ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Sequential Clustering: A Study on Covering Based Rough Set Theory.

Prabhavathy P*, and Tripathy BK

*School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India
School of Computing Science and Engineering, Vellore Institute of Technology, Vellore, India

ABSTRACT

Sequential data analysis is one of the vital research area. Several data mining techniques like classification, association, predictions can be applied in sequential data. Clustering is a challenging task in the field of machine learning, pattern recognition and web mining. Clustering is the process of grouping data based on some similarities but applying clustering approach in sequential data should focus on order as well as the content of sequence similarity. Rough set theory is one of the efficient soft computing techniques used in clustering which help researchers to discover overlapping clusters in many applications such as web mining and text mining. The rough set which holds equivalence relation is very rigid as it doesn't support incomplete information system. This leads the theory's application to a certain extent. Hence covering based rough set is introduced where the partitions of a universe are replaced by covers. Different types of covering based rough set theory exist in the literature. In this paper all three types covering rough sets is investigated and produced a comparative study of first type, second type, and third type covering based rough clustering algorithm for sequential data.

Keywords: Cover, Clustering, Rough Set, Sequential, Approximation

Corresponding author

INTRODUCTION

Data plays a significant role in the today scenario. We can identify vagueness, incompleteness, and granularity in an information systems as the data are collected using sensors, cameras and smart devices. This leads in production of unreliable solution in the data analysis. Managing these data is a challenging task for the users as well as researchers. An innovative technique into the field of uncertainty and knowledge discovery is based on Rough set theory, which provides a framework for the representation of uncertainty. Broadly we classify Clustering algorithms based on its various important issues such as algorithmic structure, nature of clusters formed, use of feature sets, etc. [4].In this paper its focused on hierarchical algorithms where it doesn't require the number of clusters, k, as an input parameter. However, a termination condition has to be defined indicating when the merger or division process should end. Finding an efficient termination condition for the merging or division process is still an open research problem [21].Clustering approaches can be classified into two types one is hard clustering which is our conventional clustering method where the objects that are similar will form in one cluster and dissimilar will be in other cluster. They are disjoint in nature. In soft clustering, an object may be a member of two or more clusters. Soft clusters usually have fuzzy or rough boundaries [20]. In fuzzy clustering, each object is characterized by partial membership whereas in rough clustering objects are characterized using the concept of a boundary region. The main features of rough set is unlike other techniques, rough set theory does not require any prior information about the data such as apriori probability in statistics and a membership function in fuzzy set theory. The classical rough set theory is based on equivalence relations, but this requirement is not satisfied in some situations of real world data.

Rough set theory is a valuable tool for data mining. In the past few years the concept of basic rough sets has been extended in many different directions. The original rough set theory proposed by Pawlak [*1,2*] is based upon equivalence relations defined over a universe. It is the simplest formalization of indiscernibility. However, it cannot deal with a number of granularity problems in real information systems which has directed to numerous significant and motivating extensions of the original concept. Generalization of rough set theory can be followed by several approaches like set-theoretical frame work, covering based rough set and subsystem based methods. One of the most recent generalizations is the notion of covering based rough sets, introduced by Zakowski [3].A cover is a generalization of the notion of partition. The covering based rough sets are models with promising potential for applications to data mining. Several properties of the different types of covering based rough sets have been derived by different researchers [5, 6, 7].

This paper focuses on comparative study of various types of covering based rough set approach in clustering sequential data. The remainder of this paper is organized as follows. The related work is discussed in section2. Section 3 defines the fundamental concepts and properties of rough set theory and section 4 defines all three types of covering based rough set theory. In Section 5, Covering based rough set clustering for sequential data is presented and compared with various types of covering based rough set. This paper concludes with future work in section 6.

RELATED WORK

Most of our traditional tools for formal modelling, reasoning and computing are crisp, deterministic and precise in character. Real situations are very often not crisp and deterministic and they cannot be described precisely. Rough Sets, introduced by Pawlak [*1,2*] has been found to be an excellent tool to model and study impreciseness in data. A rough set[*3*] is represented by a pair of crisp sets, called the lower and upper approximations of the set. The lower approximation of a rough set comprises of those elements of the universe, which can be said to belong to it definitely and the upper approximation comprises of those elements which are possibly in the set with respect to the available knowledge. Three types of rough or approximate equalities have been introduced by Novotny and Pawlak [9, 10, and 11]. The concepts of approximate equalities of sets refer to the topological structure of the compared sets but not the elements they consist of. Thus sets having significantly different elements may be rough equal.

Due to restriction in equivalence relation, application of rough set over different type of problems seems difficult, unless the clustering of problem appears to hold true for equivalence relation. The equivalence relations of rough sets were extended to generalized binary relations in several directions. Similarly, partition of universe in rough sets was extended to a covering [13]. The covering based rough sets are models with promising potential for applications to data mining. Like rough set theory Uncertainty characterization of covering in the covering approximation space, similarity measure between two covering rough sets and two generalized covering-based rough set models [14] and their properties and applications were presented. A framework for the study of covering based rough set approximations [15] was proposed. Three equivalent formulations of the classical rough sets are examined by using equivalence relations, partitions, and σ-algebras, respectively. A type of generalized rough sets based on covering and the relationship between this type of covering- based rough sets and the generalized rough sets based on binary relation [16] were studied.

In some areas, like biology, logs analysis, anomaly detection, natural language processing and telecommunications, data can be seen in the form of sequences. Various approaches were available for dealing with comparing sequential data especially sequential patterns. The two main similarity measures used for item set sequences are Edit distance and LCS [17]. The Edit distance was used for extracting sequential patterns under similarity constraints. Edit distance is not adaptable to the various definitions of similarity. Since Edit distance's operators are applied on the elements of sequence (i.e. item sets), an item set in a sequential pattern is reduced to an event type. Hence, a sequential pattern is treated as an event type sequence. The LCS measure (Longest Common Subsequence) is used for the comparison of sequences. The LCS gives the length of the longest common subsequence of two sequences. It is possible to use LCS to compare the similarity of sequential patterns without being optimal. We note three reasons why the LCS is not a optimal measure for sequential patterns (itemset sequences). Firstly, LCS does not take the position of itemsets (in order of sequence) into account in the two sequences. Secondly, LCS does not consider the length of the part which is not common. Thirdly, the number of different items in itemsets (in which the subsequence appears) does not affect the value of LCS. A S2MP (Similarity Measure for Sequential Patterns) similarity measure [17]was defined for computing the similarity between of sequential patterns. which takes the characteristics and the semantics of sequential patterns into account. This measure compares two sequential patterns both at the level of item sets and their positions in the sequences and also at the level of items in item sets.

Clustering algorithms have been classified using different taxonomies based on various important issues such as algorithmic structure, nature of clusters formed, use of feature sets, etc[19] . Clusters can be hard or soft in nature. In conventional clustering, objects that are similar are allocated to the same cluster while objects that differ significantly are put in different clusters. These clusters are disjoint and are called hard clusters. In soft clustering [20], an object may be a member of two or more clusters. Soft clusters may have fuzzy or rough boundaries. A rough cluster is defined in a similar manner to a rough set. The lower approximation of a rough cluster contains objects that only belong to that cluster. The upper approximation of a rough cluster contains objects in the cluster which are also members of other clusters. The advantage of using rough sets is that, unlike other techniques, rough set theory does not require any prior information about the data such as apriori probability in statistics and a membership function in fuzzy set theory. Joshi and Krishnapuram [21] argued that the clustering operation in many applications involves modeling an unknown number of overlapping sets, that is, the clusters do not necessarily have crisp boundaries. Web mining is one such area where overlapping clusters are required. Generally, clustering algorithms make use of either distance functions or similarity functions for comparing pairs of sequences [18]. Many of the metrics for sequences do not fully qualify as being metrics due to one or more reasons.

Basic Concepts of Rough Set Theory

Definition 1: Rough set:

Let U be the universe and let $\underline{R} \subseteq U \times U$ be an equivalence relation on U, called an indiscernibility relation. The pair $A = (U, R)$ is called an approximation space. The lower and upper approximation of set X with respect to R can be written as

$$\underline{R}(X) = \{x \in U : [x]_R \subseteq X\}$$
$$\overline{R}(X) = \{x \in U : [x]_R \cap X \neq \phi\}$$

Where $[x]_R = \{y \in U \mid xRy\}$ is the equivalence class of x. if $[x]_R \subseteq X$, Then it is certain that $x \in X$. if $[x]_R \subseteq X - \overline{R}(X)$ then it is clear that $x \not\subset X$. $[x]_R \subseteq X$ is called rough with respect to R iff $\underline{R}(X) \neq \overline{R}X$ .Otherwise X is R-discernible. figure:1 shows all the approximation space of rough sets

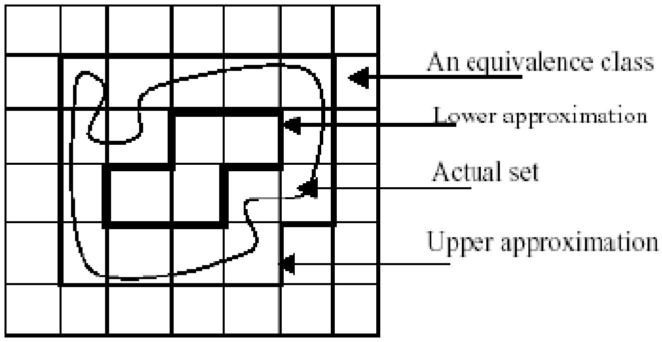

Figure.1 Approximation space of rough sets

Rough approximation satisfies the following properties

$R*(X) \subseteq X \subseteq R*(X)$,
$R*(\emptyset) = R*(\emptyset) = \emptyset; R*(U) = R*(U) = U$,
$R*(X \cup Y) = R*(X) \cup R*(Y)$,
$R*(X \cap Y) = R*(X) \cap R*(Y)$,
$R*(X \cup Y) \supseteq R*(X) \cup R*(Y)$,
$R*(X \cap Y) \subseteq R*(X) \cap R*(Y)$,
$X \subseteq Y \rightarrow R*(X) \subseteq R*(Y) \& R*(X) \subseteq R*(Y)$,
$R*(-X) = -R*(X)$,
$R*(-X) = -R*(X)$,
$R*R*(X) = R*R*(X) = R*(X)$,
$R*R*(X) = R*R*(X) = R*(X)$.

A rough cluster is defined in a similar manner to a rough set that is with a lower and upper approximation. The lower approximation of a rough cluster contains objects that only belong to that cluster. The upper approximation of a rough cluster contains objects in the cluster which are also members of other clusters.

Covering Based Rough Set Theory

According to Hu, Lin and Han [23], Rough sets theory uses the strict set inclusion definition to define the lower approximation, which does not consider the statistical distribution/noise of the data in the equivalence class This drawback of the original rough set model has limited its applications in domains where data tends to be noisy or dirty. The other drawback of rough set theory is the inefficiency in computation, which limits its suitability for large data sets in real-world applications. This has direct to numerous significant and motivating extensions of the original concept. One of the most recent generalizations is the notion of covering based rough sets, introduced by Zakowski.A cover is a generalization of the notion of partition. The covering based rough sets are models with promising potential for applications to data mining.

In recent past W. Zhu and F.Y. Wang have proposed four types of covering rough sets in which only one lower approximation and four different versions of upper approximations for such rough sets, several properties of these different types of covering rough sets are derived and analysed [13]. Covering based rough set extends a partition in rough sets to covering of the universe. A covering does not result from a rigid equivalence relation. Again, it enlarges the boundary set between lower and upper approximation. When compared, covering was observed to be more consistent with reality than partition for analysing and clustering objects. There exist six types of covering based rough sets [15]. The lower approximation for covering based rough set for first, second, third, fourth and fifth type are same, whereas the upper approximations are different. The upper approximation for fifth type of covering based rough set is defined using the definition of covering with respect to neighbourhood. For sixth type of covering, both the approximation are defined over the property of covering with respect to neighbourhood The definitions and types of covering based rough set, are as follows

Relationship between Three Types of Covering Based Rough Sets

Let U be a universe of discourse, $C = \{X \mid X \subseteq U\}$ a family of subsets of U, If no element of C is empty, and $\bigcup_{X \in C} X = U$, then C is called a covering of U. The ordered pair (U,C) a covering approximation space(14).
It is clear that a partition of U is certainly a covering of U, so the concept of covering is an extension of the concept of a partition.

Definition 2: Let (U,C) be a covering approximation space, $x \in U$, then set family is called the minimal description of x.

$$Md(x) = \{K \in C \mid x \in K \wedge (\forall S \in C \wedge x \in S \wedge S \subseteq K \Rightarrow K = S)\}$$

Definition 3:First Type covering based rough sets

$$\underline{C1}(X) = \bigcup\{K \in C \mid K \subseteq X\}$$
$$\overline{C1}(X) = C1(X) \bigcup (\bigcup\{\bigcup Md(x) \mid x \in X - C1(X)\})$$

Definition 4: Second Type covering based rough sets

Let(U,C) be a covering approximation space. For any $X \subseteq U$, the lower and upper approximations of X with respect to covering approximation space(U,C) are defined as follows:

$$\underline{C2}(X) = \{x \in U \mid (\bigcap Md(x)) \subseteq X\}$$

$$\overline{C2}(X) = \{x \in U \mid (\bigcap Md(x)) \bigcap X \neq \phi\}$$

If $\underline{C2}(X) = \overline{C2}(X)$, then X is said to be exact with respect to covering approximation space(U,C). Otherwise X is said to be covering rough set with respect to (U,C).

Definition 5: Third Type covering based rough sets

$$\underline{C3}(X) = \bigcup\{K \in C \mid K \subseteq X\}$$
$$\overline{C3}(X) = \bigcup\{\bigcup Md(x) \mid x \in X\}$$

Since user behaviour analysis is one of the important applications in clustering sequential data. Based on the above definitions First type, second type and third type covering based rough sets are applied in the application and experimented the results with msnbc web data.

Covering Based Rough Set Clustering

Vagueness in data has attracted mathematicians, philosophers, logicians and recently computer scientists. Rough set theory is an approach to deal with vagueness. In many data mining applications, the class attributes of most objects are not distinct but vague. Rough set theory is applied in various clustering algorithm . A tolerance based rough set model has been considered for document clustering and information retrieval. In general, similarity relations do not give the same kind of partitions of the universe as the indiscernibility relation. Similarity classes of each object x present in the universe U provide the similarity information for the object. An object from one similarity class may be similar to objects from other similarity classes. Therefore the basic granule of knowledge is intermixed. Extending indiscernibility to similarity relation requires weakening of some of the properties of binary relations in terms of reflexivity, symmetry and transitivity [24].

Definitions of lower and upper approximations of a set can now be easily formulated using tolerance classes. In order to do this, we substitute tolerance classes for indiscernibility classes in the basic definition of lower and upper approximations of set. Thus, the tolerance approximations of a given subset X of the universe U is defined as in definition 6.

Definition 6: Let $X \subset U$ and a binary tolerance relation R is defined on U. The lower approximation of X, denoted by $\underline{R}\ (X)$and the upper approximation of X denoted by $\overline{R}\ (X)$ are respectively defined as follows:

$$\underline{R}\ (X) = \{x \in X, R(x) \subseteq X\} \text{ and}$$

$$\overline{R}\ (X) = \bigcup_{x \in X} R(x)$$

In this paper a comparative study is made on sequential clustering algorithm using covering based rough sets for clustering web user transactions.

Let $x_i \in U$ be a user transaction consisting of sequence of web page visits. For clustering user transactions, initially each transaction is taken as a single cluster. Let the i[th] cluster be $Cl_i = \{x_i\}$. Clearly, Cl_i is a subset of U. The covering based upper approximation of Cl_i, denoted as $\overline{ClR}(C_i)$,is a set of transactions similar to x_i, that is, a user visiting the web pages in x_i may also visit other web pages present in the transactions belonging to $\overline{ClR}(C_i)$.

For any non-negative threshold value $\delta \in [0,1]$ and for any two objects $x, y \in U$, a binary relation τ on U denoted as $x\ \tau\ y$ is defined by $x\ \tau\ y$ iff $Sim(x, y) \geq \delta$. This relation R is a tolerance relation and R is both reflexive and symmetric but transitivity may not always hold.

The first upper approximation $\overline{ClR}(x_i)$ is a set of objects that are most similar to x_i. Thus, first upper approximation of an object x_i can be defined as follows:

Definition 7: For a given non-negative threshold value $\delta \in [0,1]$ and a set $X = \{x_1, x_2,,,, x_n\}$, $X \subseteq U$ the first upper approximation is

$$\overline{ClR}(\{x_i\}) = \{x_j \mid Sim(x_i, x_j) \geq \delta\}$$

Some sets in the collection resulting from the first upper approximation may share elements (the so called boundary elements). The boundary elements can guide the clustering process. The shared elements, generated after first upper approximation, may be the potential candidate of the new collection formed in the second or higher upper approximations. This can be decided by calculating the strength of shared element to all the clusters it belong.This is measured using a parameter called relative similarity. The value of the second and the higher similarity upper approximations is computed under the condition of relative similarity. For two intersecting sets $X, Y \in U$

The relative similarity of X with respect to Y is given by

$$\operatorname{Re} lSim\left(x_i, x_j\right)=\frac{\left|\overline{ClR}\left(x_i\right) \cap \overline{ClR}\left(x_j\right)\right|}{\left|\overline{ClR}\left(x_i\right)-\overline{ClR}\left(x_j\right)\right|}$$

Where $\overline{ClR}(X) \not\subset \overline{ClR}(Y)$

The relative similarity defined above, measures the ratio of size of the shared boundary between two sets and the number of elements that exclusively belong to the set under consideration. The subtraction of two sets may be zero hence the relative similarity may attain the indefinite value. Hence, to have the definite value of relative similarity in the positive real number domain, the first set should not be proper subset of the second while computing the relative similarity between two sets.

Table 1.Proposed relative similarity measures

Type of Covering based rough set	Proposed relative similarity measure	Percentage of shared boundary
First type covering based rough set	$\operatorname{Re} lSim\left(x_i, x_j\right)=\frac{\left|\overline{FCR}\left(x_i\right) \cap \overline{FCR}\left(x_j\right)\right|}{\left|\overline{FCR}\left(x_i\right)-\overline{FCR}\left(x_j\right)\right|}$	75%
Second type covering based rough set	$\operatorname{Re} lSim\left(x_i, x_j\right)=\frac{\left|\overline{SCR}\left(x_i\right) \cap \overline{SCR}\left(x_j\right)\right|}{\left|\overline{SCR}\left(x_i\right)-\overline{SCR}\left(x_j\right)\right|}$	90%
Third type covering based rough set	$\operatorname{Re} lSim\left(x_i, x_j\right)=\frac{\left|\overline{TCR}\left(x_i\right) \cap \overline{TCR}\left(x_j\right)\right|}{\left|\overline{TCR}\left(x_i\right)-\overline{TCR}\left(x_j\right)\right|}$	80%

The relative similarity is proposed for all three types of covering based rough sets as shown in table1 and forwarded to the constraint similarity upper approximation.

Now we define the proposed constrained-similarity upper approximation in the following definition:

Definition 8: Let $X\{x_1, x_2,,,, x_n\}$, $X \subseteq U$. For a fixed non-negative value $\sigma \in [0,1]$, the constrained similarity upper approximation of x_i is given by

$$\overline{ClRR}\left(\{x_i\}\right)=\left\{x_j \in \bigcup_{x_l \in \overline{CR}(x_i)} \overline{ClR}\left(x_l\right) \mid \operatorname{Re1}Sim\left(x_i, x_j\right) \geq \sigma\right\}.$$

In other words, all the sequences x_i which belong to the similarity upper approximations of elements of $\overline{CR}(x_i)$ that are relatively similar to x_i are constrained (or merged) into the next similarity upper approximation of x_i .

We repeat the process of computing successive constrained-similarity upper approximations for a given σ until two consecutive constrained-similarity upper approximations remain the same. Here, σ is a user-defined parameter called relative similarity, used to merge two upper approximations for the formation of the second and higher upper approximations. δ is a user defined threshold parameter use to define the similarity between two objects and is utilized to find the first upper approximation. The constrained-similarity upper approximation is computed for all transactions of U.

Experiments were performed on msnbc.com UCI dataset repository to check the performance of various types of covering based clustering algorithm. We recorded both the number of overlapping clusters as well as the time taken to execute the program on various sample of randomly selected data sequences. The relative similarity measures of first type,second type and third type covering based rough sets was identified and processed . Also we recorded the overlapping of clusters produced by all three types of covering based rough set. It was observed that compare than all three types covering based rough set approach second type covering based rough set gives a better result.

Table2.Overlapping Percentage of clusters for various types of covering based rough set

S.No	Type of covering based rough set theory	Overlapping clusters in percentage
1	First type covering based rough set	60%
2	Second type covering based rough set	90%
3	Third type covering based rough set	70%

It was observed that the numbers of clusters are minimum in covering based rough sequential clustering than rough set sequential clustering since overlapping is increasing in second type covering based rough set as shown in table 2.

CONCLUSIONS AND FUTURE WORK

In this paper, we studied the concept of covering based rough sets to cluster objects using the notion of similarity upper approximations. Usually, the clusters resulting from the web usage mining algorithms may not necessarily have crisp boundaries, rather they have fuzzy or rough boundaries [4].This paper results in comparison of all three types covering based rough clusters and the experiments on user navigation data which produces meaningful clusters that enable discovering of navigation patterns. Covering based Rough clusters are helpful to get early warnings of potentially significant changes in the clustering patterns. We experimented various types of covering based rough set clustering on a web navigation dataset collected from the UCI dataset repository .We observed that second type covering based provides good overlapping cluster than first and third type covers.

REFERENCES

[1] Pawlak, Z. “Rough Sets”, Int. Jour. Inf. Comp. Sc., 1982;11:341-356.

[2] Pawlak, Z. “Rough Sets: Theoretical Aspects of Reasoning about Data”, Kluwer Academic Publishers, 1991.

[3] Tripathy, B.K.; Tripathy, H.K., "Covering Based Rough Equivalence of Sets and Comparison of Knowledge," Computer Science and Information Technology - Spring Conference, 2009. IACSITSC '09. International Association of ,2009, pp.303,307

[4] Mitra, A.; Satapathy, S.R.; Paul, S., "Clustering analysis in social network using Covering Based Rough Set," Advance Computing Conference (IACC), 2013 IEEE 3rd International , 2013,pp.476-481.

[5] Zakowski, W. “Approximation in space (U, Π)”, Demonstratio Mathematica 16,1983, pp. 761-769.

[6] Zhu, W.; Fei-Yue Wang, "On Three Types of Covering-Based Rough Sets," Knowledge and Data Engineering, IEEE Transactions on ,2007;19:1131-1144.

[7] Zhu, W., "Basic Concepts in Covering-Based Rough Sets," Natural Computation, ICNC 2007. Third International Conference on , 2007; 5: 283-286.

[8] Tripathy, B.K., Mitra, A. and Ojha, J. “On Rough Equalities and Rough Equivalences of Sets”, RSCTC 2008-Akron, U.S.A., Springer-Verlag Berlin Heidelberg ,2008; 5306: 92–102.

[9] Novotny, M. and Pawlak, Z. “Characterization of Rough Top Equalities and Rough Bottom Equalities”, Bulletin of Pol. Acad. Sci.Math.,1985; 33: 91-97.

[10] Novotny, M. and Pawlak, Z. “On Rough Equalities”, Bulletin of Pol. Acad. Sci. Math., 1985;33: 99-104.

[11] Novotny, M. and Pawlak, Z. “Black Box Analysis and Rough Top Equalities”, Bulletin of Pol. Acad. Sci. Math.,1985;33:105-113.

[12] Tripathy B .K and A Mitra, “Some Topological Properties of Covering Based Rough Sets”, Proceedings of Int. Conf. On Emerging Trends In Computing (ICETiC 2009), 2009,pp. 263-268.

[13] Zhu, W.; Fei-Yue Wang, "Relationships among three types of covering rough sets," Granular Computing, 2006 IEEE International Conference on , 2006, pp.43-48.

[14] [14] Zhanhong Shi, Zengtai Gong, The further investigation of covering-based rough sets: Uncertainty characterization, similarity measure and generalized models, Information Sciences,2010;19:3745-3763.

[15] Yiyu Yao, Bingxue Yao, Covering based rough set approximations, Information Sciences, 2012:1: 91-107.

[16] William Zhu, Relationship between generalized rough sets based on binary relation and covering, Information Sciences,2009;3:210-225.

[17] Hassan Saneifar, Sandra Bringay, Anne Laurent, Maguelonne Teisseire. S2MP: Similarity Measure for Sequential Patterns. AusDM'08: The Australasian Data Mining Conference, 2008: 87:095-104.

[18] Pradeep Kumar, P. Radha Krishna, Raju. S. Bapi, Supriya Kumar De, Rough clustering of sequential data, Data & Knowledge Engineering, 2007:2:183-199.

[19] Jain, A.K. M.N. Murty, P.J. Flynn, Data clustering: a review, ACM Computing Surveys , 1999;3: 264–323.

[20] Lingras, C. West, Interval set clustering of web users with rough k-means, Journal of Intelligent Information Systems,2004,1:5–16.

[21] Joshi, A. and R. Krishnapuram, Robust fuzzy clustering methods to support web mining, in: Workshop in Data Mining and Knowledge Discovery, 1998, pp. 15-1–15-8.

[22] Bergroth, L. , H. Hakonen, T. Raita, A survey of longest common subsequence algorithm, in: Seventh International Symposium on String Processing and Information Retrieval, 2000, pp. 39–48.

[23] Xiaohua Hu C, T. Y. Lin and J Han, “A New Rough Sets Model Based on Database Systems”, Fundamenta Informaticae, IOS Press,2004,pp.1–18.

[24] A. Skowron, J. Stepaniuk, Tolerance approximation spaces, Fundamenta Informaticae, 1996:27: 245–253.

[25] Debadutta Mohanty,”Rough Set On Generalized Covering Approximation Spaces”, International Journal of Computer Science and Research,2010:1:20-35.