prabhavathy P, Tripathy B.K. Int. J. Reasoning-based Intelligent Systems 2015;Vol. 7, Nos. 3/4: 296- 304.

[25] Prabhavathy P, Tripathy B.K.International Journal of Computer Science and Business Informatics May2013; Vol. 1, No. 1: 1-15.

[26] Tripathy B.K., Nagaraj M. I.J. Information Technology and Computer Science 2015; 08: 61-67.

[27] Tripathy B.K, Nagaraju M.IIOAB journal, Special issue on Applications of Soft Computational Approaches

ISSN: 0975-8585

in Bioinformatics 2015; vol.6, issue 4:77-97.

[28] Tripathy B.K, Parida S. C. International Journal of Information Technology and Computer Science 2015;

[29] Tripathy B.K, Parida S. C. Accepted for publication, IJISA, MECS publishers, 2015;

[30] Tripathy B.K., Mitra,A. Proceedings of the IEEE – International conf. on Emerging Trends in Computing (ICETiC-2009) Virudh nagar INDIA 2009; 8-10:263-268.

[31] Tripathy B.K, Panda G. K., Mitra A. Proceedings of the Inter. Conf. in Mathematics and Computer Science Chennai, INDIA ICMCS 2009; Vol.2: 438-443.

[32] Tripathy B.K, Tripathy H.K.Proceedings of the IACSIT Spring Conference Singapore April 2009; 17-20:303-307.

[33] Tripathy B.K, Panda G.K.IEA/AIE 2009; LNAI 557:777-786.

[34] Tripathy B.K, Panda G.K.ACM proceedings of ICCIT2009 Seoul, South Korea Nov 2009; 24-26:1227-1232.

[35] Tripathy B.K, Govindarajulu K. Sixth International Conference on Computational Intelligence and Communication Networks 2014; 978-1-4799, 6929-6/14: 708 -713.

[36] Tripathy B.K, Kalia L. P. Proceedings of the UGC sponsored national conference 2012; 30-31:39-46.

[37] Pawlak Z. International Journal of Information and Computer Sciences 1982; 11:341-356.

[38] Venkatesan M, Arunkumar Thangavelu A. International Journal of Computer Applications in technology Inderscience Publisher 2013;Vol.48,No.2:144-154.

[39] Venkatesan M, Arunkumar Thnagavelu. WIREs Data Mining and Knowledge Discovery Wiley Publisher 2015;Vol.5: 142-154.

[40] Venkatesan M, Arunkumar Thangavelu, Prabhavathy P. Int. J. Communication Networks and Distributed Systems Inderscience 2015; Vol.15: 191-211.