Journal of Intelligent & Fuzzy Systems 32 (2017) 2165–2185
DOI:10.3233/JIFS-162006
IOS Press

Rough set theory for document clustering: A review

K.A. Vidhya* and T.V. Geetha
Department of Computer Science and Engineering, Anna University, Chennai, Tamil Nadu, India

Abstract. Rough set theory is a mathematical framework that can be visualized as a soft computing tool dealing with the vagueness and uncertainty of data and is applied to pattern recognition, data mining, and knowledge discovery. Document clustering is another area of research with values which are a bag of words that describe contents within clusters. This work analyzes how rough set theory is used for document clustering to fix issues that clustering methods manage. In this survey, an exhaustive literature review of the concept of rough sets, as well as how the lower and upper approximation of a set can be used for document clustering, has been presented. Rough set clusters are shown to be useful for representing real-time applications such as biomedical inferences, network data handling, and citation analysis. The survey is done in phases, showing how machine learning algorithms have been incorporated for document clustering using rough set theory, as well as how rough set theory has been extended to adapt to document clustering with feature selection techniques and feature/dimensionality reduction and, finally, ending with a view of assorted clustering tasks where rough set theory is applied. The classification of rough set theory for document clustering is depicted and its applications presented in this paper. The rough set theory works with resolving ambiguity and uncertainty in data. To the best of our knowledge, a rough set clustering survey has not been done earlier in the literature reviewed and the survey ends with a critical analysis of rough set theory in each application of clustering.

Keywords: Rough set theory, document clustering, machine learning, approximation space

1. Introduction

Clustering, identified as an unsupervised learning problem, has become a widely-used statistical application when dealing with very big datasets and text mining tasks [6, 7]. The processing of accumulated, unlabeled elements that are grouped and labelled according to similarity in certain aspects is known as clustering. A grouped set of elements with similar objects and similarity between sets of elements are measures using similarity functions. The dimensionality of text representation is very large and is even more serious when it comes to clustering snippets or tweets. The basic theory of rough sets is represented as Pawlak rough sets [1] or classical rough sets that helps satisfy the demands of recent extensions and generalizations. Previously, rough set theory was developed and intended to serve communication, with knowledge part of the objects received from an equivalence relation with discourse. The equivalence relation must satisfy properties like symmetry, transitivity and associativity. In rough set theory, data are represented as an information system, usually a decision table, where the rows of the decision table are equivalent to objects, and columns equivalent to attributes.

According to region-based equivalence classes, rough set theory classifies attribute values to three approximation values: i) lower approximation ii) upper approximation, and iii) the boundary region [2–5]. Elements bifurcated merely through data collection are known to be a lower approximation,

*Corresponding author. K.A. Vidhya, Research Scholar, Department of Computer Science and Engineering, Anna University, Chennai, Tamil Nadu, India. Tel.: +91 9500683390; E-mail: avidhya06@gmail.com.

whereas objects classified according to the probability condition are an upper approximation, and the difference between the lower and upper approximation is identified as a boundary. An advantage of rough set theory is that it does not require added information or primary information about data, statistical information or the probability of occurrences in the Dempster–Shafer theory and the boundary value, or the value of possibility in the fuzzy set theory.

A lot of software has been implemented using rough set theory in fields like medicine, pharmacology, engineering, banking and market analysis and gene expression. The best explanation of rough set-based document clustering is a major area of research in the text mining [11] process as most documents need to be clustered using unsupervised [16] learning methods. To the best of our knowledge, a survey of rough set theory-based document clustering has not been done and thus this work explains the techniques of how to overcome the problems related to clustering using rough set theory. Off Late, there has been a lot of work in the area of using rough set theory to lever uncertainty in the process of selecting clustering [25] attributes. The discernibility relation between two clusters, and the working of the rough set for the given documents below, is illustrated with the following examples in Fig. 1.

Rough Set and Document Clustering:

In a set of documents = {D1, D2, . . . D8}, the document is pre-processed and formed as a bag of words and based on the tf-idf, the term-document matrix is formed as in text-mining and document clustering analysis. In rough set analysis, the discernibility matrix is formed as mentioned below:

D1: The apple is a fruit.

D2: Apple computer.

D3: Apples are good for health.

D4: The Apple iPad has lots of games.

D5: Apple

D6: Apple computers are most unlikely to be infected by viruses.

D7: Eating apples helps fight against viral infections, a fact recorded in medical history using computers.

D8: Apple computers and iPads aren't often susceptible to virus attacks as users tend to take good care of them.

The equivalence relation R is represented with two cluster labels, C1 and C2, with four sets {S1, S2, S3, S4}

S1|R = C1|Yes{{d1}, {**d5**, **d7**, **d8**}}

S2|R = C2|Yes{{d2, d3, d4, d6}, {**d5**, **d7**, **d8**}}

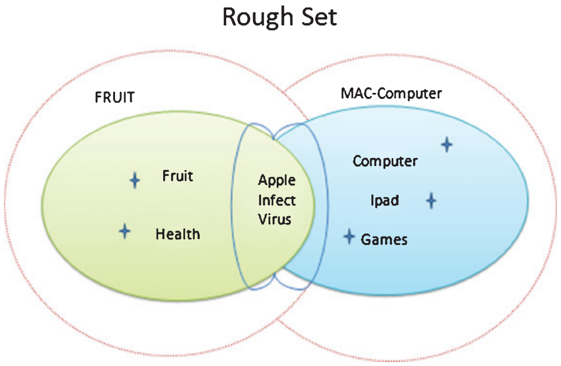

Fig. 1. RST with discernible sets.

S3|R = C1|Yes{**apple**, fruit, health, **virus**, **infect**}

S4|R = C2|Yes{**apple**, computer, ipad, games, **virus**, **infect**}

The keywords above highlighted in red are said to be indiscernible and the clusters C1, C2 formed. The tasks of document clustering include feature selection and dimensionality (or feature) reduction by which the number of features are reduced, precision and recall rate improved for search results, and clustering purity increased. The process of selecting attributes as subsets from the universe of attributes is termed feature selection. This process aims to maximize the use of unique features and eliminate unrelated ones to frame a new, useful learning model. The benefits of feature selection, usually done in two phases, significantly decreases the computation time of the algorithm and increases the accuracy of the resulting mode.

The rest of the paper is organized as follows. Section 1 discusses feature selection and dimensionality reduction methods using rough set theory. Section 3 discusses centroid calculation, cluster label discovery, cluster content discovery, similarity, dissimilarity, and the merits and demerits of methods for clustering using rough set theory. Section 3 describes the hybrid approach of rough set theory for document clustering. Section 4 discusses how rough set theory has been modified to achieve good cluster quality and solve clustering issues. Section 5 discusses how the theory of the rough set has been modified to ease the clustering process. A machine learning algorithm has been used along with rough set theory for the clustering process, the tools and techniques in the literature reviewed have been presented and, finally, the survey ends with various applications being discussed in Sections 6 and 7 respectively.

1.1. Motivation and challenges

Motivation behind this survey, is growth of enormous digital documents that need to classified or clustered according to the similarity of the document to mine useful information for predictive analysis of the data. Clustering remains an essential task for grouping the similar documents but there are many challenges in carrying out this task namely

1. Feature Selection had been an arduous task in clustering, deciding important features which would form the clusters narrows down our survey to solve the issues in clustering.
2. Deciding on the distance measure for numerical attributes is much easier whereas the identification of measure for categorical attributes is difficult, requires an efficient method to form better clusters.
3. Identifying the count of clusters is a challenging task if the number of class labels is not known initially, choosing the initial-k in k-means clustering remains an open issue in cluster analysis. If not decided with the quantity of clusters, the heterogeneous data may merge and similar types may be placed in different clusters, which can be disastrous if the approach used is hierarchical.
4. Deciding on the class labels when there is a lot of missing values in an unstructured data is an open issue that has to be addressed. While addressing the issues in clustering there is always an uncertainty in all the clustering tasks. To overcome the uncertainty in the clustering tasks, rough set theory which is a statistical has been used to efficiently improve the clustering results. This study aims to give an inference of how rough set theory has been used in tackling the issues in clustering tasks. The classification diagram of how the survey is organized is depicted in Fig. 2.

2. Feature selection and feature reduction using RST for document clustering

In text document clustering, feature selection or feature reduction plays a great role as the number of

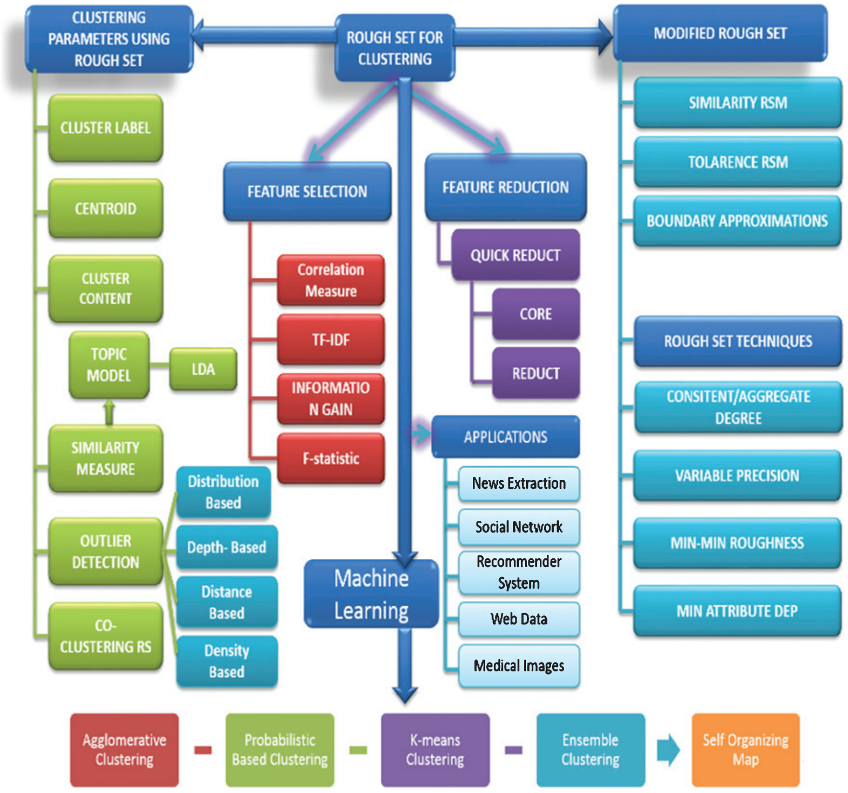

Fig. 2. Classification diagram – rough set theory for document clustering.

features needed to be selected are to be informative enough to be clustered into a certain group. The role of rough set theory in feature selection is to select the appropriate attribute contributing to the formation of clusters. Dimensionality reduction for clustering can be done by feature extraction and feature selection methods. Feature extraction develops new features by transforming the original features into a new, permanent feature that possesses effective information for the target concept. In contrast, feature selection only selects features that offer higher gains in terms of information and eliminates unimportant features to the target decision attribute. The process of feature extraction is quite complicated.

2.1. Feature selection

Feature selection is the process of selecting certain attributes, from a set of original attributes that contribute to efficient decision making. Feature selection identifies the most significant features, removes irrelevant attributes, and constructs a good decision table. Rough set-based feature selection aims at reducing the computation time of the machine learning algorithm, thereby increasing the accuracy of the learning model. Finding significant features is an NP-hard problem but, in most real-time environments, it’s not necessary to have the entire features taken into account to ensure first-rate accuracy for the resulting model. Table 1 gives information about the methodology of feature selection incorporating rough set theory for document clustering.

2.2. Feature reduction for clustering using RST

A significant problem on hand is to find particular subsets of condition attributes to form clusters. Attributes from such subsets are considerably redundant and can be removed without causing deterioration of clustering quality and without the induction of rules inherently found. These subsets of the condition attributes are referred to as a reduct of the clustering values, and a reduct of the rough set that contains the least condition attributes is called the minimal reduct. If the removal of an attribute causes inconsistency in the degree of clustering quality, the attribute is deemed essential. Attribute reduction can be done using RST, usually to find its effectiveness in numerous domains focussing on the reduct of the rough set model being conducted. Finding minimal reduct points is an NP-hard problem and assorted algorithms were proposed to address this issue. Currently, there are three main approaches to obtain the reduct in an RS model.

The following are certain algorithms that have been listed for feature reduction using the rough set. The initial attribute-removing algorithm haphazardly removes attributes from the feature collection until a core reduct is obtained. The heuristic algorithm adds attributes to the core set according to heuristic principles such as the quality classification degree. Skowron et al. proposed a discernibility matrix-based algorithm through discernibility matrices for describing knowledge systems. Pros and cons characterize the latter two algorithms. Although the addition of an attribute to the algorithm constitutes lesser com-

Table 1
Rough set in document clustering

Rough Set Clustering and Feature Selection	Author
Rough set theory was used to construct, under supervision, reducts for reducing the number of features without labels. Apart from the hierarchical method, the feature selection method can also be applied, combined with the clustering method.	Questier et al. [15]
A heuristic algorithm based on rough set theory to learn about the feature subset has been proposed	Zhang et al. [16]
An algorithm was proposed to find the set of all reducts in a much shorter time, compared to the elimination method.	Starzyk et al. [17]
An unsupervised algorithm was proposed for clustering datasets without knowing the decision attribute for feature selection.	Jaganathan et al. [18]
Rules were extracted to carry out two-fold data clustering, data discretization, and attribute selection.	Abidi et al. [19]
The author analyzed a feature selection method with certain basic properties covering generalized rough sets, and a set of axioms proposed to characterize the covering lower approximation operation for document clustering.	Zhu et al. [20]
Rough set feature selection has been implemented using a distance measure for document clustering.	Parthalain Neil et al. [8]
Rough set-based techniques were used to select clustering attributes.	Mazlack et al. [50]
The ITDR was proposed as an alternative method for categorical data clustering, varying from the baseline method and rough set attribute reliance upon the rough entropy being calculated by the categorically-valued information system.	In-Kyoo Park et al. [40]

puting complexity, it may result in a failure to find a reduct and, besides, the calculation does not necessarily detect the minimal reduct. On the other hand, the discernibility matrices-based algorithm is definitively able to find the minimal reduct with limitations in computation complexity that require the traversal of all possible combinations of attributes, thereby compounding the risks of a combinatorial explosion.

2.3. Variable precision rough set (VPRS) model for attribute reduction

Generally in clustering the documents, determining the initial number of clusters is currently hard and fixing the number of features to get a meaningful clusters is even more heuristic task. In this regard, W.Zaikro proposed attribute reduction with a variable precision rough set to conceal the number of objects in the β-lower approximations of all decision classes by defining the dependency function. This kind of reduct is called an approximate reduct or β-reduct, and was noted not to preserve the β-lower approximation of each decision class in VPRS, proposed by Inuiguchi and Beynon [53]. Therefore, the derived decision rules from the β-reduct may be in conflict with the ones from the original decision system proposed by Zhang et al. [35]. Considering these drawbacks of the β-reduct, the concepts of β-lower and upper distributed reducts based on VPRS were presented in [34] to preserve β-lower and upper approximations, respectively. Always, the derived decision rules from β-lower and upper distributed reducts are compatible with the ones derived from the original system. Furthermore, in the β-lower and upper approximations, distributed discernibility matrices were also provided to find the β-lower and upper distributed reducts, respectively. Table 2 discusses the feature reduction techniques and its implications while using Variable Precision Rough Set.

2.4. Analysis of rough set in feature selection and feature reduction

The Rough set based attribute reduction works well for incomplete dataset in structure learning as it does not require any information for clustering the data. In Bayesian network there are two types of learning relationship between the entities, it can be either structure learning [13] or parameter learning. The learning of the Bayesian structural learning can be done with the rough attribute reduction method. In a large database choosing the non-redundant and irrelevant attributes will lead to slow learning and fall in accuracy. To avoid that rough set based feature selection, the dependency measure of the approximation region can be used to estimate the feature that are highly informative. The dependency measure is the indiscernibility measure that can be used to estimate the clusters.

Accelerated quick reduct algorithm can be used to select a minimal set of subsets for the given datasets but the drawback is that it works well only for smaller datasets. Rough set based PCA (Principal Component Analysis) helps in reducing the number of features and the error rate is much reduced compared to the other dimensionality reduction algorithms. This aims in acquiring the important attributes that convey an appropriate generalization technique. Rough Set Theory addresses the high-dimensional data, where there are missing attribute and redundant features has to be identified and eliminated in order to improve the efficiency and cluster purity. Variable precision based attribute reduction has to ad dress two problems, reduct anomaly issue which causes the inconsistency in the positive approximation region there by formed the rules which are irrelevant.

Finally to conclude the attribute reduction problem is an NP-Hard problem, there should be an appropriate strategy for better trade-off between the

Table 2
Feature reduction using rough set theory

Feature Reduction and Variable Precision Rough Set	Author
Dependency function, the positive region, and the β-lower approximation may not be monotonic in VPRS after deleting a condition attribute. Consequently, the algorithm of finding the β-reduct may not be convergent by applying the measure of dependency function.	J. Zhou et al.
The algorithm of attribute reduction was proposed using the β-variable precision rough entropy, where both the precision parameter and the correction coefficient were given in advance.	L. Sun et al. [22]
While selecting the biggest dependency attribute during each iteration using VPRS's dependency as a criterion to improve the current QuickReduct algorithm, the authors admitted that a few minor errors had occurred during the process.	X. Pan et al. [16]
The binary discernibility matrix has been proposed to find attribute reduction for VPRS.	X. Zhang et al. [34]
A distributed discernibility matrix has been proposed that can also be used to find all reducts. Though finding all reducts using this technique is an NP-hard problem, the method draws the same mathematical foundation as does VPRS.	Zhang et al. [35]

number of minimal reducts and thereby reducing the computational overhead. The inconsistency in the positive region in VPRS can be easily maintained by the set covering problem using a set covering heuristic function. Rough set for feature reduction works well for large datasets assuming the pre-clustering step can be discarded.

3. Rough set for clustering tasks

Rough set theory has been used for clustering tasks such as calculating centroids, fixing negative and positive regions, finding cluster labels, finding topics using the similarity measure calculation and cluster content discovery. The detection of outliers is a major issue in the clustering process where there is uncertainty to be addressed, such as whether the outlier can be put in an existing cluster or two outliers can form a new cluster. The topic modeling task can be done with rough boundaries with the probability of each document's existence under a particular topic. Apart from the issues above in clustering, dominance rough set clustering has been analyzed and presented in this section.

3.1. Centroid calculation

The effects of the lower and upper bounds have been modified to be adapted to conventional K-means, and the three properties of rough sets are not necessarily independent or complete. However, itemizing them will be useful in understanding the rough set adaptation of developing statistical and categorical clustering methods. Yao and Zhao [27], in their study, discussed the "decision theoretical rough set model" and provided the core properties of rough sets for a positive boundary and negative regions. While applying rough sets to the K-means clustering method, there is a need for the additional concept of lower and upper bounds. While calculating the centroid of clusters out of the conventional K-means, a modification in K-means in order to include the possible effects to these bounds is required. Besides, enumerating the same will prove useful at a later stage in comprehending the rough set adoption of evolving neural and statistical clustering methods.

$$c = \frac{\sum_{x \in lower(c)} X}{|lower\ c|} \tag{1}$$

Else if lower(c) = $\emptyset$ and bnd(c) $\neq \emptyset$

$$c = \frac{\sum_{x \in bnd(c)} X}{|bnd\ c|} \tag{2}$$

Else

$$c = w_{lower} * \frac{\sum_{x \in lower(c)} X}{|lower\ c|} + w_{upper} * \frac{\sum_{x \in bnd(c)} X}{|bnd\ c|} \tag{3}$$

The parameters w_{lower} and w_{upper} correspond to the relative importance of lower and upper bounds, and $w_{lower} + w_{upper} = 1$. If the lower bound and upper bound are most likely equal, they are conventional clusters. Therefore, the boundary region $bnd(c)$ will be empty, and the second term in the equation ignored. Consequently, the equation above will be reduced to conventional centroid calculations.

3.2. Co-clustering rough set

In normal clustering, the focus is either on row or column dimensions, while co-clustering targets both object and feature dimensions and seeks "co-clusters" of interrelated objects and features by sporadically changing object clustering into feature clustering. Co-clustering is an unsupervised data analysis algorithm that utilizes the duality of both dimensions, discovers hidden patterns, reduces dimensionality considerably, generates a representation easily and reduces running time substantially. Cho Hyuk et al. [52], in their study, developed an algorithm that can update both row and column clustering data dynamically only for each available data point. The streaming data collected from various sensor networks or other mobile devices are handled by the algorithm proposed by authors. Although relatively new, co-clustering is considered much more desirable than traditional clustering. Further, they learnt that because of these desirable characteristics and co-clustering's theoretical maturity, it has become popular for diverse applications in web mining, image retrieval, recommendation, bioinformatics, and so on, in a relatively short time.

3.3. Qualitative/quantitative rough clustering

Analyzing the unsupervised clustering semantic process can be difficult with large datasets to be clustered. The relations between the two attributes contributes to the semantic analysis of the qualitative clustering process. Pawan Lingras et al. designed a framework based on dominance relations which help to analyse clustering qualitatively, a method useful for combining clustering schemes related to multifarious categories. The qualitative combination helps to analyze the quantitative counterpart which can also be used on behalf of the quantitative combination. They worked again to extend the framework to include a rough set-based clustering called a dominance-based framework that can be used to axiomatize the process of clustering. Pawan Lingras et al. [26] proposed two clustering schemes: one based on spending and the other on visits of the customers who arrive to the shop. The authors had combined these two clustering schemes to get a better understanding of the value of customers. Generally speaking, this type of combination can be attained through combining two sets of attributes and clustering objects.

This is quite possible by assigning attributes different weights according to their relative importance, and the authors admitted that it was a traditional combination with a quantitative approach. The dominance relation-based framework is identified as a combination qualitative method that is useful when it can also be an alternative to the quantitative approach by reducing emphasis on numeric distances. Masson and Denoeux, in their work, stated that fuzzy clustering can be extended by applying the ‘belief function’ theory to denote ‘evidential clustering’ that detects lower and upper bounds, just the same as the rough set theory. The fuzzy degree of membership becomes very descriptive in cases where interpreting clustering occurs. Rough set-based clustering techniques provide a better solution than conventional clustering and fuzzy clustering.

3.4. Dominance rough set clustering

Based on the decision class to which objects called criteria belong, they are named fully-ordered domains in dominance rough set theory. The DRSA (dominance-based rough set approach) was used to identify inconsistencies in the rough set theory by Greco et al. for their study, given that DRSA stirs interest in scholars who have undertaken a study of rough set theory. DRSA has applications in medical research, especially in the diagnosis of liver disorders and breast cancer. Besides, this mechanism is utilized to predict bankruptcy risks, credit rating and house pricing. Chen and Tzeng proposed an efficient implementation of DRSA. A combination of dominance relations from multiple criteria is a key feature of recent research using the dominance relation. The present study uses certain concepts proposed by Alpigini et al. [28] and Chen and Tzeng for analyzing crisp and rough clustering. In dominance-based rough sets, attributes with fully ordered domains are called criteria.

Objects are ordered based on the decision class they belong to. However, the need for a dominance-based rough set theory is presented, in tandem with how rough sets have been utilized to document clustering, in the following table. The first attempt on rough set-based techniques to select the clustering attribute was proposed by Mazlack et al. [50]. They proposed two techniques, bi-clustering and total roughness (TR), based on the bi-valued attribute, maximum total roughness in each attribute, respectively. One of the most successful pioneering rough clustering techniques is MMR, proposed by Parmar [48].

3.5. Outlier detection

In real-world datasets, time-based triggered datasets are the most readily available real-time ones, and detecting outliers which are grossly different from the spatiotemporal data left behind is considered to be the principal challenge in knowledge discovery and data mining applications. Alessia Albanese et al. [43] discussed the “outlier detection problem” in spatiotemporal data and described a rough set approach to identify top outliers in an unlabeled spatiotemporal dataset. Alessia Albanese et al. proposed the “rough outlier set extraction” (ROSE) that depends on a rough set based on the theoretic representation of the outlier set using the lower and upper approximations of the rough set theory.

In the following Table 3 the role of rough set in outlier detection has been discussed, the merits and the demerits of RST has been quoted in Table 3. The authors also identified a new set called the kernel set, which is the subset of the proper set that conveys the key data with reference to the structure and results retrieved from it. Table 3 discusses the drawbacks of the outlier methods and how rough set has been used to overcome those drawbacks. Daneshpazhouh et al. [32] proposed a modified entropy-based outlier detection [33] method which calculates negative and positive instances of data, motivated by the LSA algorithm for outlier detection. Daneshpazhouh et al. [32] proposed an objective function to find the small subset of elements k where this minimizes the entropy of the set. George Peters et al. [23] introduced its compliance to classical k-means the numerical stability and its performance in the outliers. Zhu et al. [31]

Table 3
Rough set for outlier dectection

Rough Set	Outlier Method	Drawbacks
Jiang et al. (2009) [51] introduced the distance-based outlier detection to rough set theory and proposed definitions of distance metrics for distance-based outlier detection in rough set theory.	Distance-based outlier detection	The widely-used distance-based method has certain drawbacks. For instance, it cannot assign a degree of outlierness for each object, it is difficult to find local outliers using the method, and the computation complexity of the method is usually too high.
Jiang et al. (2005) proposed a boundary-based method for outlier detection based on notions in rough sets. Objects in boundary regions are deemed likelier to be outliers than objects in lower approximations. (Chen et al., 2010).	Boundary-based method	The boundary-based method is reckoned intuitive and meaningful but it does not have a record of good performance for outlier detection.
F. Jiang et al. [51] combined opinions from distance-based and boundary-based methods for outlier detection to obtain a hybrid method for outlier detection, aiming to complement the advantages of the two previous methods and solve problems associated with both together.	Hybrid model: distance-based and boundary-based on rough set theory	Time complexity is relatively low and is most suitable for dealing with large datasets.
Feng Jiang et al. [51] proposed a novel definition of outliers in information systems of rough set theory: sequence-based outliers.	Distance-based measure	Proposed the overlap metric and value difference metric in rough set theory, both of which are especially designed to deal with nominal attributes.
Yumin Chen et al. [59] proposed a neighborhood-based metric on outlier detection and compared neighborhood outlier detection with a k-nearest neighbor.		Distance-based measure
Other Outlier Methods	Drawbacks	
Johnsin et al. (1998) revealed that different layers of k-d convex hulls and flag terms in the outer layer as outliers could be computable, based on computational geometry.	Depth-based method	Suffers from the dimensionality curse and cannot cope with large ks, compared to the depth-based method.
Data carried as input could be classified by detecting outliers as by-products (Jain et al., 1999).	Clustering-based method	As the main objective is clustering, it is not optimized for outlier detection.
A classical method in statistics, based on a standard distribution model (normal, Poisson, etc.), objects which deviate from the model are recognized as outliers (Barnett and Lewis, 1994).	Distribution-based method	Its greatest disadvantage is that the distribution of measurement data is unknown in practice.

proposed an entropy based approach to calculate the approximation space for rough set clustering task.

3.6. Topic modelling

Although existing topic models can solve problems of polysemy and synonymy, it is unsuitable for computing the similarity of two short texts directly. To resolve this, Zhang et al. [42] proposed a topic-model information system using rough sets, where topic-word distribution can be transformed to an information system, Topic Word Information System = (W_n, R_s, V_A, f), where W_n is a set of N words, R_s a set of T topics, V_A a set of topic values, and f an information function that assigns values from topics to words.

For the i-th topic, $f : R_s \rightarrow V_{Ak}$, where V_{Ak} is the range of its value. The element of Φ is a real number between 0 and 1. The topic value in TWIS is usually discrete. Therefore, a real number is mapped to a discrete value, e.g., “low,” “middle” and “high.” Two words having a similar topic-word distribution are considered to be similar under a certain topic. Words in the positive region consistently reflect the related topic. If the words in two positive regions are discernible, the common words may express different topics.

The procedure for finding synonyms and polysemy is described below, where the first step involves: Computing the union set of the three sets above, i.e., $RA = RP \cup RQ \cup RC$, where RP and RQ are the distances between the given two short texts and RC the similarity measure between the texts.

Extracting the subset of TWIS by replacing Wn with RA;

Computing two positive regions, POST S (RP) and P OST SRQ;

Finding synonyms: a. Obtain ΔRP = RP − POST S (RP) and ΔRQ = RQ − POST S (RQ) b. For each p ∈ ΔRP and q ∈ ΔRQ, if (p, q) ∈ IND (T S), then p and q are synonymous; and

Finding polysemy: a. Obtain pairs {(p, q)|p ∈ POST S (RP) and q ∈ POST S (RQ)} b. If a pair exists (p, q), not in IND (T S), then c ∈ RC is polysemous.

3.7. Advantages and drawback of rough set in addressing clustering tasks

The clustering tasks like co-clustering try to handle streaming data where the complete dataset is not available to start with and online samples of data are fetched from the main memory incrementally, thus updating information from the co-clusters formed by each iteration. To put it plainly, it also does this in a single pass, i.e. it just processes each instance (i.e., data point) at a time. Data that is available over a certain period of time, coupled with the characteristics and patterns of the data in question, may still be captured; thus, grouping and class information latent in the whole data processed over time can be identified.

The similarity measure that works for finding short texts using rough sets has been proved to work for concise datasets. Polysemy and synonyms are found using rough sets for modeling the topic of the given short text generated by the latent Dirichlet allocation and the weights of these words adjusted correspondingly. The rough set based outlier detection works better for all datasets compared with the other outlier detection algorithms. The rough set approach is computationally less intensive compared other outlier detection methods. This yields better results for varying dataset sizes. Thus the usual clustering tasks problems are solved using a rough set which yields better results when compared to state of art clustering methods.

4. Extended rough set theory

Rough set theory is modified and extended for document clustering to enhance the results of clustering that can be achieved using the similarity rough set model (SRSM), tolerance rough set model (TRSM), WordNet similarity rough set model and neighbourhood rough set model.

4.1. Similarity RSM (SRSM)

The similarity rough set model, a mathematical model extended from Pawlak's rough set model, uses a similarity relation instead of an equivalence relation. It is also an expansion of the tolerance rough set model with a tolerance relation. Equivalence, tolerance, and similarity are binary relations [38] that can be used to represent relations between terms in document clustering. An equivalence relation must satisfy reflexive, symmetric and transitive properties, while a tolerance relation does not have to satisfy a transitive one. A similarity relation must be reflexive, but is not required to be symmetric and transitive [11, 12]. The new similarity rough set model is defined as follows. The universe U of the approximation space $(U,\ R)$ is the set of all terms T used in the document vectors. The binary relation R is defined by

$$t_j\ Rt_i \Leftrightarrow f_D(t_i,\ t_j) \geq \alpha f_D(t_i) \tag{4}$$

$f_D(t_i,\ t_j)$ is the number of documents in the document set D in which terms t_i and t_j co-occur, $f_D(t_i)$ is the number of documents in D in which term t_i occurs and α a parameter $(0 < \alpha < 1)$. The relation R defined above is a similarity relation that satisfies only reflexivity. The co-occurrence of terms was chosen for document clustering based on TRSM because it offers a meaningful interpretation of dependency and the semantic relation of index terms in the context of information retrieval, and is relatively simple and computationally efficient. However, in large collections of documents, occurrences may have a big difference between terms. There are high-frequency and low-frequency terms. In TRSM, the tolerance class of terms depends on the constant threshold θ. If the frequency of a term is higher, the number of documents in which it co-occurs with other terms will be also large and, as a result, many terms end up crowding the tolerance class.

4.2. Tolerance RSM (TRSM)

The extension of the vector space model with the use of rough set theory and co-occurrence terms are identified as TRSM and SRSM, ie., tolerance rough set model [6] and similarity rough set model, also applied effectively for document clustering. However, several studies have shown that SRSM performs better than TRSM and assorted conventional methods. The TRSM method is used to work out term categorizations in information retrieval, text mining, and sometimes as a base for sundry model documents. With its ability to deal with vagueness and fuzziness, the rough tolerance set seems to be a promising tool to model relations between terms and documents. In a plethora of information retrieval-related problems, especially in document clustering, defining the relation (i.e., similarity or distance) between document-document, term-term or term-document is essential. In the vector space model, it has been noticed [34] that a single document is usually represented by relatively few terms, resulting in zero-valued similarities which decrease the quality of clustering. Cluster performance must be enriched to increase the effectiveness of document clustering [7].

In general, observations of the three properties of an equivalence relation R (reflexive, xRx; symmetric, xRy $\longrightarrow$ y Rx; and transitive, xRy $\wedge$ yRz $\rightarrow$ xRz for $\forall$x, y, z $\in$ U) demonstrate that the transitive property does not always hold in certain application domains, particularly in natural language processing and information retrieval. This can be illustrated by considering words from Roget's Thesaurus, where each word is associated with a class of other words that have similar meanings. It is clear that these classes are not disjoint (equivalence classes) but overlapping, and the meaning of the words is not transitive. Overlapping classes can be generated using tolerance relations in association with either symmetric or reflexive properties. Besides, using tolerance relations (where generalized spaces are also known as tolerance spaces) that contain overlapping classes of objects in totality results in the identification of tolerance classes.

Usually, tolerance spaces are defined in the format quadruple $R = (U, I, v, P)$, in which U is a universe of objects. Objects I: U $\longrightarrow$ 2U is an uncertainty function, ν: 2U $\times$ 2U $\longrightarrow$ [0, 1] a vague inclusion, and P: I(U) $\longrightarrow$ $\{0, 1\}$ a structural function. The assumption is made that an object x is perceived by information Inf(x) about it. The uncertainty function I: U $\longrightarrow$ 2U determines I (x) as a tolerance class of all objects that are considered to have information similar to x.

The above can be any function that satisfies the condition x $\in$ I (x) and y $\in$ I (x) iff x $\in$ I (y) for any x, y $\in$ U, where the function should correspond to I $\subseteq$ U $\times$ U, implying that xIy iff y $\in$ I (x). Since 'I' satisfies the properties of reflexivity and symmetry, 'I' is a tolerance relation. The vague inclusion $\nu : 2U \times 2U \rightarrow [0, 1]$ measures the degree of inclusion of sets; in particular, it relates to the question of whether the tolerance class I (x) of an object x $\in$ U is included in a set X. There is only one requirement of monotonicity with respect to the secondary argument of ν, that is, $\nu(X, Y) \leq \nu(X, Z)$ for any $X, Y, Z \subseteq U$ and $Y \subseteq Z$. Finally, the structurality function is introduced by analogies with mathematical morphology. In the construction of the approximation spaces, only tolerance sets are considered, being structural elements.

4.3. WordNET similarity RSM

Several approaches have employed WordNet-based semantic similarity to enhance the performance of document clustering. The approach is so handled that term weights in document vectors – in respect to their relationship with other terms co-occurring in the text – have been readjusted, with the term weights using the modified VSM model. WordNet and SRSM-related methods provide better results when compared to the standard VSM model. A new method has been identified as a representative model for documents using rough set theory and WordNet related semantic similarity.

In document clustering, the effect of semantic similarity between terms is large, and must be taken into account to enhance the performance of VSM. In SRSM, the semantic relation between terms is calculated using the co-occurrence of terms. However, there seem to be cases where terms have high co-occurrence but low semantic similarity. WordNet-based approaches measure the relatedness of terms using a lexical database. Based on the ontology structure of terms or definitions of terms in WordNet, Nguyen Chi Thanh et al. computed the scores of semantic relatedness. However, as a general dictionary, WordNet does not cover all terms and term meanings in every specific subject. Moreover, the semantic relations of terms may be different in different fields. Thanh et al. recommended that both approaches be exploited to get better clustering results, and proposed [39] a new relation that integrates WordNet knowledge to eliminate terms having no similar meaning but a high frequency of co-occurrence.

$$\begin{aligned} t_j R t_i \Leftrightarrow f_D(t_i, t_j) \geq \alpha f_D(t_i) \\ \cap((t_i \textit{ not in word net}) \cup \\ (t_i \textit{ not in word net}) \cup \textit{ sim } (t_i, t_j) > \theta)) \quad (5) \end{aligned}$$

The relation R is defined by a similarity relation because it is reflexive, non-symmetric and non-transitive. The basic idea is that term tj is similar to ti when tj is similar to ti from the viewpoint of co-occurrence and also similar in WordNet's semantics. If ti or tj is not in WordNet, we use only the co-occurrence similarity.

4.4. Variable Precision RSM

Ziakro further developed an RST innovation called the variable precision rough set (VPRS) model, which incorporates probabilistic decision rules. As noted by Kattan and Cooper, this is an important extension with reference to computer-based decision methods, but it is quite obvious that the patterns of classes frequently overlap, implying that the predictor information might be incomplete. This lack of information results in probabilistic decision making, where perfect prediction accuracy is not expected. VPRS has the additional desirable property of owing partial classification, compared to the complete classification required by RST, relative to the traditional rough set approach.

The level of confidence in clustering is such that expectations are high, so we can expect to get a predictable analysis of data brought out by using the majority inclusion relation. VPRS paves the way for resolving clustering problems using uncertain data and a non-functional relationship among attributes, and also provides flexibility to the rigid boundary definition of Pawlak's rough set model [1] by suggesting improvements to the model for enhanced suitability. Hence this provides a new method to approach noisy data. Table 4 depicts the proposed VPRS method for document clustering tasks.

Inference: There may be, similarly, a few terms in the tolerance class of a word with low frequency. Consequently, the size of a term's tolerance class is affected by its frequency in the collection though not the nature of the relationship between word meanings. The similarity rough set model was proposed in document clustering to resolve problems with TRSM.

5. Rough clustering techniques

In this section, rough set clustering techniques have been discussed and presented in sequence, assigning consistent and aggregate degrees of elements in the rough set in tandem with another technique where the rough set space is divided into n number of subspaces with mean and standard deviations in which attributes are categorized as follows.

5.1. Consistent/aggregate degrees

Duo Chen et al. [21] presented rough set-based hierarchical clustering using consistent and aggregate degrees, and their functions in the clustering process are analyzed at length. The clustering level calculation formula is so designed that two factors, consistent degrees and aggregate degrees, are taken into account. In the case of the traditional RST, only the consistent DT and inconsistent DT are given, and the consistent measure of the DT is not defined. In fact, in a clustering DT $(U,\ AU\ \{d\})$, each $a_k \in a$ represents the membership distribution of each object in U, corresponding to the cluster Dj in clustering model P.

This paper holds that this distribution can include information on coordination of the condition attribute set A to the clustering model P in the DT, further reflecting that the clustering accuracy can be used as a kind of measure in the clustering process. On the basis of the analysis above, we have offered a definition of a consistent degree. The root-mean-square is employed to define $AGD(P,\ j)$, whereas the arithmetic mean value is not adopted to make the definition, chiefly to avoid the yielding of trivial values (i.e. a simple 0 or 1). This method borrows the predictiveness

Table 4
VPRS model for document clustering

Herawan et al. [6]	Used VPRS in the process of selecting clustering attributes.
Wu Chen et al.	Used VPRS in designing multi-granulation rough sets.
Park et al. [40]	Used accuracy of approximation using VPRS and Min-Min roughness.
Dominik Ślezak et al.	Initiated non-parametric modification of the VPRS model, called the Bayesian rough set (BRS) model.
Ziarko	Proposed the VPRS model.
Kryszkiewicz et al.	Proposed the tolerance relation rough set model.
Stenfanowski et al.	Proposed the similar relation rough set model.
Wang Guoyin et al.	Proposed the limited tolerance relation rough set model.
Greco et al.	Proposed a rough set model based on advantageous relationships.
Qian et al.	Proposed several basic views for establishing a multi-granulation rough set model in incomplete information systems.

definition method listed in the literature reviewed. But predictiveness and the aggregate degree differ in three points: the first is that predictiveness is defined in terms of probability distribution, while the aggregate degree is defined in terms of the membership matrix; the second is that the aggregate degree has had normalization manipulated, thus preventing it from quickly reaching rather large numerical values with an increase in cluster numbers; and the third is that ushering in the Ψ function is convenient for a match with a consistent degree.

5.2. Rough sub space RSM and min-min roughness

The theory of rough sets holds that the attributes of categorical data are decomposed into a number of rough subspaces. This design helps retain quality in clustering data to provide a positive solution by squeezing the resulting partitions. Can Gao et al. introduced an index to evaluate the clustering algorithm for categorical data. Another attempt has been made by Mazlak et al. [4] using RST to select partitioning attributes for clustering. The total roughness method is applied to identify narrowed partitions within the clustering. Even then, the procedure for partitioning starts with binary-valued attributes and the total roughness criterion is used only for multi-valued attributes. This is a handicap, given that partitioning is done on a binary attribute though the total roughness for a multi-valued attribute is lower. MMR overcomes this drawback by clustering objects on all attributes. In addition, MMR suggests a new method of measuring data similarities for the roughness concept by attempting a measure-termed mean roughness which can be compared to the method proposed by Mazlak et al. [4], based on RST. In other words, the same has been reproduced below:

Suppose $a_i \in A,\ V(a_i)$ has k-different values, say γ_k, k = 1, 2, 3 . . . n. Let $Y(a_i = \gamma_k)$, k = 1, 2, 3 . . . n be a subset of objects having k-different values of attribute a_i. Min-roughness of the set $Y(a_i = \gamma_k)$, k = 1, 2, . . . n, with respect to a_i, where i ≠ j denoted $MR_H = (Y_{i[\alpha]}|X_j)$ is defined by the following equation

$$MR_H(Y_{i[\alpha]}|X_j) = \min(H(Y|a_i = \alpha|X_j)) \tag{6}$$

Min-mean roughness of attributes of $a_i \in A$ with respect to $a_j \in A$ where i ≠ j is denoted by $MMR_H = (a_i|a_j)$ is evaluated as follows

$$\begin{aligned} &MMR_H(Y_i|X_j)\\ &\quad = MR_H(Y_{i[\alpha]}|X_j) + \dots\\ &\qquad + MR_H(Y_{i[a_{i[\nu][\alpha]}]}|X_j)/V(a_i) \end{aligned} \tag{7}$$

Where V (a_i) is the set of values of attributes $a_i \in A$.

Given n attribute min-mean-min roughness of attributes of $a_i \in Y$ with respect to $a_j \in A$ where i ≠ j refers to the min of $MMR_H = (a_i|a_j)$ is denoted by s evaluated as follows $MMMR_H = (Y_i|X_j)$ is obtained by the following formula:

$$\begin{aligned} &MMMR_H(Y_i|X_j)\\ &\quad = \min(MMR_H(Y_1|X_1), \dots\ \dots,\\ &\qquad MMR_H(Y_m|X_n) \end{aligned} \tag{8}$$

However, Min Li et al. [54] gave a new measure for min-min roughness by taking the total mean distribution and mean distribution into consideration. Let IS = (U, A, V, f) be a categorical information system, and MMR defined as the minimum of the min–roughness of all attributes in A. That is, MR determines the best crispness each attribute can achieve, and MMR determines the best split among all attributes. The MMR principle stipulates choosing one among a list of candidate attributes with the minimum min–roughness. On the basis of the definitions mentioned above, Parmar et al. [48] proposed a top down hierarchical clustering algorithm named MMR (Min–Min Roughness), which iteratively divides a group of objects with the goal of achieving better clustering crispness. The algorithm takes k, the number of clusters, as an input and terminates when this pre-defined number, k, is reached.

MMR is considered to be a robust clustering algorithm that deals with the uncertainty in the process of clustering categorical data. The MMR algorithm is valued for the unique advantages it offers, such as i) its ability to handle uncertainty in the clustering process; ii) its robustness as a clustering algorithm that enables users to obtain stable results with only a single input: the number of clusters; and iii) its capability in terms of handling large datasets. The Gordian technique of the MMR algorithm is utilizing the concept of roughness to determine the clustering attribute from all candidate attributes. However, the roughness (or the accuracy of approximation) cannot reflect the power of discernibility to boundary objects. Inspired by this, a novel concept of distribution approximation precision, which is proved to be a far more effective uncertainty measure.

5.3. Dynamic Rough Clustering

The central idea of dynamic clustering [25] is to detect changing patterns in data over time and adapt clustering parameters accordingly. To address this challenge, soft computing approaches seem to be of particular interest since the transition from stable data structures to dynamic behaviour cannot be crisply defined. Hence, soft computing techniques such as neural networks [56, 58], evolutionary computing, and fuzzy sets, among others, play a crucial role in dynamic clustering. Neural networks, for instance, offer excellent learning abilities and in this sense become of interest when dealing with varied facets of pattern updating. Evolutionary computing with its inherent adaptation capabilities provides powerful mechanisms for dynamic clustering. Georg Peters et al. enriched the field of dynamic clustering by introducing a dynamic clustering method based on rough set theory, in particular, rough k-means [49]. Dynamic clustering approaches – one of the first dynamic cluster methods – has been presented by Georg Peters et al. where groups of objects called "samplings" adapt over time and evolve into interesting clusters. Here, a particular element of the analysed phenomenon – the notion of changing objects – is of special importance. In Table 5 the working mode of dynamic clustering has been discussed for efficient document clustering.

In dynamic clustering, static as well as dynamic input data can be analysed. Input data (i.e., feature values) are called static if no time-dependent variation is considered. In the opposite case, dynamic input data has to be clustered – for example, feature values become trajectories instead of real values, as shown below. Several clustering systems have been proposed that treat static input data using dynamic elements during classifier design, i.e., the respective algorithm adapts dynamically while being applied to a set of static input data. CHAMELEON is such a system, using hierarchical clustering where the merging decision on each hierarchy adapts to changing cluster characteristics.

A decision that has to be taken in objective function-based clustering methods, such as k-means and fuzzy k-means, is to determine the number of clusters (here: k) before running the respective algorithm. This issue becomes particularly interesting in dynamic clustering where the cluster numbers can vary over time. Several papers have dealt with this decision of identifying an appropriate cluster number. The clustering methods mentioned earlier use dynamic elements during their application to a set of static input data, i.e., still in the area of static clustering.

The optimization of initial cluster parameters is an issue that is especially important for static clustering since it directly impacts the final solution. In dynamic clustering, however, optimizing initial parameters is not as important as updating parameters during the respective cycles, which is much more relevant.

6. RST with machine learning algorithms for clustering

Machine learning methods have been applied for text-based document clustering [56] attempts to group documents into clusters to improve cluster purity and form clusters where each cluster might represent a label that is different from the labels of the other clusters. Document clustering algorithms are

Table 5
System developed for dynamic clustering

Dynamic Clustering Systems	Working Mode
Functional fuzzy c-means (FFCM) presented by Joentgen et al. [30]	The similarity between objects (and thereby their distance) is determined using fuzzy sets.
Rule-based fuzzy system	Dynamically-changing classes fuzzy clustering has been used, for example, as a preprocessing tool that iterates between fuzzy clustering and tuning of the fuzzy rule base.
Dynamic evolving neural-fuzzy inference system (DENFIS)	Used for dynamic time-series prediction [39], it is, constructed from the evolving clustering method (ECM), a first-order Takagi-Sugeno-type fuzzy rule set for prediction.
Dynamic data assigning assessment (DDAA), proposed by Georgieva and Klawonn.	A prototype-based clustering algorithm, the noise-clustering technique finds good single clusters one by one and simultaneously separates noisy data.
Kohonen self-organizing feature maps (SOFM), combined with rough set theory. [29]	The interval-set clustering provides interesting results in the setting of temporal analysis.
Mamat [14] pointed out that determining optimal initial parameters remains one of the key challenges, especially in rough clustering.	Weights of the approximations, as well as a threshold, have to be determined besides the number of clusters.

divided, in general, into two categories: partitional clustering and hierarchical clustering. Partitional clustering divides a document collection into groups in a single level, while hierarchical clustering creates a tree structure of documents. Agglomerative hierarchical clustering and iterative partitional clustering are two major categories of a clustering algorithm. In the following section, machine learning algorithms that have been implemented using the rough set theory concept are presented in detail.

6.1. Need for rough set clustering?

George Peters [61] discussed the need for rough set clustering and is there any improvement in cluster quality because of rough set approximation technique? The authors have apparently bought out the effectiveness of clustering objects, compared to the original k-means method. In *k*-means the number of correctly clustered objects is to be maximized which corresponds to minimizing the number of incorrectly clustered objects. However, in rough k-means the correctly and in-correctly clustered objects are not coherently connected thus in rough clustering the number of incorrectly clustered objects can be explicitly minimized. Many real time applications problems consider the task of minimizing the number of incorrectly clustered objects is more important than maximizing the number of correctly clustered objects.

6.2. Agglomerative clustering

Agglomerative clustering refers to the bottom-up approach where one starts with a small cluster and pairs of clusters are amalgamated while traversing through the hierarchy. Duo Chen et al. [21] presented the Rough Set-based Agglomeration Hierarchy Clustering Algorithm (RAHCA), which first lets $P = U/R\{d\} = \{\{x1\}, \ldots, \{xn\}\}$, and then conducts a merger of the clustering level and similarity measure. This operation can be carried out till the number of clusters m or the aggregate degree threshold λ given by users outputs the clustering results, i.e. clustering model P. This algorithm can also be conducted till all the objects are merged into a single cluster. It is in this way that the algorithm outputs the dynamic clustering map. A data table $(U,\ A)$ is prepared with a number of clusters m (or aggregate degree threshold (λ) and the output is the clustering model P. The equivalence classes $U/R\{ak\}$ should be first computed for the computation of Mk, and then $r \times |U|$ elements in Mk must be conducted, where r is the number of the clusters in P with the maximum $|U|$. The maximal number of iterations is $|U| - 1$, and each iteration needs to compute $LEV\ r$ times in the second step and $SIM\ r - 1$ times in the first step, both of which can be computed in $O(r|A||U|)$, and the updating operations in the previous step finished in $O(|A||U|)$, such that the time complexity of step 3 is $O(r|A||U|2)$. So then, the time complexity of the algorithm can be estimated as $O\ (|A||U|)$.

6.3. Rough K-means clustering

Zhao et al. [27] described various generalizations of rough sets by relaxing the assumptions of an underlying equivalence relation. Such a trend toward generalization is also evident in the rough mereology proposed by Polkowski et al, the use of information granules in a distributed environment by Skowron and Stepaniuk. The present study uses such a generalized view of rough sets. If one adopts a more restrictive view of rough set theory, the rough sets developed in this paper may have to be looked upon as interval sets.

The rough set dynamic clustering has been used to perform rough k-means. At the beginning of each cycle, we perform rough k-means and obtain parameters that are used to classify new objects and new data. In the second step, new data collected over a certain period – for instance, a season – is received and new data merged with the current dataset. The new data needs to be classified to check for structural changes, based on the results obtained in the previous step. It is decided that if the data structure has changed, objects of dying clusters must be deleted [55, 58]. A cluster is said to be dying when it does not receive a sufficient number of new objects. As a consequence, the author proposes to eliminate the dying cluster and its sure members. The number of clusters is to be updated to upgrade the initial parameters of rough k-means [37] as well. In rough clustering, we do not consider all the properties of rough sets [3]. However, the family of upper and lower approximations is required to follow certain basic rough set properties, such as:

An object $\mathbf{v}$ that is a member of a lower approximation of a set is also part of its upper approximation.

$$(\mathbf{v} \in \mathbf{A}(\mathbf{xi}) \rightarrow \mathbf{v} \in \mathrm{A}(\mathbf{xi}))$$

This implies that a lower approximation of a set is a subset of its corresponding upper approximation.

$$(\mathbf{A}(\mathbf{Xi}) \subseteq \mathbf{A}(\mathbf{Xi}))$$

If an object **v** is not part of a lower approximation, it belongs to two or more upper approximations.

The equation above implies that an object resides in a single boundary region rather than in multiple regions. We comprehend that while basic properties are not required to be either independent or complete, enumerating them is useful in understanding the rough set adaptation of the K-means algorithm. The rough K-means approach has been exposed by Peters [37] and miscellaneous deficiencies of Lingras [29] and West's original proposal highlighted. The alternative suggested by Peters is the use of ratios of distances, as opposed to differences between distances, similar to those used in the rough set-based Kohonen algorithm described in [29]. The use of ratios, rather than differences, is a better solution. Depending upon the values fed as input vectors, differences may vary but the ratios are not susceptible to the input values. Peters [37] has analysed and identified a significant additional modification to rough K-means for the improvement of the algorithm in sundry aspects.

The refined rough K-means algorithm makes calculations easy and simple for the centroid by ensuring that the lower bound of each cluster has at least one object. Further, it improves the quality of the clusters, as clusters with an empty lower bound have a limited basis for existence. Peters tested the refined rough K-means to analyze convergence and dependency on the initial cluster assignment. A study of the Davies–Boulden index shows that the boundary region can be interpreted as a security zone, as opposed to the unambiguous assignments of objects to clusters in conventional clustering. Despite all the experiments done, there is a need for additional areas where the rough K-means needs still more refinement, as in the selection of parameters.

Another interesting application is interval set clustering for web users, aims to provide analytical solutions in web mining, as clustering in the web mining needs to address varied set of issues. For this Pawan Lingras [9] have proposed a variation in k-means clustering using rough set. The author has done a modification of the K-means algorithm to create interval of clusters will provide an efficient method for representing clusters with vague and imprecise boundaries. The author applied the variation in k-means based rough set theory to cluster the web page visitor of sixteen week period time into three classes of upper and lower bound rough set to ease the process of web mining.

6.4. Mixture model and ensemble clustering

The EM clustering allows overlapping of clusters and thus the degree of uncertainty could be overcome using the rough set theory. The QuickReduct algorithm for determining the reduct set is proposed using a knowledge of the rough set theoretical background. The algorithm starts with an empty set and, in each iteration, adds the attribute that results in the greatest increase in the rough set dependency metric to the reduct set. Using rough set theory for feature reduction and the application of the mixture model is an effective representation of the probability density function and consists of k component density functions. The objective of a mixture model is to fit density functions to a given dataset to approximate data distribution. The EM algorithm can be used to solve the problem of mixture models where Θ is the model parameter and unknown-random variable $Y = \{y_i\}^N$ presents each object that belongs to a particular model. That means $y_i = k$ if the ith object belongs to the component k. The complete data log-likelihood expression for this density from data X and Y in this work is given by the Gaussian distribution used, alongside the EM algorithm to determine the mean (μ_l), covariance matrix (Σ_l) and sampling probability (W_l) for each cluster. The attribute set affects the distribution of data and leads to different model parameters.

$$\log(L(\Theta/X,\ Y)) = \log(P(X,\ Y/\Theta)) = b \quad (9)$$

$$\sum\nolimits_i^k \log\ (P(x_i|y_i)P(y)) = \sum\nolimits_i^k \log\ W_{yi}P(x_i|\theta_{yi}) \tag{10}$$

For appropriate clustering of the EM algorithm, the rough set has been applied for the feature selection phase which is the pre-processing procedure for improving cluster quality and reducing impurity.

6.5. Probabilistic rough set

For modelling probabilistic rough sets, rough membership functions and rough inclusion can be formulated with the use of conditional probabilities and threshold values. The membership function can be analysed for conditional probabilities or posterior probabilities. The parameterized approximations or probabilistic inclusion can be made using the parameters applied to a rough membership function. H. Zhang et al. analysed three probabilistic rough set models the decision-theoretic rough set model, the variable precision rough set model and the Bayesian rough set model [24] and proposed three probabilistic ones.

The primary differences among those models show how they are different in terms of equivalent formulations of probabilistic approximations and interpretations of rough approximations. Zaikro et al. proposed a Bayesian rough set model in an attempt to provide an alternative interpretation of the required prior probabilities defined. However, the two models designed have limitations in setting up new parameters for the loss function. In designing the Bayesian rough set model, dual decision approximation has been proposed which extends the one-parameterized approximation model by assigning a threshold value on a Bayesian confirmation measure. The Bayesian rough set model [24] and decision rough set model bring new insights into probabilistic rough set approximations.

6.6. Self-organizing map and rough set

The SOM-Kohonen self organizing map is a tool for pattern recognition that maps high dimensional space into a small number of dimensions by organizing similar elements close together to form clusters. Sap et al. [36] proposed a rough set-based SOM. They took two neurons, defined to be indiscernible, in the upper approximation and proposed a two-level clustering algorithm using SOM [36]. They found that the first-stage rough SOM is found to perform better and more accurately than the proposed crisp clustering method (incremental SOM) and reduces errors. An incremental clustering algorithm for dynamic information processing has been proposed where the data point is either added or deleted dynamically. Overlapped neurons are found with respect to SOM's rough set clustering and will be applied to SOM results to calculate errors and uncertainty. The aim of the proposed approach is making the rough set clustering of SOM as precise as possible.

From the rough set algorithm it can be observed that if two neurons are defined as indiscernible (those neurons in the upper approximation of two or more clusters), there is a certain level of similarity they have with respect to the clusters they belong to and that similarity relation has to be symmetric. Thus, the similarity measure must be symmetric. According to the rough set clustering of SOM, overlapped neurons and respectively overlapped data (those data in the upper approximation) are detected. In the experiments, to calculate errors and uncertainty, the previous Equations (5 and 6) will be applied to the results of SOM (clustered and overlapped data). The aim of the proposed approach is making the rough set clustering of the SOM to be as precise as possible. The approach is based on the rough set theory that employs a soft clustering which can detects overlapped data from the data set and makes clustering as precise as possible.

Bazan Jan. G [25] proposed dynamic reducts with large stability coefficients are "good" candidates for decision rule generation. They allow to construct rules with better classification quality of unseen objects than reducts with smaller stability coefficients. Intuitively, any dynamic rule is appearing in all (or almost all) of experimental subtables. The decision rules can be computed from the so called k-relative discernibility matrix used to generate decision rules with the minimal number of descriptors.

6.7. Advantages and disadvantages of modified rough set with machine learning approach for clustering

The drawback of the probabilistic rough set models is a lack of a systematic procedure for setting the required parameters. Although it is possible to link the loss function automatically, mathematically their physical meaning in terms of establishing parameters and the probability function has to be explored. While incorporating rough set with agglomerative clustering the concept of equivalence classes can be used to divide and merge classes. However predicting the number of rules to determine the uncertainty level leads to the increase in the number of examples.

Thus the agglomerative method uses a small set of examples to divide and merge the subclasses. Rough set theory focusses on the idea of using global properties to find the similarity between the objects in the form of initial coarse labels thus the clusters formed is unaffected by the local discrepancies. The Bayesian Rough set Model can be used when there is no user defined parameters available, as the relevant approximation region can be decided using the prior probabilities that are available.

The multi-decision problems can be easily addressed by the monotonic approximation quality measures availability. This machine learning model can be applied for medical diagnosis, fault detection, and economic forecasting problems. In the learning problem, the irrelevant and redundant features lead to low accuracy and slow learning which leads to an NP-complete problem, the rough set can be used to reduce the features which are redundant by removing the uncertainty in the data thereby improving the cluster purity and external cluster validity. Rough set, when combined with the Expectation Maximization algorithm, will yield lesser and more accurate features.

For analyzing the grouped categorical data, the total mean distribution algorithm and maximum total mean distribution algorithm can be used for clustering the features, implies the cohesion between the clusters should be high and the coupling between the clusters should be less. The above algorithm overcomes the Min-Min Roughness which uses the splitting criterion as the leaf node with more objects which in turn leads to undesirable clustering results. The rough set based agglomerative clustering algorithm works for the categorical data however it requires certain data from the users like mentioning the number of clusters and the aggregate threshold value, however, the algorithm does not work with mixed numeric and categorical data. The rough K-means does work well for the categorical dataset, which converges much earlier from global optimum to local objective function compared to other algorithms like c-means and fuzzy c-means. Evolutionary algorithms for clustering generates a set of clusters, however, the limitation is that the initial cluster

Table 6
Tools and techniques for rough set document clustering

Rough set theory		
Tools	Description	Usage
Rosetta	ROSETTA is a toolkit for analyzing tabular data within the framework of the rough set theory. ROSETTA is designed to support overall data mining from initial browsing and pre-processing of data via computation of minimal attribute sets and generation of all patterns for validation and analysis of induced rules or patterns.	The computational kernel is also available as a co33mmand-line program suitable for being invoked from, for example, Perl or Python scripts. http://www.lcb.uu.se/tools/rosetta/index.php
Rose2	Rose2 (Rough Sets Data Explorer) software implements basic elements of the rough set theory and rules discovery techniques. Created at the Laboratory of Intelligent Decision Support Systems of rough set-based knowledge discovery and decision analysis, Rose2 system is a successor of the Roughcast and RoughClass systems.	RoughDAS is, historically, one of the first successful implementations of the rough set theory. http://idss.cs.put.poznan.pl/site/rose.html
Rough Set Library	A rough set library (rsl.tar.Z) of public domain software is available from the Warsaw University of Technology, Institute of Computer Science, via ftp.	(jhs@ii.pw.edu.pl).
Cluto Toolkit	The CLUTO toolkit is a tool used for rough set working using the similarity rough set and WordNet similarity model.	The algorithms provided in CLUTO toolkit are based on the partitional, agglomerative, and graph-partitioning paradigms. They are denoted as rb, rbr, direct, agglo, graph, and bagglo.
Document clustering		
WEKA 3:Data Mining Software in JAVA	Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code.	Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. http://www.cs.waikato.ac.nz/ml/weka/
Carrot2	Carrot2, an open source search results clustering engine, can automatically organize small collections of documents (search results, but not only) into thematic categories.	Carrot2 is implemented in Java, but a native C#/ .NET API is also available. Other non-Java platforms, such as PHP or Ruby can call Carrot2 clustering through their REST interface.
Text-Garden	Text-mining software tools enable easy handling of text documents for the purpose of data analysis including automatic model generation and document classification, document clustering, document visualization, dealing with Web documents, crawling the Web and much more.	The code, written in *C*++, originally ran on the Windows platform and can be run on Linux/Unix using Wine or a similar utility. http://ailab.ijs.si/dunja/textgarden/
MALLET	MALLET is a Java-based package for statistical natural language processing, document classification, clustering, topic modeling, information extraction, and other machine learning applications to text.	MALLET includes tools for sequence tagging for applications such as named-entity extraction from text. http://mallet.cs.umass.edu/

Table 7
Applications of RST for document clustering

Author	Application	Methods	Analysis
Nguyen Chi Thanh et al. [39]	Document representation and clustering	TRS, SRT, Medline Abstract	The evaluation of the clustering result was 0.363, 0.332, 0.894 for entropy, mutual information, and *F*-measure respectively.
In-Kyoo Park et al. [40]	Job searches	VPRS, Entropy-based	The selected attribute is Welfare Program with a value of 0.3287. The selected attribute is Talent and Hobby with a value of 0.0883 because less entropy guarantees more stability.
Tuttut Herawan et al. [14]	Diabetics data clustering	VPRS, MMR, Clustering	The clustering of datasets gives us the overall cluster purity at 0.646909.
Zhifei Zhang et al. [42]	Social media-Short texts	Topic model, polysemy, synonyms	Performance has improved about 3.7% on the Search-Snippet dataset and 3.5% on the Web-Title dataset.
Alessia Albanese et al. [41]	Wisconsin Breast Cancer Dataset Clustering	Clustering, spatio-temporal outlier detection, kernel set	Computationally less intensive compared to other methods, thus supporting small to large datasets.
Can Gao et al. [44]	Biomedical UCI Data Clustering	K-modes, FCM, Ensemble	Average accuracy and CCI values of the random subspace ensemble algorithm increase by 10 and 7.3%, respectively.
Pawan Lingras et al. [16]	Highway section, Web user, Supermarket chain data clustering	Rough K-means clustering	Rough k-means clustering has been used to maximize enrichment for optimal clustering quality.
Siriporn Chimphlee [45]	Web access prediction	Rough clustering, the Markov Model	Access prediction using rough set and the Markov model fetches better results.
Nayak, Rudra Kalyan [46]	Gene expression data	Attribute clustering, Rough set	Comparatively, proficient results have been achieved by the proposed attribute clustering method.
Ming Chang Lee [47]	Customer value evaluation	Rough set, Entropy, Decision tree	The proposed method is useful for removing redundant attributes.
Li Yuan and Feipeng Xu [10]	Knowledge transfer Risk analyzing of IT outsourcing	Multiple weight-based, clustering analysis and similarity co-efficient matrix	Risk management has been reduced for IT outsourcing.
Cariappa M. M. & Mydhili K. Nair [12]	Genetic algorithm for web service composition	Evolutionary computational technique	GA suffers from an innate problem of increased execution time when the initial population (input data) is high, as well as decreased hit rate (success rate).
Raibei Mamat et al. [14]	Fourteen UCI datasets and a supplier dataset	Addressed the problem of clustering attribute selection using maximum attribute selection	Lower computational time than the other three rough set-based algorithms.
Famei He [56]	Movie-lens 100K dataset	Detecting attacks, push attack model, nuke attack model, and the recommender system	The best classification accuracy of average attacks is nearly 100%. LHS coverage = 0.972689, RHS coverage = 0.983015, confidence = 98%).
Pattaraintakorn et al. [57]	Orwik 2009/2010 OMOP Cup: Methods Competition	Web-based recommender health system, database of rules, database of facts and an inference engine	Use has been limited by confidentiality and reliance on a single system over human medicine.
Inbarani et al. [59]	Weblogs from msnbc.com	User profiling, web personalization	The Gaussian rough method yields an optimum number of clusters.
Weihua Liao [60]	GIS subzone discrete data	Similarity measurement, rough set, discrete and continuous data	The similarity of subzones is calculated with a threshold of.8 and above and neighbourhood rough sets used for calculating upper and lower approximations.
Tian-Wei Sheu et al. [41]	Taiwanese electronic companies were studied from 2001 to 2006. 30 had failed, 89 were healthy.	Similarity computation, rough set cluster analysis, and predictive accuracy	The accuracy of the approximation is 96.63%, while others are 95.51%, 89.89%, 89.89% and 89.89%. These empirical results show that the rough set model offers the best predictive accuracy.
Kevin Voges et al. [38, 62]	Shopping Oriented Data	Rough Clusters, Interpretation of shopping Data	Numbers above 0.90 suggest a strong positive orientation towards that particular aspect of shopping, while numbers below 0.90 suggest negative orientation

size has to be mentioned, which does not optimize to a feasible approach in large datasets. In rough set based evolutionary algorithm solution for clustering the template based description helps to cover the largest portion of the dataset. Two phase self-organizing map combined with rough set yields good clustering results for pattern recognition application. Rough set handles the dynamic incomplete information systems using incremental learning matrix by learning various extended relation

7. Tools and application of RST and document clustering

Table 6 depicts the tools that are available for implementing Rough Set Theory and Document Clustering, Multifarious tools for implementing clustering techniques using machine learning algorithms, rough set tools, and APIs have been used for clustering numeric and nominal datasets, presented in the following table. A brief description of the tool and its functionality has been listed in the table. Using the references set out in the following table would make it easier to implement rough set theory for document clustering tasks.

8. Applications of rough set theory

Rough set theory has been used for applications for clustering data from which inferences are drawn for knowledge gains, classification and searching tasks. Evaluation results from each application have been presented and the methods critically reviewed and categorized in the following table. Rough set clustering has been used in a variety of real-time applications in order to decide whether rough sets can be used to improve clustering quality without compromising on accuracy. Table 7 discusses the rough set theory being applied for various real time applications in research.

The above table discusses various real time applications where rough set clustering has been implemented to improve the clustering quality and cluster purity. The overall inference from the table is that the use of rough set does yield better results in clustering tasks, in varied applications might be, as the prior knowledge is not much needed, in finding the initial clusters. The rough clusters formed with a boundary approximations addressed various problems like detecting attacks for movie recommender systems, the taiwanese based companies for predictive analysis.

9. Conclusion and future work

An exhaustive review of rough set theory for document clustering is presented in this paper. The hybrid methods, designed with machine learning and rough set theory for document clustering is also discussed and the advantages of the methods listed in tabular form. Rough set clustering extensions – like the similarity rough set model, tolerance rough set model, and word similarity method – are presented along with the variable precision rough set model which works for clustering tasks to improve clustering quality by means of appropriate outlier detection. Rough set techniques for document clustering like min-min roughness and dynamic clustering have been analysed for their suitability to applications. Existing tools for rough set theory and document clustering are presented with website links and evaluation metrics commonly used for document clustering. Real-time applications using rough set theory for document clustering have been listed, providing insights for researchers using rough set theory for document clustering. Going forward, we plan to implement a novel algorithm for clustering, using rough set theory.

Apart from the complexity problem, it is worthy to note that applying ideas of the rough set model in cluster analysis seams to lead to stable and valid results in general. Furthermore, they are new insights how concepts such as “cluster”, “error” or “outlier” and even “validation” may be understood and defined. We think that this is a value on its own.

In future, we have planned to design a rough set based similarity measure for improving the cluster purity and cluster validity for entity resolution task. We are planning to explore how rough set based clustering works for big data with the less computational complex solutions for the large datasets which are highly heterogeneous.

Funding

I am greatly thankful for UGC in supporting this study, which was funded by DST PURSE UGC, NEW DELHI, INDIA (9500/PD2/2014-2).

References

[1] Z. Pawlak, Rough sets, *International Journal of Computer & Information Sciences* **11**(5) (1982), 341–356.

[2] X.T. Hu, T.Y. Lin and J. Han, A new rough sets model based on database systems, *Lecture Notes in Artificial intelligence* **2639** (2003), 114–121.

[3] Z. Pawlak, Rough set, *Communication of the ACM* **38**(11) (1995), 88–95.

[4] Z. Pawlak, Rough sets approach to knowledge based decision support, *European Journal of Operational Research* **99** (1997), 48–59.

[5] A. Pawlak, Rough sets, Rough sets and Data Mining. Kluwer Academic Publisher, Dordrecht. 1997b, pp. 3–8.

[6] T. Herawan, M.M. Deris and J.H. Abawajy, A rough set approach for selecting clustering attribute, *Knowledge-Based Systems* **23**(3) (2010), 220–231.

[7] T. Qu, J. Lu, H.R. Karimi and E. Xu, A novel research on rough clustering algorithm. In *Abstract and Applied Analysis*, vol. 2014, Hindawi Publishing Corporation, 2014.

[8] N. Parthalain, Q. Shen and R. Jensen, A distance measure approach to exploring the rough set boundary region for attribute reduction, *IEEE Transactions on Knowledge and Data Engineering* **22**(3) (2010), 305–317.

[9] P. Lingras and C. West, Interval set clustering of web users with rough k-means, *Journal of Intelligent Information Systems* **23**(1) (2004), 5–16.

[10] L. Yuan and F. Xu, Research on the multiple combination weight based on rough set and clustering analysis—the knowledge transfer risk in IT outsourcing taken as an example, *Procedia Computer Science* **17** (2013), 274–281.

[11] W.B. Michael, Survey of Text Mining: Clustering, *Classification and Retrieval*, 2007.

[12] M.M. Cariappa and M.K. Nair, Applying rough set theory to genetic algorithm for web service composition, *Int J Comput Sci Inform* **2**(4) (2012).

[13] Y. Terada, et al., On the possibility of structure learning-based scene character detector. *2013 12th International Conference on Document Analysis and Recognition*, IEEE, 2013.

[14] R. Mamat, T. Herawan and M.M. Deris, MAR: Maximum Attribute Relative of soft set for clustering attribute selection, *Knowledge Based System* **52** (2013), 11–20.

[15] Questier, I.A. Rollier, B. Walczak and D.L. Massart, Application of rough set theory to feature selection for unsupervised clustering, *Chemometrics and Intelligent Laboratory Systems* **63** (2002), 155–167.

[16] P. Lingras, Unsupervised rough set classification using GAs, *Journal of Intelligent Information Systems* **16**(3) (2001), 215–228.

[17] J. Starzyk, D.E. Nelson and K. Sturtz, Reduct generation in information systems, *Bulletin of International Rough Set Society* **3**(1/2) (1999), 19–22.

[18] P. Jaganathan, et al., Classification rule discovery with ant colony optimization and improved Quick Reduct algorithm, *IAENG International Journal of Computer Science* **33**(1) (2007), 50–55.

[19] S.S.R. Abidi and K.M. Hoe, Symbolic exposition of medical data-sets: A data mining workbench to inductively derive data-defining symbolic rules, *Computer-Based Medical Systems, 2002 (CBMS 2002) Proceedings of the 15th IEEE Symposium on IEEE*, 2002.

[20] W. Zhu and F.-Y. Wang, Reduction and axiomization of covering generalized rough sets, *Information Sciences* **152** (2003), 217–230.

[21] D. Chen, et al., A rough set-based hierarchical clustering algorithm for categorical data, *International Journal of Information Technology* **12**(3) (2006), 149–159.

[22] L. Sun, J. Xu and Y. Li, A feature selection approach of inconsistent decision systems in rough set, *Journal of Computers* **9**(6) (2014), 1333–1340.

[23] G. Peters, Outliers in rough k-means clustering. *International Conference on Pattern Recognition and Machine Intelligence*. Springer Berlin Heidelberg, 2005.

[24] H. Zhang, J. Zhou, D. Miao and C. Gao, Bayesian rough set model: A further investigation, *International Journal of Approximate Reasoning* **53** (2012), 541–557.

[25] J.G. Bazan, A comparison of dynamic and non-dynamic rough set methods for extracting laws from decision tables, *Rough sets in Knowledge Discovery* **1** (1998), 321–365.

[26] P. Lingras and G. Peters, Applying rough set concepts to clustering. *Rough Sets: Selected Methods and Applications in Management and Engineering*. Springer London, 2012, pp. 23–37.

[27] Y. Zhao, R.R. Ni and Z.F. Zhu, RST transforms resistant image watermarking based on centroid and sector-shaped partition, *Science China Information Sciences* **55**(3) (2012), 650–662.

[28] J.J. Alpigini, et al., eds. *Rough Sets and Current Trends in Computing: Third International Conference, RSCTC 2002*, Malvern, PA, USA, 2002. Proceedings. Vol. 2475. Springer, 2003.

[29] P. Lingras, M. Hogo and M. Snorek, Interval set clustering of web users using modified Kohonen self-organizing maps based on the properties of rough sets, *Web Intelligence and Agent Systems: An International Journal* **2**(3) (2004), 217–225.

[30] A. Joentgen, et al., Dynamic fuzzy data analysis based on similarity between functions, *Fuzzy Sets and Systems* **105**(1) (1999), 81–90.

[31] P. Zhu and Q. Wen, Entropy and co-entropy of a covering approximation space, *International Journal of Approximate Reasoning* **53**(4) (2012), 528–540.

[32] A. Daneshpazhouh and A. Sami, Entropy-based outlier detection using semi-supervised approach with few positive examples, *Pattern Recognition Letters* **49** (2014), 77–84.

[33] Z. Xue, Y. Shang and A. Feng, Semi-supervised outlier detection based on fuzzy rough C-means clustering, *Mathematics and Computers in Simulation* **80**(9) (2010), 1911–1921.

[34] X. Zhang, et al., Comparative study of variable precision rough set model and graded rough set model, *International Journal of Approximate Reasoning* **53**(1) (2012), 104–116.

[35] W. Zhang, J. Xiu, S. Mi and Z.W. Wei, Approaches to knowledge reductions in inconsistent systems, *International Journal of Intelligent Systems* **18**(9) (2003), 989–1000.

[36] M.N.M. Sap and E. Mohebi, Hybrid self organizing map for overlapping clusters, *International Journal of Signal Processing, Image Processing and Pattern Recognition* **1**(1) (2008), 11–20.

[37] G. Peters, Some refinements of rough k-means clustering, *Pattern Recognition* **39**(8) (2006), 1481–1491.

[38] K.E. Voges, N. Pope and M.R. Brown, Cluster analysis of marketing data examining on-line shopping orientation: A comparison of k-means and rough clustering approaches, *Heuristics and Optimization for Knowledge Discovery* (2002), 207–224.

[39] N.C. Thanh and Y. Koichi, Document representation and clustering with wordnet based similarity rough set model, *International Journal of Computer Science Issues (IJCSI)* **8**(5) (2011).

[40] I-K. Park and G.-S. Choi, A variable-precision information-entropy rough set approach for job searching, *Information Systems* **48** (2015), 279–288.

[41] T. Sheu, T. Chen, C. Tsai, J. Tzeng, C. Deng and M. Nagai, Analysis of students' misconception based on rough set theory, *Journal of Intelligent Learning Systems and Applications* **5**(2) (2013), 67–83. doi: 10.4236/jilsa.2013.52008

[42] Z. Zhang, D. Miao and X. Yue, Similarity measure for short texts using topic models and rough sets, *Journal of Computational Information Systems* **9**(16) (2013), 6603–6611.

[43] A. Albanese, S.K. Pal and A. Petrosino, Rough sets, kernel set, and spatiotemporal outlier detection, *IEEE Transactions on Knowledge and Data Engineering* **26**(1) (2014), 194–207.

[44] C. Gao, W. Pedrycz and D. Miao, Rough subspace-based clustering ensemble for categorical data, *Soft Computing* **17**(9) (2013), 1643–1658.

[45] S. Chimphlee, et al., Rough Sets Clustering and Markov model for Web Access Prediction. *Proceedings of the Postgraduate Annual Research Seminar*, 2006.

[46] R.K. Nayak, et al., Rough set based attribute clustering for sample classification of gene expression data, *Procedia Engineering* **38** (2012), 1788–1792.

[47] L. Ming-Chang, Customer value evaluation based on rough set with information gain and generate decision tree, *British Journal of Mathematics & Computer Science* **4**(15) (2014), 2123.

[48] D. Parmar, T. Wu and J. Blackhurst, MMR: An algorithm for clustering categorical data using rough set theory, *Data & Knowledge Engineering* **63**(3) (2007), 879–893.

[49] G. Peters, R. Weber and R. Nowatzke, Dynamic rough clustering and its applications, *Applied Soft Computing* **12**(10) (2012), 3193–3207.

[50] L.J. Mazlack, A. He and Y. Zhu, A rough set approach in choosing partitioning attributes. *Proceedings of the ISCA 13th International Conference (CAINE-2000)*.

[51] F. Jiang, Y. Sui and C. Cao, Some issues about outlier detection in rough set theory, *Expert Systems with Applications* **36**(3) (2009), 4680–4687.

[52] H. Cho and M.K. An, Co-clustering-based clustering and segmentation for pattern discovery from time course data, *International Journal of Information and Electronics Engineering* **4**(5) (2014), 358.

[53] M.J. Beynon and M.J. Peel, Variable precision rough set theory and data discretisation: An application to corporate failure prediction, *Omega* **29**(6) (2001), 561–576.

[54] M. Li, et al., Hierarchical clustering algorithm for categorical data using a probabilistic rough set model, *Knowledge-Based Systems* **65** (2014), 60–71.

[55] K.A. Vidhya and G. Aghila, Text mining process, techniques and tools: An overview, *International Journal of Information Technology and Knowledge Management* **2**(2) (2010), 613–622.

[56] F. He, X. Wang and B. Liu, Attack detection by rough set theory in recommendation system, *Granular Computing (GrC), 2010 IEEE International Conference on IEEE*, (2010), 692–695).

[57] P. Pattaraintakorn, G.M. Zaverucha and N. Cercone, Web based health recommender system using rough sets, survival analysis and rule-based expert systems. *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*. Springer Berlin Heidelberg, 2007.

[58] K.A. Vidhya and G. Aghila, Hybrid text mining model for document classification, *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on*, **1** (2010), 210–214.

[59] Inbarani, H. Hannah and K. Thangavel, Rough set based User profiling for Web Personalization, *International Journal of Recent Trends in Engineering* **3**(10) (2009).

[60] W. Liao, The rough method for spatial data subzone similarity measurement, *Journal of Geographic Information System* **4** (2012), 37–45. http://dx.doi.org/10.4236/jgis.2012.41006, Published Online January 2012 (http://www.SciRP.org/journal/jgis).

[61] G. Peters, Is there any need for rough clustering? *Pattern Recognition Letters* **53** (2015), 31–37.

[62] K. Voges, N. Pope and M. Brown, A rough cluster analysis of shopping orientation data. *Proceedings Australian and New Zealand Marketing Academy Conference*, Adelaide. 2003.4.