International Journal of Computer Applications (0975 – 8887)
Volume 56– No.10, October 2012

Rough Set Approach in Machine Learning: A Review

Prerna Mahajan,
Research Scholar
Bansthali University,Rajasthan
, India

Rekha Kandwal,
India Meteorological Department,
Ministry of Earth Sciences & Science and Technology,New Delhi, India.

Ritu Vijay
Department of Electronics,AIM & ACT
Bansthali University,Rajasthan

ABSTRACT

The Rough Set (RS) theory can be considered as a tool to reduce the input dimensionality and to deal with vagueness and uncertainty in datasets. Over the years, there has been a rapid growth in interest in rough set theory and its applications in artificial intelligence and cognitive sciences, especially in research areas such as machine learning, intelligent systems, inductive reasoning, pattern recognition, data preprocessing, knowledge discovery, decision analysis, and expert systems. This paper discusses the basic concepts of rough set theory and point out some rough set-based research directions and applications. The discussion also includes a review of rough set theory in various machine learning techniques like clustering, feature selection and rule induction.

General Terms

Information and decision systems, (in)discernibility , approximation spaces, rough sets, rough membership functions, reducts, decision rules, dependencies of attributes. Clustering, Rule Induction, Feature Selection

Keywords

Clustering, Rule Induction, Feature Selection

1. INTRODUCTION

The concept of rough sets was introduced by Pawlak [77], as an extension of set theory in early eighties. It is an approach to approximate concepts under uncertainty. The theory has been widely used for attribute selection, data reduction, rule discovery, genetics and many knowledge discovery applications in the areas such as data mining, machine learning and medical diagnoses [42, 54,63,64]. One may regard the theory of rough sets to be complementary to other generalizations of set theory, such as fuzzy sets and multisets [34,78,81,125]. In recent years, there has been a fast growing interest in this new emerging theory. The successful applications of the rough set model in a variety of problems have amply demonstrated its usefulness and versatility [79,109,110,126,136]. It is turning out to be rationally significant to artificial intelligence and cognitive science, especially in the representation of and reasoning with vague and/or imprecise knowledge, machine learning, knowledge acquisition, decision analysis, knowledge discovery from databases, expert systems and pattern recognition [23,41,46,59,65,66,82,109]. It is of particular importance to decision support systems and knowledge discovery. Unlike many other approaches, the main advantage of RST is that it does not need any preliminary or additional data about information systems.

The main objective of this paper is to present an extensive review of the rough set based approaches for knowledge discovery. We discuss the basic mathematical constructs and terminology of RST. We also present the various quality metrics of RST proposed in research for handling uncertainty and efficient classification. The discussion also includes a review of rough set theory in various classification techniques like clustering, feature selection and rule induction.

The foundation of RST is on the fact that, in the universe of discourse, every object is associated with some information. For eg. if students have failed in an examination, marks of the examination form information about students, objects that can be characterized by the same information are considered similar to each other, in view of the available information about them. This similarity (Indiscernibility) relation forms the basis of RST. The set of all similar (indiscernible) objects is called a crisp (precise) set, otherwise the set is called rough (imprecise or vague). Consequently, each rough set has boundary-line cases i.e., objects which can not with certainty be classified either as members of the set or of its complement [77]. This means that boundary-line cases cannot be properly classified by employing available knowledge. RST is a promising approach to deal with qualitative information and provides an approach based on an individual object [60].

2 Rough set Theory: Terminology & Notations

Rough sets analyze uncertainty in data. They were used to determine the crucial attributes of objects and build the upper and lower approximate sets of objects sets[15]. In real world data varies in size and complexity, which is difficult to analyze and also hard to manage from computational view point. The major objectives of Rough Set analysis are to reduce data size and to handle inconsistency in data. The following section discuss the major concepts of rough sets used to approximate inconsistent information and to handle redundant data.

2.1 Information Table

In Rough set data model, information is stored in a table where a fact or an object is represented by single row (tuple) .The information about the real world object is represented by the corresponding tuple in the table. Each column in the table represents an attribute (a variable, a property, etc) that can be measured for each object. Such a table is called an information system.

It can be represented as a pair $IS=(U, A)$, where, $U=\{x_1,x_2,\ldots,x_n\}$ is a non empty finite set of objects called the universe and $A=\{a_1,a_2,\ldots,a_m\}$ is a non- empty finite set of attributes such as a $U \rightarrow V_a$, $a \in$. A. The set v_a is called the value set of a. We can split the set of attributes in two subsets $C \subset A$ and D=A- C, respectively the conditional set of attributes and the decision (or *class*) attribute(s). Condition attributes represent measured features of the objects, while the decision attribute is an *a posteriori* outcome of classification

An example of information table is given in Table 1:

Columns of the table are labeled by attributes outlook, Temperature, Humidity, Windy and Play and rows by players(objects) $p_1, p_2, p_3, \ldots, p_{14}$.

Each row of the table can be seen as information about specific player. For example player P3 can be characterized in the table by the following attribute-value set

{(Outlook,overcast),(Temperature,83),(Humidity,86),

(Windy, False),(Play, Yes)}

Table 1: An example of an Information table

	Outlook	**Temperature**	**Humidity**	**Windy**	**Play**
P_1	sunny	85	85	FALSE	no
P_2	sunny	80	90	TRUE	no
P_3	overcast	83	86	FALSE	yes
P_4	rainy	70	96	FALSE	yes
P_5	rainy	68	80	FALSE	yes
P_6	rainy	65	70	TRUE	no
P_7	overcast	64	65	TRUE	yes
P_8	sunny	72	95	FALSE	no
P_9	sunny	69	70	FALSE	yes
P_{10}	rainy	75	80	FALSE	yes
P_{11}	sunny	75	70	TRUE	yes
P_{12}	overcast	72	90	TRUE	yes
P_{13}	overcast	81	75	FALSE	yes

2.2 Indiscernibility relation

It is the starting point of rough set theory. It intends to express the fact that due to the insufficient knowledge we are unable to distinguish (discern) some objects using the available information. It is one form of redundancy in data. Two tuples are indiscernible with respect to each other if 'A' decision table is an information system IS= (U, A), for every set of attributes 'B $\subseteq$ A', an equivalence relation, denoted by IND_{IS} and called the B- indiscernibilty relation , is defined by:

$$IND_B(U) = \{ (x, y) \in U^2 \mid \forall a \in B \; a(x) = a(y) \} \quad (1)$$

If $(x,y) \in IND_{IS}$, then objects x and y are indiscernible from each other by attributes from B. We can easily prove that indiscernibility is an equivalence relation. We denote the B-indiscernibility class of x as $[x]_B$. Equivalence relations lead to the universe being divided into equivalence class partition and union of these sets make the universal set.

Let us observe that each subset of attributes divides the set of all objects in the tables into classes having the same features i.e clumps of objects which are indiscernible in view of the available data. For example, in the Table 1, players p_3, p_7, p_{12}, p_{13} are indiscernible in terms of the attribute outlook. Thus each subset of attributes induces on the set of objects an equivalence relation, whose equivalence classes form granules (clusters, groups) of objects having the same features. These clusters will be referred to as elementary sets, which are basic building blocks of rough set theory.

2.3 Equivalence Relation

Let R be an equivalence relation over U, then the family of all equivalence classes of R is represented by U/R . $[x]_R$ means a category in R containing an element x $\in$ U. Consider P $\subseteq$R, and P$\neq \phi$, then IND(P) is an equivalence relation over U. For any x $\in$ U, the equivalence class of the relation IND(P) is denoted as $[x]_P$.

2.4 Approximation of Rough Sets

The equivalence relation and the induced equivalence classes provide the available information or knowledge about the objects under consideration. The primary idea of the rough sets theory is the approximation spaces and lower and upper approximations of a set. A rough set is defined through its lower and upper approximation. Let X be a concept such that X$\subseteq$U, X can be approximated using only the information contained within B by constructing the B-lower and B-upper approximations of X:

$$\underline{B}X = \{x \in U : [x]_B \subseteq X\} \quad (2)$$

$$\overline{B}X = \{x \in U : [x]_B \cap X \neq \phi\} \quad (3)$$

$$BN_B(X) = \overline{B}X - \underline{B}X \quad (4)$$

Where $\underline{B}X$ and $\overline{B}X$ is called as the B-lower and B-upper approximations of X, respectively [31].

Based on the lower and upper approximations of a set X$\subset$U, the universe U can be divided into three disjoint regions, the positive region POS(X), the negative region NEG(X), and the boundary region BND(X):

$$POS(X) = \underline{apr}(X),$$

$$NEG(X) = U - \overline{apr}(X),$$

$$BND(X) = \overline{apr}(X) - \underline{apr}(X), \quad (5)$$

We can say that any element $x \in POS(X)$ certainly belongs to X, and that any element $x \in NEG(X)$ does not belong to X whereas the upper approximation of a set X is the union of the positive and boundary regions, namely, $\overline{apr}(X) = POS(X) \cup BND(X)$. We cannot decide with certainty whether or not an element $x \in BND(X)$ belongs to X. For arbitrary element $x \in \overline{apr}(X)$ one can only conclude that x possibly belongs to X.

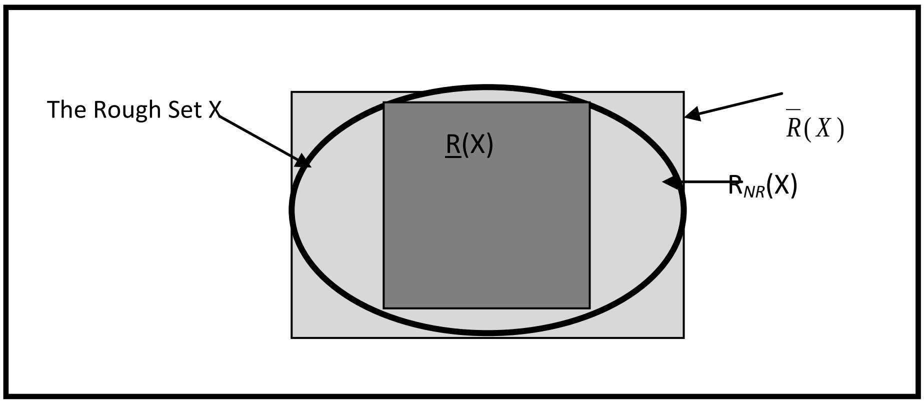

Figure 1 : Rough Set Model [77]

2.5 Rough Membership Function

Rough sets can be also defined employing, instead of approximation, rough membership function .The rough membership function expresses conditional probability that x belongs to X given R and can be interpreted as a degree that x belongs to X in view of information about x expressed by R[80].

$\mu^{R}_{X} : U \rightarrow < 0, 1 >$

$$\mu^{R}_{X}(x) = \frac{|X \cap R(x)|}{|R(x)|} \quad (6)$$

And $|X|$ denotes the cardinality of *X*.

Rough set theory deals with the concept of vagueness and uncertainty clearly, which are otherwise very often confused terms. ***Vagueness*** is the property of sets and can be described by approximations, whereas ***Uncertainty*** is the property of elements of a set and can be expressed by the rough membership function. Rough sets can be defined in two ways as given below [81]

Def 1: Set X is rough with respect to R if $\underline{R}X \neq \overline{R}X$.

Def 2: Set X is rough with respect to R if for some x : $0 \leq \mu^{R}_{X}(x) \leq 1$

2.6 Reducts & Core It refers to a subset of attributes which can, by itself, fully characterize the knowledge in the database. , which means we can remove some superfluous data from information system while preserving its basic properties. A reduct can be thought of as a *sufficient* set of features – sufficient, that is, to represent the category structure.

Let C, D $\subseteq$ A, be sets of condition and decision attributes respectively. We will say that C′ $\subseteq$ C is a D-reduct (reduct with respect to D) of C, if C′ is a minimal subset of C such that

$\gamma(C,D) = \gamma(C',D)$

The intersection of all D-reducts is called a D-core (core with respect to D). Because the core is the intersection of all reducts, it is the set of attributes which is possessed by *every* legitimate reduct and therefore consists of attributes which cannot be removed from the information system without causing collapse of the equivalence-class structure. The core may be thought of as the set of *necessary* attributes – necessary, that is, for the category structure to be represented.

2.7 Functional dependence For given A= (U,A), C,D ⊆ A, by C→D is denoted the *functional dependence* of D on C in A that holds iff IND(C) ⊆ IND(D). In particular, any B-reduct C determines functionally D. Also dependencies to a degree are considered [79].

2.8 Decision systems and rules Decision systems can be used to match classification of objects by an expert with a classification in terms of accessible features. A *decision system* is a tuple A^d=(U,A,d), where (U,A) is an information system with the set A of condition attributes, and the decision (attribute) d: U→V_d, where d∉A. In case A → d holds in A^d, we say that the decision system A^d is deterministic and the dependency A→ d is A^d-exact. Then, for each class $[x]_A$ there exists a unique decision d(x) throughout the class. Otherwise, the dependency A→ d in A^d holds to a degree. A *decision rule* in A^d is any expression $\wedge\{a=v_a : a\in A \text{ and } v_a \in V_a\} \rightarrow d=v$ where d is the decision attribute and $v\in V_d$. This decision rule is true in (U,A,d) if for any object satisfying its left hand side it also satisfies the right hand side, otherwise the decision rule is true to a degree measured by some coefficients[79]. Strategies for inducing decision rules can be found in [89,90].

2.9 Definable and rough concepts (sets) Classes of the form $[x]_B$ can be regarded as the primitive B-*definable concepts w*hose elements are classified with certainty by means of attributes in B[129]. This property extends to more general concepts, i.e., a concept X ⊆ U, is B-*definable* iff for each y in U, either $[y]_B \subseteq U$ or $[y]_B \cap X = \emptyset$. This implies that X has to be the union of a collection of B-indiscernibility classes, i.e., $X = \cup \{[x]_B : x \in X\}$. Then we call X a B-*exact* (*crisp*, *precise*) concept. One observes that unions, intersections and complements in U to B-exact concepts are B-exact as well, i.e., B-exact concepts form a Boolean algebra for each B⊆A. In case when a concept X is not B-exact, it is called B-*rough*, and then X is described by *approximations* of X that are exact concepts [79] , i.e., one defines the B-*lower approximation* of X, and the B-*upper approximation* of X by $B_*(Y) = \{x\in X: [x]_B \subseteq X\}$ and $B^*(Y) = \{x\in X: [x]_B \cap X\neq\emptyset\}$, respectively. The set $B^*(Y) - B_*(Y)$ is called the B-*boundary region* of X.

$$\overline{R}X = \cup\left\{x \in U \left| [x]_R \cap X \neq \varphi \right.\right\} \quad (7)$$

2.10 Rough mereology The approach based on inclusion functions was generalized to the *rough mereological approach* [75,88,90]. The inclusion relation $x\mu_r y$ with the intended meaning "x *is a part of y to a degree at least* r" has been taken as the basic notion of the *rough mereology* that is a generalization of the Leśniewski mereology. Rough mereology offers a methodology for synthesis and analysis of complex objects in distributed environment of intelligent agents, in particular, for synthesis of objects satisfying a given specification to a satisfactory degree or for control in such complex environment. Moreover, rough mereology has been recently used for developing foundations of the *information granule calculi* [75], aiming at formalization of the Computing with Words and Perceptions paradigm, recently formulated in [131]. More complex information granules are defined recursively using already defined information granules and their measures of *inclusion* and *closeness*. Information granules such as classifiers [35] or approximation spaces can have complex structures. Computations on information granules are performed to discover relevant information granules, e.g., patterns or approximation spaces for complex concept approximations.

3 Quality Metrics of Rough Sets Theory

RST offers various metrics for the analysis of information systems. A good measure will help in deciding various parameters used in analysis. Accuracy, quality of approximation and mean etc. are major representatives of these techniques.

3.1 Accuracy of Approximation

Accuracy is a metric which tells how much a set is rough. If a set has $\underline{B}X = \phi$ and $\overline{B}X = U$, the approximation tells nothing about X, because for any element $x \in U$ we can not decide whether $x \in X$ or not. If, on the contrary $\underline{B}X = \overline{B}X = X$, the set is crisp (precise), means for every element $x \in U$, we certainly know if $x \in X$ or not. Accuracy of a rough set is expressed by the following formula:

$$\alpha_B(X) = \frac{|\underline{B}(X)|}{|\overline{B}(X)|} \quad (8)$$

Accuracy value for rough sets is $0 \leq \alpha_B(X) \leq 1$, and for crisp (precise) sets $\alpha_B(X) = 1$.

The accuracy of roughness given above can also be interpreted using the well-known Marczeweski-Steinhaus(MZ) metric[127]. MZ metric when applied to the lower and upper approximations of a subset $X \subseteq U$ in information system *IS*, we get the following equation:

$$D(\underline{B}X, \overline{B}X) = 1 - \frac{|\underline{B}(X) \cap \overline{B}(X)|}{|\underline{B}(X) \cup \overline{B}(X)|} = 1 - \frac{|\underline{B}(X)|}{|\overline{B}(X)|} = 1 - \alpha_B(X) \quad (9)$$

The notion of the dependency of attributes in information systems is given in the following definition.

3.2 Quality of Approximation. The following coefficient expresses the percentage of objects which can be correctly classified into class *X*.

$$\gamma(X) = \frac{|R'_*(X)|}{|X|} \quad (10)$$

Moreover,

$0 \leq \alpha(X) \leq \gamma(X) \leq 1$ and $\gamma(X) = 1$ if $\alpha(X) = 1$ [102]. *R'* is a similarity relation, from the indiscernibility relation *R* by relaxing the original conditions for indiscernibilty[111].

3.3 Dependency of Attributes

An Information system can be represented as a pair IS=(U, A), where, $U=\{x_1,x_2,\ldots,x_n\}$ is a non empty finite set of objects called the universe and $A=\{a_1,a_2,\ldots,a_m\}$ is a non- empty finite set of attributes such as a $U \rightarrow Va$, $a \in A$. The set Va is called the value set of a. We can split the set of attributes in two subsets $C \subset A$ and D=A- C, respectively the conditional set of attributes and the decision (or *class*) attribute(s). Condition attributes represent measured features of the objects, while the decision attribute is an *a posteriori* outcome of classification. Formally dependency between attributes can be defined in the following way:

(i) D depends on C with a degree *k* where k is ($0 \leq k \leq 1$) i.e $k = \gamma(C,D)$

(ii) If *k* =1, then D depends totally on C that is all elements of universe *U* can be uniquely classified to various classes of *U/D* employing *C*.

(iii) If *k is* ($0 \leq k \leq 1$) then D is partially dependent on , that is only some elements of the universe *U* can be uniquely classified to various classes of *U/D* employing *C* [79].

$$\gamma(C,D) = \frac{|POS_C(D)|}{|U|} \quad (11)$$

where

$$POS_c(D) = \cup_{x \in U/D} \underline{C}X$$

3.4 Mean Roughness

It is defined by Mazlack et.al.[72] as the average roughness of all sub-partitions of an attribute *i*

Given $a \in A$, $V(a_i)$ refers to the set of values of attribute a_i, X is a subset of objects having one specific value, α, of attribute a_i, that is, $X\ (a_i = \alpha)$, $\underline{X_{a_j}(a_j = \alpha)}$ refers to the lower approximation, and $\overline{X_{a_j}(a_j = \alpha)}$ refers to the upper approximation with respect to $\{a_j\}$,

Then $R_{a_j}(X)$ is defined as the roughness of X with respect to $\{a_j\}$, that is

$$R_{a_j}(X \mid a_i = \alpha) = 1 - \frac{\left|\underline{X_{a_j}(a_j = \alpha)}\right|}{\left|\overline{X_{a_j}(a_j = \alpha)}\right|}, \text{ where } a_i, a_j \in A$$

and $a_i \neq a_j$ (12)

Let $\left|V(a_i)\right|$ be the number of values of attributes ai, the mean roughness on attribute a_i with respect to $\{a$j$\}$ is defined as

$$\text{Rough}_{a_j}(a_i) = \frac{R_{a_j}(X \mid a_i = \alpha_1) + \ldots R_{a_j}(X \mid a_i = \alpha_{|V(a_i)|})}{\left|V(a_i)\right|}$$

(13)

The lower the mean roughness is, the higher the crispness of the sets.

3.5 Min- Roughness (MR)

Defined by Parmar et al. [76] MR, min- roughness of attribute a_i ($a_i \in A$) for given n attributes, refers to the minimum of the mean roughness i.e,

MR(a_i)= Min (Rough a_i (a_i) ,.., Rough a_j (a_i),…),

where a_i, $a_j \in A$, $a_i \neq a_j$, $1 \leq$ i ,j $\leq$ n (14)

Min- roughness (MR) determines the best crispness each attribute can achieve.

3.6 Min- Min- Roughness (MMR)

Defined by Parmar et al.[76] the MMR is the minimum of the Min- Roughness of the n attributes for given n attributes i.e.

MMR = Min (MR(a_1), ……, MR(a_i),…) where $a_i \in A$, i goes from 1 to $|A|$. (15)

MMR determines the best split on the attributes, which divides the group of objects resulting in better crisp sets (clusters).

3.7 Total Roughness (TR)

Defined by Mazlack et al.[72], Total Roughness is used to represent the overall crispiness of a partitioning towards every attribute [72]. It ranges from 0 to 1. The larger the *Total Roughness (k),* the crisper the partition.

The partitioning attributes will be chosen based on *Total Roughness.*

Total Roughness(k)=

$$\frac{\left(\prod_{i=1 \text{ to } k}^{m} Rough(i)\right)}{m} \quad (16)$$

where m is the number of attributes.

3.8Rough Involvement Function. It is very similar to rough membership function, but can also reflect the involvement of objects to classes. It is the ratio which quantifies the percentage of objects correctly classified into the class X which are related to the object x [6].

$$\nu_X(x) = \frac{\left|X \cap R'(x)\right|}{|X|} \quad (17)$$

4. Related Approaches

Rough set theory supplies essential tools for knowledge analysis. It allows for creating algorithms for knowledge reduction, concept approximation, decision rule induction, and object classification. In the next section we discuss the adoption of rough set theory in various classification techniques.

4.1 Rough Set Theory in Feature Selection

Feature selection plays a vital role in data mining. It focuses on the most important features necessary for the data representation and rejects the irrelevant features. Feature selection aims at finding the optimal subset of features of a data according to some criterion. The main objectives of removing the irrelevant features and selects only the relevant one are defined as follows:[26]

1) Noisy variables are detrimental to the generalization of learning algorithms, as the algorithms expand computational effort to train on variables with low signal-to-noise ratios.

2) So-called 'deceptive 'variables may mislead learning algorithms into generalizing on the wrong underlying concepts.

Feature selection has mainly two objectives:

- Maximize information content in the selected subset of variables.
- Minimize the cardinality of that subset.

These requirements complicate the task of Feature Selection (FS) algorithms. Many Feature selection algorithms based on forward selection (adds variables incrementally until the desired selection quality is achieved) and backward elimination (starts with the entire set of variables and incrementally removes variables till the quality remains consistently high, whereas Bidirectional hill climbing allows the addition or removal of a variable at any given stage, as needed to maximize quality [26]. There are basically two categories of Feature Selection algorithms:

Filters: are pure preprocessors. They rely on evaluating the information content of variables, and thus draw heavily from Information Theory. Filters are very generic but employ no knowledge of the classifying properties of the data [33]. However the filter approach is ineffective in dealing with the feature redundancy. Some of the algorithms in the Filter approach methods are Relief[56],Focus[4], Las Vegas Filter (LVF)[70], Selection Construction Ranking using Attribute Pattern (SCRAP)[93], Entropy-Based Reduction (EBR) [55], Fractal Dimension Reduction (FDR) [114] .

Wrappers: work in combination with a classifier. They determine the quality of subsets of variables on the basis of how efficiently those variables classify the training samples. Wrappers are more accurate approach than filters however they lack efficiency and generality in comparison to filters. Some of the Wrapper approach methods are LasVegas Wrapper (LVW) and Neural network-based feature selection [70,71].

Feature Selection is one of the important aspect in Rough set theory which uses the concept of reduct for feature reduction. More formally, a reduct is a minimal subset of attributes $B \subset A$ such that IND(B) = IND(A), where IND(X) is the X- indiscernibility relation. A reduct is a minimal subset of attributes $B \subseteq A$ such that it preserves the partitioning of universe and hence has the ability to perform classification [58]. The concept of reducts in the feature selection and reduction of attributes has been studied and used by various authors [3,25,57,58,75,134]. Rough sets have been extensively used for feature selection. Their use has been proposed in various contributions [10, 11, 79,113]. The primitive approach is to determine the core subset for discrete attribute dataset, which contains strongly relevant features and reducts, also a subset of core and weakly relevant features, so that each reduct is sufficient to determine the concepts described in data set. Reducts can be further used for feature selection for example a minimal reduct would be a reduct containing a minimal set of attributes. Concept of dynamic reducts was proposed by [10,11] in order to find a more robust and generalized feature subset. The selection of dynamic reduct is based on the cross- validation method. The methods of dynamic reducts generation have been used for dynamic selection of relevant features as well as in the process of selection of relevant decision rules. Some other methods based on non invasive data analysis and rough sets are reported in [36].

Many good methods of calculating reducts have been developed, some of them are based on genetic algorithms, which allows the calculation of reducts with an acceptable computational cost[122,123,124] and others based on heuristic methods[13, 29,30,31].Another evolutionary approach for feature selection based on RST proposed by Caballero et al.[24]. Two algorithms are presented namely epigraph2 based on evolutionary method and epigraph3, greedy algorithm with heuristic approach. Another RST based feature selection approach is given by Zhang and Yao[133] namely PASH(Parametrized Average Support heuristic). This algorithm considers the overall quality of the potential set of rules. It selects features causing high average support of rules over all decision classes. In addition it also has the parameters that are used to adjust the level of approximations. Methods based on feature weighting and instance selections based on rough set theory have been given by Salamo and Colobardes[100,101,102,103].

4.2 Rough Set Theory in Clustering

Clustering is regarded as a fundamental task in data mining which groups the similar objects in the same cluster. Clustering is being used in various data analysis tasks such as unsupervised classification, data summation and in data segmentation which divides large datasets into smaller homogeneous subsets (clusters) that can be easily managed, classified separately and analyzed. To date many researchers have worked on various clustering techniques for data with categorical, continuous or mixed attributes.

Rough Clustering is an emerging technique which is based on a simple extension of rough sets theory to cluster analysis, and applicable where group membership is unknown. Rough clustering solutions allow multiple cluster membership of objects. In this section we discuss the research work done in the area of rough clustering.

Clustering based on Rough set theory can be achieved by mapping the clustering dataset to the decision table. The basic concept of representing a set as lower and upper approximations of rough sets can be used in a broader context such as clustering. For rough clustering an appropriate distance measure should be used such that the strict requirement of indiscernibility relation used in normal clustering is relaxed [37]. Rough clustering has been used successfully in forestry[86], medicine [49, 86], imaging[73], web mining[68], supermarkets[69] and traffic engineering applications[67].

Rough sets are used to develop efficient heuristics searching for relevant tolerance relations that allow extracting objects in data. Rough sets are used to develop efficient heuristics searching for relevant tolerance relations that allow extracting objects in data. An attribute-oriented rough sets technique reduces the computational complexity of learning processes and eliminates the unimportant or irrelevant attributes so that the knowledge discovery in database or in experimental data sets can be efficiently learned. Using rough sets, has shown to be effective for revealing relationships within imprecise data, discovering dependencies among objects and attributes, evaluating the classificatory importance of attributes, removing data re-abundances, and generating decision rules Kusiak[60,61]. Some classes, or categories, of objects in an information system cannot be distinguished in term of available attributes. They can only be roughly, or approximately, defined.

Rough set theory can be used to represent the overlapping clusters. Rough sets provide more flexible representation than conventional sets, at the same time they are less descriptive than the fuzzy sets. Rough clusters extends the crisp(precise) notion of cluster, that is in rough clusters some objects are located at the lower approximation of a cluster that is objects that only belong to that cluster implying full membership to it, while others are laid at its upper approximation that is objects which are also members of other clusters. In this way rough cluster manages uncertainty about membership of objects to clusters. In recent years, there has been a fast growing interest in this new emerging theory, few of the successful results of rough clustering are discussed here.

Mazlack et al.[72] have proposed a rough set technique for selecting a cluster attribute. They have given two techniques namely Bi-clustering and Total Roughness (TR) technique which are based on the bi-valued attribute and maximum total roughness in each attribute set. Another successful rough set based clustering technique given by Parmar et al.[76] is MMR(Minimum-Minimum Roughness). This technique is based on lower & upper & quality of approximation of sets [83]. Another technique Maximal Attributes Dependency (MADE) is proposed by Herawen et al.[47], for categorical data clustering. This technique calculates rough attribute dependencies in categorical valued Information System and is used to select clustering attribute based on the maximum degree. Another significant hierarchical clustering algorithm for categorical data based on RST is given by Chen et al.[28]. Authors have proposed an attribute membership matrix is introduced and clustering level is calculated using consistent degree and aggregate degree of the clusters found. The similarity among clusters is calculated using categorical similarity measure based on Euclidean distance.

Upadhyaya, Arora and Jain[117] have proposed a rough set based indiscernibility relation combined with indiscernibility graph which results in discovery of natural clusters in data. In this method objects are grouped on the basis of similarity rather than being identical. Hakim, Winarko and Winarko[44] have proposed a method of clustering binary data based on the combination of indiscernibility and its indiscernibility level. Herawan, Yanto and Deris[48] rough set approach for clustering has been used for supplier base management.

Besides the approaches discussed above, several related approaches to rough clustering have been proposed. These related approaches mainly includes class of Rough Partitive Algorithms which include switching regression models,where the clusters are represented by functions instead of objects[92]. Peters and Lampart[84] suggested rough *k*medoids and Peters[85] also proposed a rough switching regression model, which—together with the rough *k*means—form a class of rough partitive algorithms. Other related approaches include Genetic Algorithm Based Rough Clustering. There are three versions of the GA based rough clustering, first one proposed by Lingras[67], another one by Mitra et.al,[73] and an evolutionary *k*-medoid by Peters et al.[87]. Kohonen Network Based Rough Clustering incorporates rough sets into the Kohonen algorithm which requires an addition of the concept of lower and upper approximations in the equations, which are used for updating the weights of the winners [68].

Rough Support Vector Clustering (RSVC) is a soft clustering method derived from the SVC paradigm [7]. It achieves soft data clustering by a natural fusion of rough set theory and SVC. In RSVC, theQP problem involved in SVC is modified to impart a rough set theoretic flavor. The modified QP problem obtained for RSVC turns out to be the same as the one involved in SVC. Therefore, the existing solution strategies used for solving the SVC–QP problem can be used for solving the RSVC–QP problem as well. The cluster labeling method of RSVC is a modified version of the one used in SVC.

Peters and Weber [86] proposed a dynamic approach to rough clustering where the initial parameters of the algorithm are updated in cycles to better adapt to changing environments like the seasonal changes in customer behavior. Several further approaches to rough clustering have been proposed. They include early approaches to clustering based on the set interpretation of rough sets by do Prado et al.[32] and Voges et al.[118,119]. Yao et al.[128], suggested to relax some of the properties of rough clustering, in particular the need for the membership, to at least two clusters of objects in boundary areas, and introduced an interval-based clustering approach.

4.3 Rough Set Theory in Rule Induction

Decision tree induction (ID3, C4.5 and its later versions), Bayesian approach, back propagation neural networks, rough set framework, and evolutionary algorithms are some of the important classification techniques to discover the decision rules. Rule discovery methods received a great deal of attention and were reported in many papers and surveys. Commonly known algorithms of discovering association rules by Agrawal et al.[1], Agrawal and Srikant[2], Zaki[132] and Han et al.[45] are based on using parameters of support and confidence - the most popular measures of interest and significance. These factors are actually hidden and do not occur in the association rules explicitly. Moreover, the traditional methods do not differentiate between average and very strong rules, which exhibit the deep relations between variables under consideration.

In order to find out more strong association rules from the set of all discovered ones, an extra effort is needed [9,14]. Rough set theory offers another kind of inductive learning in designing rough set decision rules from data written in the form of attribute tables [77,91]. The decision rules can be generated as certain or approximate ones. However, the level of uncertainty cannot be exposed, same as in the previous case of association rules. There are different algorithms of managing incomplete data, i.e. data with missing attribute values, when designing rough set decision rules [40]. It has proved itself successful in automated reasoning of rule-based systems. It deals with the theory of uncertain reasoning in order to model human-like reasoning problems of real life. Uncertainty, vagueness, ambiguity, and impreciseness are invariably some of problems found in relationships between attributes of real world systems, and these can be taken into account effectively by rough set theory. In recent past, rough set theory has found high degree of applicability in development of the rule-based systems.

RST is a mathematical approach to managing vague and uncertain data or problems related to information systems, indiscernibility relations and classification, attribute dependence and approximation accuracy, reduct and core attribute sets, and decision rules [108]. By using the data analysis concepts of ‘‘reduct” and ‘‘core”, the patterns or internal structures of a set of condition-decision data records can be easily reduced and extracted as a set of minimal decision rules without using any prior knowledge[77].

RST identifies the meaningful decision rules, in two steps. Firstly, the attribute reduction algorithm pre-processes rule induction. For this, it removes redundant information or features and selects a feature subset that has the same discernibility as the original set of features. This approaches aims at identifying subsets of the most important attributes influencing the raw data. For example, Hu et al.[51], computed the significance of an attribute using heuristic ideas from discernibility matrices and proposed a heuristic reduction algorithm (DISMAR). Hu[50] gave a rough set reduction algorithm using a positive region-based attribute significance measure as a heuristic (POSAR). Wang and Li[120] developed a conditional information entropy reduction algorithm (CEAR). Wang et al.[121] proposed a rough set attribute reduction algorithm that incorporated a search method based on particle swarm optimization (PSO) on brain glioma data to find minimal rough set reducts. Nguyen

[74] presented a heuristics approach based on Boolean reasoning to analyzing structure of malicious decision tables.

Secondly, a rough set rule induction algorithm generates decision rules, which can reveal profound knowledge and provide new insights [120]. For example, Tsumoto[115] introduced an approach to knowledge acquisition, which induced probabilistic rules based on rough set theory (PRIMEROSE) and developed a program that extracts rules for an expert system from a clinical database. Tsumoto[116] also proposed PRIMEROSE4.5 (Probabilistic Rule Induction Method based on Rough Sets Ver 4.5) as an extension of earlier version of PRIMEROSE4 reported by Tsumoto. In the earlier work of Tsumoto, only rigid set-inclusion relations were considered for grouping, while rough-inclusion relations were introduced in the second approach, allowing it to outperform the earlier approach [115]. The LEM2 algorithm was proposed to extract a minimum set of decision rules, and the rule induction algorithm was useful for both classification and medical knowledge discovery [39,53,112]. This algorithm could reveal regular and interpretable patterns of the relations between glioma MRI features and the degree of malignancy, which were helpful for medical evolution.

Law and Au [62] presented an approach which included rough classification, information system (IS), information reduction, and decision rules induction to model the relations in a set of mixed numeric and non-numeric data on tourism shopping. Shen and Chouchoulas[106] proposed a highly modular framework for data-driven fuzzy rule set induction called rough set attribute reduction (RSAR) incorporating a dimensionality-reduction step based on rough set theory.

The incremental technique is a way to solve the issue of add-in data. Previously proposed version of this technique include an incremental protocol design system that contains an incremental protocol verification technique and an Estelle translator [52], an incremental learning algorithm for classification [130], ordered incremental training for GA-based classifiers [135], a neural network architecture for incremental learning [107], continuous and incremental data mining association rules using a frame metadata model [38], a statistics- based approach to control the quality of sub clusters in incremental gravitational clustering [27].

There are also numerous studies including incremental rough set theory. For example, Blaszczynski and Slowinski[22] proposed a new RST method of incremental rule induction, called DomAprioriUpp, which is a technique of post-processing of decision rule sets. Asharaf et al.[8] proposed a novel RS-based incremental approach to clustering interval data. Bazan et al.[12] showed how among formulas used for classifier construction from decision rules can be need to search for new patterns relevant for the incremental concept approximation of rough set theory. Richards and Compton[99] described Ripple-Down Rules (RDR) and its approach to verification and validation, concentrating particularly on recent extensions which use Rough Set Theory for verification and Formal Concept Analysis for validation . Guo et al.[43] proposed a novel incremental rules extraction algorithm called ‘‘RDBRST” (Rule Derivation Based On Rough sets and Search Tree) [43]. Shan and Ziarko[105] proposed an incremental RS learning algorithm, although it does not support inconsistent data. To solve this problem, Bian[21] presented an improved algorithm based on Shan’s algorithm , using an extended decision matrix to deal with inconsistent data that cannot be solved by the Shan and Ziarko algorithm [105].

Most of the above discussed techniques discover high level symbolic rules in the form of simple Production Rules (PRs) (If Premise Then Decision). Though PRs are simple to interpret and implement, they fail on the grounds of exception handling and approximate reasoning. PRs provide decision rules at a single conceptual level and ignore conceptual hierarchies among the classes of the dataset being mined. Moreover, the PRs fragment the discovered knowledge into large number of rules reducing the overall comprehensibility of the rule set. The above deficiencies of PRs have been recently addressed by discovering the decision rules in the form of Censored Production Rules (CPRs), Hierarchical Production Rules (HPRs), Hierarchical Censored Production Rules (HCPRs) and Hierarchical Censored Production Rules with Fuzzy Hierarchy (CPRFHs) [5,16, ,17,18,19,20,94,95,96,97,98].

5. Summary

Rough set theory mainly deals with methods to classify imprecise, uncertain, and incomplete information or knowledge expressed in terms of data acquired from experience. It mainly differentiates between objects that may definitely be classified into a certain category and those that may possibly be classified. It allows for creating algorithms for knowledge reduction, concept approximation, decision rule induction, and object classification.

This paper presents a review of the research done in rough set theory. Different research is focused on one or more methods in rough sets. We have focused on the adoption of rough set theory in data preprocessing, clustering and rule induction. However, there is still room for us to further investigate and develop. Accommodating uncertainty in data incurs an additional computational overhead. In particular, the exhaustive application of rules seems quite time-consuming and expensive. A number of simplifications and evolutionary approaches of rough set for the decision-making process have been proposed and researchers are working on it. Research on integrating RST with other contemporary techniques granular computing neural network, genetic algorithms, evolutionary methods is going on and can be found in literature.

6. REFERENCES

[1] Agrawal, R., Imielinski, T. and Swami, A. Mining association rules between sets of items in large databases. in *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*,(Washington, D.C,1993), ACM Press ,22(2), 805-810.

[2] Agrawal, R. and Srikant, R. Fast algorithms for mining association rules in large databases. in J. B. Bocca, M. Jarke, and C. Zaniolo,(Eds.), *Proceedings of the 20th International Conference on Very Large Data Bases, VLDB,(* Santiago, Chile ,1994), Morgan Kaufmann , 487-499.

[3] Ahn,B.S., Cho, S.S. and Kim C.Y. The integrated methodology of rough set theory and artificial neural networks for business failure predictions. *Expert Systems with Applications,* 18, 2,(2000),65-74.

[4] Almuallim,H. and Dietterich,T.G. Learning with many irrelevant features. in *Proceedings of the 9th National*

Conference on Artificial Intelligence, *(* Anaheim, California ,1991), AAA*I* Press,547-552.

[5] Al-Maqaleh, B. M. and Bharadwaj, K. K. Genetic programming approach to hierarchical production rule discovery. *Transactions on Engineering, Computing and Technology*,6(2005), 271–274.

[6] Arco, L., Bello, R., Caballero, Y. and Falcon, R. Rough Text Assisting Text Mining: Focus on Documents Clustering Validity. In R. Bello,R. Falc N , W. Pedrycz (Eds.) *Granular Computing at the Junction of Rough Sets and Fuzzy Sets* , (Heidelberg, 2008) , Springer-Verlag ,229-248.

[7] Asharaf, S., Shevade, S.K. and Murty, N.M. Rough support vector clustering. *Pattern Recognition*, 38(2005), 1779–1783.

[8] Asharaf, S., Shevade, S.K. and Murty, N.M. Rough set based incremental clustering of interval data, *Pattern Recognition Letters*, 27 (2006), 515-519.

[9] Baqui,S.,Just,J. and Baqui,S.C. Deriving strong association rules using a dependency criterion, the lift measure. *International Journal of Data Analysis. Technical Strategy,* 1,3(2009),297–312.

[10] Bazan, J., Skowron, A. and Synak, P. Dynamic reducts as a tool for extracting laws from decision tables. in *Proceedings of the Symposium on Methodologies for Intelligent Systems, Lecture Notes in Artificial Intelligence-869,* (Berlin ,1994),Springer-Verlag,346-355.

[11] Bazan, J. A comparison of dynamic and non-dynamic rough set methods for extracting laws from decision system. In Polkowski, L., Skowron, A. Eds. *Rough Sets in Knowledge Discovery* , Physica-Verlag, Heidelberg ,1998, 321–365.

[12] Bazan, J. G., Peters, J. F., Skowron, A. and Nguyen, H. S. Rough set approach to pattern extraction from classifiers. *Electronic Notes in Theoretical Computer Science*, 82,4(2003),1–10.

[13] Bell, D. and Guan, J. Computational methods for rough classification and discovery. *Journal of the American Society for Information Science*, 49,5(1998), 403-414.

[14] Berzal,F.,Blanco,I.,Sanchez,D. and Vila,M. A. A new framework to assess association rules. in *Symposiom on Intelligent Data Analysis, Lecture Notes in Computer Sciences.* 2189(2001), 95–104.

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

[16] Bharadwaj, K. K. and Varshneya, R. Parallelization of hierarchical censored production rules. Information and Software Technology, 37,8(1995),453-460.

[17] Bharadwaj, K. K. and Al-Maqaleh, B. M. Evolutionary approach for automated discovery of censored production rules.*Enformatika*, 10(2005), 147–152.

[18] Bharadwaj, K. K. and Al-Maqaleh, B. M.Evolutionary approach for automated discovery of censored production rules with fuzzy hierarchy. In *Proceedings of the International Multi-Conference of Engineers and Computer Scientists*(Hong Kong,2007), Newswood Limited ,716-721.

[19] Bharadwaj, K. K. and Ratnoo,S. Parallel Genetic algorithm approach to automated discovery of hierarchical production rules. In J. Mehnen (Ed.) ,*World online international conference on soft computing and applications (WSC 2008), Applications of Soft Computing*, Springer, Berlin/Heidelberg , 2008,505-408.

[20] Bharadwaj,K.K. and Ratnoo,S. A parallel genetic programming based intelligent miner for discovery of censored production rules with fuzzy hierarchy. Expert System and Applications, 37,6(2010), 4601-4610.

[21] Bian, X.Certain rule learning of the inconsistent data. *Journal of East China Shipbuilding Institute*,12,1(1998),25–30.

[22] Blaszczynski, J. and Slowinski, R. Incremental induction of decision rules from dominance-based rough approximations. *Electronic Notes in Theoretical Computer Science*, 82,4(2003),1–12.

[23] Bonikowski, Z. and Wybraniec-Skardowska, U. Vagueness and Roughness. Transactions on Rough Sets-IX,5390, Springer-Verlag ,Heidelberg - Berlin, Germany,2008,1-13.

[24] Caballero,Y., Bello,R., Alvarez,D. and García,M.M. Two new feature selection algorithms with Rough Sets Theory, in Bramer, M. (ed.). *Artificial Intelligence in Theory and Practice* , Springer, Santiago, Chile ,2006,209-216.

[25] Carlin, U., Komorowski,J. and Ohrn, A. Rough set analysis of medical datasets and A case of patient with suspected acute appendicitis. *European Conference of Artificial Intelligence(ECAI 98) ,Workshop on Intelligent data analysis in medicine and pharmacology,* (Chichester, 1998),John Wiley and Sons,18-28.

[26] Caruana, R. and Freitag, D. Greedy attribute selection. in *Proceedings of the 11th International Conference on Machine Learning,(* NJ, USA,1994) ,Morgan Kaufmann,28-36.

[27] Chen, C. Y., Hwang, S. C. and Oyang, Y. J. A statistics-based approach to control the quality of subclusters in incremental gravitational clustering. *Pattern Recognition*, 38(2005), 2256–2269.

[28] Chen,D., Cui, D.W. ,Wang, C.X. and Wang, Z.R. A Rough Set-Based Hierarchical Clustering Algorithm for Categorical Data. *International Journal of Information Technology*, 12,3(2006),149-159.

[29] *Choubey,S.K.*, Deogun, J.S., Raghavan, V.V. and Sever, H.A comparison of feature selection algorithms in the context of rough classifiers.in *Proceedings of the Fifth IEEE International Conference on Fuzzy Systems*, (New Orleans, LA, USA,1996),IEEE Conference Publications,1122-1128.

[30] Deogun, J.S., Raghavan, V.V. and Sever, H. Exploiting upper approximations in the rough set methodology. In U.M. Fayyad and R. Uthurusamy(Eds.), *Proceedings of First International Conference on Knowledge Discovery and Data Mining*, (Canada, 1995),AAAI Press,69-74.

[31] *Deogun*, J.S., Choubey, S. K., Raghavan, V. V. and Sever, H. Feature selection and effective classifiers.

Journal of the American Society for Information Science, 49, 5 (1998), 423-434.

[32] Do Prado H.A, Engel ,P.M. and Filho, H.C. Rough clustering: an alternative to find meaningful clusters by using the reducts from a dataset. In J. Alpigini, J. Peters, A. Skowron, N. Zhong(Eds.)*Proceedings of the Rough Sets and Current Trends in Computing (RSCTC'02), Lecture Notes in Artificial Intelligence, LNAI- 2475*, (Heidelberg, 2002), Springer-Verlag, 234–238.

[33] Dong,J.Z. Zhong,N. and Ohsuga,S. Using rough sets with heuristics to feature selection. In N. Zhong, A. Skowron, and S.Ohsuga(Eds.), *Proceedings of the 7th International Workshop on New Directions in Rough Sets, Data Mining, and Granular-Soft Computing (RSFDGrC-99), Lecture Notes in Artificial Intelligence, Berlin,LNAI*- 1711, (Heidelberg, 1999), Springer-Verlag, 178–187.

[34] Dubois, D. and Prade, H. Rough fuzzy sets and fuzzy rough sets. *International Journal of General Systems*, 17(1990), 191-209.

[35] Duda , R. , Hart , P. and Stork , D. *Pattern classification.2* nd *edition* . John Wiley & Sons, New York, 2001 .

[36] Duentsch, I. and Gediga, G. Statistical evaluation of rough set dependency analysis. *International Journal* of *Human-Computer Studies*, 46(1997), 589–604.

[37] Emilyn, J. and Ramar, K.Rough Set Based Clustering Of Gene Expression Data: A Survey .*International Journal of Engineering Science and Technology*,2,12(2010),7160-7164.

[38] Fong, J., Wong, H. K. and Huang, S. M. Relevance feedback icon continuous and incremental data mining association rules using frame metadata model. *Knowledge-Based Systems*, 16(2003), 91–100.

[39] Grzymala-Busse ,J.. LERS A system for learning from examples based on rough sets. *Intelligent Decision Support Handbook of Applications and Advances of the Rough Sets Theory* , (Dordrecht, 1992) , Kluwer,3-18.

[40] Grzymala-Busse,J. and Grzymala-Busse,W. An experimental comparison of three rough set approaches to missing attribute values. *Transactions on Rough Sets, Lecture Notes in Computer Science*,LNCS, 6(2007),31-50.

[41] Grzymala-Busse, J. W. Three Approaches to Missing Attribute Values: A Rough Set Perspective. *Data Mining: Foundations and Practice* , Studies in Computational Intelligence,vol.118, Springer-Verlag ,Heidelberg , Germany , 2008,139—152.

[42] Grzymala-Busse J.W. Mining Numerical Data - A Rough Set Approach .*Transactions on Rough Sets-XI*, 5946(2010), 1--13.

[43] Guo, S., Wang, Z. Y., Wu, Z. C. and Yan, H. P.A novel dynamic incremental rules extraction algorithm based on rough set theory. in *Proceedings of the fourth International Conference on Machine learning and cybernetics, (*Guangzhou, 2005), IEEE Computer Society,18-21.

[44] Hakim,F., Winarko,S. and Winarko,E. Clustering Binary Data Based on Rough Set Indiscernibility Level. *Biomedical Soft Computing and Human Sciences*,16,2 (2010),87-95.

[45] Han,J.,Pei,J.,Yin,Y. and Mao,R .Mining frequent patterns without candidate generation. *Data Mining and Knowledge Discovery*, 8(2004),53–87.

[46] Hassanien ,A. E., Suraj Z., Slezak, D. and Lingras, P. Rough Computing. Theories, Technologies, and Applications, series: In memoriam Professor Zdzislaw Pawlak, IGI Global Hershey ,New York,2008.

[47] Herawan ,T., Ghazali,R., Yanto, I.T.R. and Deris, M.M. Rough Set Approach for Categorical Data Clustering. *International Journal of Database Theory and Application*, 3,1(2010), 33-52.

[48] Herawan, T., Yanto, I. and Deris, M..ROSMAN: ROugh Set approach for clustering Supplier base MANagement. *Biomedical Soft Computing and Human Sciences*, 162(2010),105-114.

[49] Hirano, S. and Tsumoto, S. Rough clustering and its application to medicine. *Information Sciences*, 124(2000), 125–137.

[50] Hu, X. Knowledge discovery in databases: An attribute-oriented rough set approach. Ph.D. Thesis, Regina University,1995.

[51] Hu, K. Y., Lu, Y. C. and Shi, C. Y. Feature ranking in rough sets. *Artificial Intelligence Communications*, 16,1(2003),41–50.

[52] Huang, C. M., Hsu, J. M., Lai, H. Y., Huang, D. T. and Pong, J. C.An Estelle-based incremental protocol design system. *Journal of Systems and Software*, 36, 2(1997),115–135.

[53] Jan, P., Grzymala-Busse, J. W. and Zdzislaw, S. H. Melanoma prediction usingData mining system LERS. In *Proceedings of the 25th Annual International Computer Software and Applications Conference (COMPSAC'01)*, (Chicago, IL, USA,2001), IEEE Computer Society,615-620.

[54] Jaworski W. Rule Induction: Combining Rough Set and Statistical Approaches. in *Proceedings of the Sixth International Conference on Rough Sets and Current Trends in Computing (RSCTC'2008)*,Ohio, USA, Lecture Notes in Artificial Intelligence- 5306, (Ohio, USA,2008),Springer-Verlag,170-180.

[55] Jensen, R. and Shen,Q. A Rough Set – Aided system for Sorting WWW Book-marks. In N.Zhong et al.(Eds.), *Web Intelligence: Research and Development*, Springer,Maebashi , Japan,2001,95-105.

[56] Kira, K. and Rendell,L.A. The Feature selection Problem: Traditional Methods and a New Algorithm.in *Proceedings of Association for the Advancement of Artificial Intelligence(AAAI*),(SanJose, CA,USA,1992), MIT Press,129-134.

[57] Kohavi, R. and Frasca, B. Useful feature subsets and Rough set Reducts. In *Proceedings of the Third International Workshop on Rough Sets and Soft Computing,(*, San Jose, CA,1994), San Jose State University ,310-317.

[58] Komorowski ,J. Pawlak ,Z. ,Polkowski ,L. and Skowron, A. A Rough set perspective on Data and Knowledge. In Klosgen, W. and Zytkow, J. (Eds). *The Handbook of Data mining and Knowledge discovery,* (Oxford, 1999),Oxford University Press,134-149.

[59] *Kryszkiewicz, M.*Rough Set Approach to Incomplete Information Systems. *Journal of Information Sciences. 112(1998), 39–49.*

[60] Kusiak, A.Feature transformation methods in data mining. *IEEE Transaction on Electronics Packaging Manufacturing*, 24,3(2001),214–221.

[61] Kusiak, A.Rough set theory: A Data Mining tool for semiconductor manufacturing. IEEE Transactions on Electronics Packaging Manufacturing , 24 ,1(2001) , 44-50 .

[62] Law, R. and Au, N. Relationship modeling in tourism shopping: A decision rules induction approach. *Tourism Management*, 21(2000), 241–249.

[63] Li, D., and Zhang, W. Gene selection using rough set theory. in *Proceedings of 1st International Conference on Rough Sets and Knowledge Technology(RSKT 2006),* (Chongqing, China,2006), Springer-Verlag,778-785.

[64] Li, J.,Pattaraintakorn,P. and Cercone, N . Rule Evaluations, Attributes, and Rough Sets: Extension and a Case Study. In Peters, J.F.; Düntsch, I.; Grzymala-Busse, J.; Orlowska, E.; Polkowski, L. (Eds.) *Transactions on Rough Sets VI: Commemorating Life and Work of Zdislaw Pawlak, Part I*, LNCS 4374, Springer-Verlag ,Berlin, Heidelberg ,2007,152-171.

[65] Liang, J.Y. and Xu, Z.B. The algorithm on knowledge reduction in incomplete information *systems*. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* ,24(2002), 95–103.

[66] Lingras, P. and Yao, Y. Data mining using extensions of rough set model.*Journal of American Society for Information Science* , 49,5(1998),415–422 .

[67] Lingras P. Unsupervised rough set classification using Gas. *Journal of Intelligent Information System*, 16(2001), 215–228.

[68] Lingras, P., Hogo, M. and Snorek, M. Interval set clustering of web users using modified Kohonen self-organizing maps based on the properties of rough sets. *Web Intelligence and Agent System: An International Journal*, 2,3(2004),217–230.

[69] Lingras, P., Hogo, M., Snorek, M. and West, C. Temporal analysis of clusters of supermarket customers: conventional versus interval set approach. *Information Sciences* , 172(2005), 215–240.

[70] Liu,H. and Setiono,R. Feature selection and classification – a probabilistic wrapper approach. In *Proceedings of the 9th International Conference on Industrial and Engineering Applications of AI and ES*, (Fukuoka, Japan,1996), Gordon and Breach Science Publishers,419-424 .

[71] Liu,H. and Setiono,R. Neural Network feature selector. *IEEE Transactions on Neural Networks*, 8 ,3(1997),645-662.

[72] Mazlack, L.J., He, A., Zhu, Y. and Coppock, S.A rough set approach in choosing partitioning attributes. In *Proceedings of the ISCA 13th, International Conference of* Computer Applications in Industry and Engineering(*CAINE-2000), (*Hawaii, USA,2000),ISCA,1-6.

[73] Mitra, S. An evolutionary rough partitive clustering. *Pattern Recognition Letters*, 25(2004), 1439–1449.

[74] Nguyen, H. S. On the decision table with maximal number of reducts. *Electronic Notes in Theoretical Computer Science*, 82,4(2003),pp. 1–8.

[75] Pal, S.K., Polkowski, L. and Skowron, A. (eds.) .*Rough--Neural Computing- Techniques for Computing with Words*. Springer-Verlag ,Berlin,2004.

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

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

[78] Pawlak, Z. Rough sets and fuzzy sets. *Fuzzy Sets and Systems*, 17(1985),99-102.

[79] Pawlak ,Z. *Rough Sets: Theoretical Aspects of Reasoning About Data,* Kluwer Academic Publishers, Boston, MA,1991.

[80] Pawlak, Z. and Skowron,A. Rough membership functions, In R. R Yaeger, M. Fedrizzi and J. Kacprzyk (eds.).*Advances in the Dempster Shafer Theory of Evidence*(pp. 251-271), John Wiley & Sons, Inc, Toronto, Singapore,1994,251-271.

[81] Pawlak, Z. Hard and soft sets . *In* W.P. Ziarko(Ed.), *Rough Sets, Fuzzy Sets and Knowledge Discovery*, Springer-Verlag ,London,1994,130-135.

[82] Pawlak, Z., Grzymala-Busse, J.W., Slowinski, R. and Ziarko, W. Rough Sets. *Communications of ACM* , 38,11(1995),89–95.

[83] Pawlak, Z. and Skowron, A. Rudiments of rough sets. *Information Sciences*, 177 ,1(2007),3–27.

[84] Peters, G. and Lampart, M. A partitive rough clustering algorithm. *Proceedings of the Fifth International Conference on Rough Sets and Current Trends in Computing (RSCTC'06)*, Lecture Notes in Artificial Intelligence, LNAI-4259, (Kobe, Japan,2006), Springer,657-666.

[85] Peters, G.Rough clustering and regression analysis.in *Proceedings of 2007 IEEE Conference on Rough Sets and Knowledge Technology (RSKT'07)*, Lecture Notes in Artificial Intelligence, LNAI-4481, pp.292–299 (Toronto, Canada,2007), John Wiley & Sons Inc,292-299.

[86] Peters, G. and Weber, R. A dynamic approach to rough clustering. in *Proceedings of the Seventh International Conference on Rough Sets and Current Trends in Computing (RSCTC'08)*, Lecture Notes in Artificial Intelligence, LNAI-5306, (USA,2008), Springer,379-388.

[87] Peters, G., Lampart, M. and Weber, R. Evolutionary rough k-medoid clustering. *Transactions on Rough Sets* VIII, Lecture Notes in Computer Science, 5084(2008),289–306.

[88] Polkowski, L. and Skowron, A. Rough mereology: a new paradigm for approximate reasoning. *International Journal of Approximate Reasoning*, 15(1996), 333–365.

[89] Polkowski, L. and Skowron, A. (eds.). *Rough Sets in Knowledge Discovery I & II*'. , Physica-Verlag, Heidelberg,1998.

[90] Polkowski, L., Tsumoto, S. and Lin, T.Y. (eds.). Rough Set Methods and Applications. *New Developments in Knowledge Discovery in Information Systems,* Physica-Verlag ,Heidelberg,2000.

[91] Predki,B., Slowinski, R.,Stefanowski, J.,Susmaga,R. and Wilk,Sz. ROSE. Software implementation of the rough set theory. *Lecture Notes in Computer Science*, 1424(1998),605–608.

[92] Quandt, R. The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of American Statistics Assocociation,* 53(1958), 873–880.

[93] Raman,B. and Loerger,T.R. Instance-based filter for feature selection. *Journal of Machine Learning Research*, l(2002),1-23.

[94] Ratnoo,S. and Bharadwaj, K. K. A Parallel Genetic Algorithm approach to automated discovery of censored production rules. in *Proceedings of IASTED International Conference on Artificial Intelligence and Applications (AIA07)*,(Innsbruck, Austria,2007), ACTA Press, 435–441.

[95] Ratnoo,S. and Bharadwaj, K. K. Distributed mining of censored production rules in data streams. In *Proceedings of the 7th WSEAS International Conference on Artificial intelligence, Knowledge engineering and Databases (AIKED08)*, (Cambridge, UK,2008), University of Cambridge, 500– 505.

[96] Ratnoo,S. and Bharadwaj, K. K. A Parallel genetic programming approach for automated discovery of censored production rules with fuzzy hierarchy. *In Proceedings of International Conference on Information Processing*, (Bangalore, India,2008), Springer,624-633.

[97] Ratnoo,S. and Bharadwaj,K.K. Discovery of Exceptions: A Step towards Perfection,NSS In Y. Xiang, J. Lopez, H. Wang and W. Zhou (Eds.),*Third International Conference on Network and System Security*,(Queensland, Australia,2009), IEEE Computer Society,540-545.

[98] Ratnoo,S and Bharadwaj,K.K. Combining Rough Set and Evolutionary Approach for Automated Discovery of Censored Production Rules with Fuzzy Hierarchy. Journal of Intelligent Systems ,19,1(2010), 47-78.

[99] Richards, D. and Compton, P. An alternative verification and validation technique for an alternative knowledge representation and acquisition technique. *Knowledge-Based Systems*, 12,1–2(1999), 55–73.

[100] Salamo, M. and Golobardes, E. Analysing rough sets weighting methods for case-based reasoning systems. Journal of Artificial Intelligence,1,15(2002), 34–43.

[101] Salamo, M. and Golobardes, E. Deleting and building sort out techniques for case base maintenance, In Craw, S., Preece, A. (Eds.). *Proceedings of the. 6th European Conference on Case-Based Reasoning,* (Scotland, UK,2002), Springer,365-379.

[102] Salamo, M. and Golobardes, E. Unifying weighting and case reduction methods based on rough sets to improve retrieval. in *Proceedings of the 5th International. Conference on Case-Based Reasoning*, (Trondheim, Norway ,2003), Springer,494-508.

[103] Salamo, M. and Golobardes, E. Global, local and mixed case base maintenance techniques. In *Proceedings of the 6th Congres Catala d'Intel_ligencia Artificial*, (Amsterdam, Netherlands,2004) , IOS Press ,127–134.

[104] Salton, G.The SMART retrieval system. Prentice-Hall, Englewood Cliffs ,1971.

[105] Shan, N. and Ziarko, W. An incremental learning algorithm for constructing decision rules. In R. S. Kluwer (Ed.). *Rough sets, fuzzy sets and knowledge discovery, (*Berlin, 1994), Springer-Verlag,326-334.

[106] Shen, Q. and Chouchoulas, A. A modular approach to generating fuzzy rules with reduced attributes for the monitoring of complex systems. *Engineering Applications of Artificial Intelligence*, 13,3(2000),263–278.

[107] Shiotani, S., Fukuda, T. and Shibata, T. A neural network architecture for incremental learning. *Neurocomputing*, 9(1995), 111–130.

[108] Shyng, J.-Y., Wang, F.-K., Tzeng, G.-H. and Wu, K.-S. Rough set theory in analyzing the attributes of combination values for the insurance market. *Expert Systems with Applications*, 32,1(2007),56–64.

[109] Skowron, A. and Stepaniuk, J. Intelligent systems based on rough set approach. in *Proceedings of the International Workshop Rough Sets: State of the Art and Perspective,Extended Abstracts, (*Kiekrz, Poland,1992),IDSR,62-64.

[110] Slowinski, R. (ed.). Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory. Kluwer Academic Publishers, Boston,1992.

[111] Slowinski, R. and Vanderpooten, D. Similarity relations as a basis for rough approximations. In *P.P.* Wang, (Eds), *Advances in Machine Intelligence and Soft-Computing*, (NC, USA,1997), Duke University Press,17–33.

[112] Stefanowski, J. On rough set based approaches to induction of decision rules. In A. Skowron & L. Polkowski (Eds.). *Rough sets in knowledge discovery* , Physica Verlag ,Heidelberg ,1998,500-529.

[113] Swiniarski, R. and Nguyen, J. Rough set expert system for texture classification based on 2D spectral features. In *Proceedings of the Third Biennial European Joint Conference on Engineering Systems Design and Analysis ESDA_96* , (Montpellier, France,1996), ESDA ,3-8.

[114] Traina Jr,C., Traina,A., Wu,L. and Faloutsos,C. Fast Feature selection using the fractal dimension. in *Proceedings of the 15th Brazilian Symposium on Databases(SBBD),* (*Maringa, Brazil,2000),* Brazilian Computer Society Press,177-188.

[115] Tsumoto, S. Extraction of Experts′ Decision Rules from Clinical Databases using Rough Set Model. *Journal of Intelligent Data Analysis*, 2,3(1998),215-227.

[116] Tsumoto, S. Automated extraction of hierarchical decision rules from clinical databases using rough set model. *Expert Systems with Applications*, 24,2(2003),189–197.

[117] Upadhyaya, S., Arora A. and Jain ,R. Rough Set Theory: Approach for Similarity Measure in Cluster Analysis. in *Proceedings of 2006 International Conference on Data Mining*, (Hong Kong, China,2006), IEEE Computer Society , 353-356.

[118] Voges, K.E, Pope, N.K and Brown, M.R. Cluster analysis of marketing data examining on-line shopping orientation: a comparison of k-means and rough clustering approaches', In: Abbass HA, Sarker RA, Newton CS,(Eds.) *Heuristics and Optimization for Knowledge Discovery,* Idea Group publishing, Hershey, PA,2002,208-225.

[119] Voges, K.E, Pope, N.K and Brown, M.R. A rough cluster analysis of shopping orientation data. in *Proceedings of Australian and New Zealand Marketing Academy Conference*, (Melbourne,2003), Adelaide,1625-1631.

[120] Wang, Q. H. and Li, J. R. A rough set-based fault ranking prototype system for fault diagnosis. *Engineering Applications of Artificial Intelligence*, 17,8(2004),909–917.

[121] Wang, X., Yang, J., Jensen, R. and Liu, X. Rough set feature selection and rule induction for prediction of malignancy degree in brain glioma. *Computer Methods and Programs in Biomedicine*, 83(2006), 147–156.

[122] Wroblewski, J .Finding minimal reducts using genetic algorithms. In Wang, P.P. (Ed), Proceedings of the International Workshop on Rough Sets Soft Computing at Second Annual Joint Conference on Information Sciences, (North Carolina, USA,1995), Rough sets & Computer Science,186-189.

[123] Wroblewski, J.Theoretical foundations of order-based genetic algorithms, *Fundamenta Informaticae*, IOS Press,28(1996) , 423-430.

[124] Wroblewski, J. Genetic algorithms in decomposition and classification problems. In Polkowski, L. & Skowron, A. (Eds.), *Rough sets in Knowledge Discovery 1: Applications, Case Studies and Software System,* Physica-Verlag ,Heidelberg ,1998,472-492.

[125] Yao, Y.Y. On combining rough and fuzzy sets. in Lin, T.Y. (Ed.). *Proceedings of the Computer Science Conference CSC'95 Workshop on Rough Sets and Database Mining*, (Nashville, Tennessee, USA, 1995), San Jose State University,165-172.

[126] Yao, Y. Y. , Li, X. , Lin, T. Y. and Liu, Q. Representation and Classification of Rough Set Models. *Soft Computing: Rough Sets, Fuzzy Logic, Neural Network, Uncertainty Management and knowledge Discovery*, Society for Computer Simulation ,CA, USA ,1995,44-47.

[127] Yao, Y.Y. Two views of the theory of rough sets in finite universes. *Approximate Reasoning*, 15,4(1996), 191–317.

[128] Yao, Y.Y, Lingras, P., Wang, R and Miao, D. Interval set cluster analysis: a re-formulation. in Proceedings of the *International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular computing (RSFDGrC'09)*, Lecture Notes in Computer Science, LNCS-5908, (Berlin, Heidelberg,2009),Springer-Verlag,398-405.

[129] Yao, Y.Y. Three-way decisions with probabilistic rough sets. .*Information Sciences*, 180,3(2010),341-353.

[130] Ye, N. and Li, X. A scalable, incremental learning algorithm for classification problems.Computers and Industrial Engineering, 43(2002), 677–692.

[131] Zadeh, L.A. A new direction in AI: Toward a computational theory of perceptions. *AI Magazine*, 22,1(2001),73-84.

[132] Zaki, M.J. Scalable algorithms for association mining. *IEEE Transactions on Knowledge and Data Engineering*, 12,3(2000),372–390.

[133] Zhang, M. and Yao, J.T. A rough set approach to feature selection. *Proceedings of the 23rd International Conference* of the *North American Fuzzy Information Processing Society* (*NAFIPS* 2004), (Banff, Canada,2004), North America Fuzzy Information Processing Society,434-439.

[134] *Zhong, N*., Dong, J.Z. and Ohsuga, S. Using Rough sets with heuristics for feature selection. *Journal of Intelligent Information Systems*, 16(2001), 199-214.

[135] Zhu, X. *Semi-Supervised Learning with Graphs*. PhD thesis, Carnegie Mellon University,USA,2005.

[136] Ziarko, W.P. (Ed.) .Rough Sets, Fuzzy Sets and Knowledge Discovery. *i*n Proceedings of International Workshop on Rough Set and Knowledge Discovery,. Canada, Springer-Verlag, 1994.