International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-3 Issue-5, November 2014

Rough Sets for Feature Selection and Classification: An Overview with Applications

Amit Saxena, Leeladhar Kumar Gavel, Madan Madhaw Shrivas

***Abstract*: Rough set theory provides a useful mathematical concept to draw useful decisions from real life data involving vagueness, uncertainty and impreciseness and is therefore applied successfully in the field of pattern recognition, machine learning and knowledge discovery. This paper presents an overview of basic concepts of rough set theory. The paper also surveys applications of rough sets in feature selection and classification.**

***Keywords*: Pattern recognition, feature selection, classification.**

I. INTRODUCTION

Rough set theory proposed by Pawlak [1], [2], has become a well-established theory to resolve problems related to vagueness, uncertainty and incomplete information in variety of applications related to pattern recognition and machine learning. The problems belonging to these areas widely include classification [3], [4], [5], feature selection [6], [7], [8], [9], [10], [11], clustering [12], [13], [14], data mining, knowledge discovery [15], Image processing[16], and prediction[17]. The theory of rough sets can be described in two ways: constructively and algebraically (axiomatically) [18]. The constructive approach is found suitable for practical applications of rough sets, while the algebraic approach is appropriate for studying the structures (theory) of rough set algebras. Subsequently a new extension of rough set theory, called α –RST [19], presented a suitable framework to deal with vague data and for quantifying fuzzy concepts. Two new operators introduced for the rough set theory [20] can be used to convert two inequalities into equalities. Hence, many properties in rough set theory can be improved and in particular, the union, the intersection, and the complement operations can be redefined based on these two equalities. A new roughness measure of a fuzzy set based on the notion of the mass assignment of a fuzzy set and its α-cuts are proposed by Huynh et al. [21]. It is shown that this roughness measure inherits interesting properties of Pawlak's roughness measures of a crisp set. The Variable Precision Rough Set (VPRS) model extends the basic rough set theory to incorporate probabilistic information [22]. A non-parametric modification of the VPRS model called the Bayesian Rough Set (BRS) model tends to serve well for data mining applications whereas the predictive model is suitable for primary importance. Knowledge acquisition using rough set theory in the systems having incomplete information is proposed in literature [15]. Two kinds of partitions, lower and upper approximations, are formed for the mining of certain and association rules in incomplete decision tables. As a result one type of *optimal certain* and two types of *optimal association* decision rules is generated. Definable concepts are very important in investigating properties of various generalized rough set models [23]. The rough set concept has led to its various generalizations approach to multi-criteria decision making for synthesis and analysis of concept approximations in the distributed environment of intelligent agents [24].

Based on rough membership and rough inclusion functions [25], Bayesian decision-theoretic analysis is adopted to provide a systematic method for determining the precision parameters by using more familiar notions of costs and risks. Jing Tao Yao [26] presented a list of decision types based on rough set regions created by two models viz. Pawlak and probabilistic. A general framework is formed for the study of fuzzy rough sets which uses both approaches (constructive and axiomatic) and classical representation of Interval Type 2 (IT2) fuzzy [27] and rough approximation operators. The association between special IT2 fuzzy relations and IT2 fuzzy rough approximation operators is investigated [28]. The composite rough set model for composite relations was developed to deal with attributes of multiple different types simultaneously [29]. Multigranulation rough set (MGRS) theory provides a new perspective for decision making analysis based on the rough set theory. The new model based on MGRS and decision-theoretic rough sets together is called a multigranulation decision theoretic rough set model [30]. Jia et al. [31] proposed an optimization representation of decision-theoretic rough set model to minimizing the decision cost. The MGRS model based on the decision strategy *Seeking common ground while eliminating differences* (SCED), also called pessimistic rough set model was proposed in literature [32] specifying the relationship between optimistic and pessimistic multigranulation rough sets. Susmaga [33] introduced the constructs in a uniform definition framework of Dominance-based Rough Sets Approach (DRSA) which is a collection of twenty four reduced attribute subsets. The DRSA systematically discusses the basic theory of the probabilistic rough fuzzy set.

Revised Manuscript Received on 30 November 2014.
* Correspondence Author
Prof. Amit Saxena*, Department of CSIT, Guru Ghasidas Vishwavidyalaya, Bilaspur, India.
Leeladhar Kumar Gavel, Department of CSIT, Guru Ghasidas Vishwavidyalaya, Bilaspur, India.
Madan Madhaw Shrivas, Department of CSIT, Guru Ghasidas Vishwavidyalaya, Bilaspur, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

Subsequently the 0.5-probabilistic rough fuzzy set model, variable precision probabilistic rough fuzzy set model and Bayesian rough fuzzy set model are defined [34].

II. ROUGH SET THEORY: BASIC DEFINITIONS

Rough set theory was developed by Zdzislaw Pawlak [1], [2]. It deals mainly with classification analysis of data tables. The main goal of the rough set analysis is to synthesize approximation of concepts from the acquired data which contains vagueness, missing values or redundancy of features. In this section, some terms which are frequently used in rough sets are defined.

A. Information and decision systems

A data set is represented as a table where each row represents a case, an event, a pattern or simply an object. Every column represents an attribute (a variable, an observation, a property, a feature) that can be measured for each object; the attribute may also be supplied by a human expert or user. This table is called an information system. More formally, it is a pair $I = (U, A)$ where U is a non-empty finite set of objects called Universe and A is a non-empty finite set of attributes such that $a: U \rightarrow V_a$ for every $a \in A$. The set V_a is called the value set of a.

In many applications, the class of the attribute of several patterns (or objects) is known in advance. This set of patterns is called training data. The class of an unknown pattern (also called test data), can be predicted from the priory knowledge of the training data; this process is known as supervised learning. Information systems of this type are called decision systems. Mathematically a decision system is any information system of the form $D = (U, A \cup \{d\})$, where $d \notin A$ is the decision attribute. The element of A are called condition attributes or simply conditions.

Table. I. An Example Dataset

$x \in U$	a	b	c	d $\Rightarrow$ (class)	e
0	S	R	T	T	R
1	R	S	S	S	T
2	T	R	R	S	S
3	S	S	R	T	T
4	S	R	T	R	S
5	T	T	R	S	S
6	T	S	S	S	T
7	R	S	S	R	S

An example of a decision system can be found in Table I. The table consists of four conditional features *(a, b, c, d)*, a decision feature *(e)* also called class, and eight objects (or patterns). A decision system is consistent if for every set of objects whose attribute values are the same, the corresponding decision attributes are also identical.

Indiscernibility

A decision system (i.e. decision table) represents the knowledge about the model. This table may be redundant in at least two ways. The same or indiscernible objects may be represented several times or even some of the attributes may be superfluous.

As we know, for a binary relation $R \subseteq X \times X$ to be an equivalence relation, it should be reflexive (i.e. an object is in relation with itself xRx), symmetric (if xRy then yRx) and transitive (if xRy and yRz then xRz) is called an equivalence relation. The equivalence class of an element $x \in X$ consists of all objects $y \in X$ such that xRy.

Let $I=(U, A)$ be an information system, then with any $B \subseteq A$, there is associated an equivalence relation $IND_I(B)$.

$$IND_I(B) = \{(x, x') \in U^2 \mid \forall a \in B a(x) = a(x')\} \quad (1)$$

$IND_I(B)$ is called the B-indiscernibility relation.

If $(x, x') \in IND_I(B)$, then object x and x' are indiscernible from each other by attributes from B. The equivalence classes of the B-indiscernibility relation are denoted $[x]_B$.

For the illustrative example, if $B=\{b, c\}$ then object 1, 6, 7 (values S S) and objects 0, 4 values (R T) are indiscernible; $IND_I(B)$ creates the following partition of U. $U/IND_I(B)=\{\{0, 4\}, \{1, 6, 7\}, \{2\},\{3\},\{5\}\}$

B. Lower and upper approximation

Let $I = (U,A)$ be an information system and let $B \subseteq A$ and $X \subseteq U$. We can approximate X using only the information contained in B by constructing the B-lower and B-upper approximations of X, denoted $\underline{B}(X)$ and $\overline{B}(X)$ respectively.

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

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

C. Positive, negative and boundary regions

Let P and Q be sets of attributes including equivalence relations over U, then the positive, negative, and boundary region are defined as

$$POS_P(Q) = \cup_{X \in U/Q} \underline{P}X \quad (4)$$

$$NEG_P(Q) = U - \cup_{X \in U/Q} \overline{P}X \quad (5)$$

$$BND_P(Q) = \cup_{X \in U/Q} \overline{P}X - \cup_{X \in U/Q} \underline{P}X \quad (6)$$

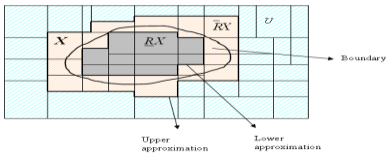

Fig. 1 A Rough set

The positive region comprises all objects of U that can be classified to classes of U/Q using the information contained within attributes P. The boundary region, $BND_P(Q)$, is the set of objects that can possibly, but not certainly, be classified in this way. The negative region, $NEG_P(Q)$, is the set of objects that cannot be classified to classes of U/Q.

For example, let $P = \{b, c\}$ and $Q=\{e\}$, then

$$POS_P(Q) = \cup\{\emptyset, \{2, 5\}, \{3\}\} = \{2, 3, 5\}$$

$$NEG_P(Q) = U - \cup\{\{0, 4\}, \{2, 0, 4, 1, 6, 7\ 5\}, \{3, 1, 6, 7\}\} = \emptyset$$

$$BND_P(Q) = \cup\{\{0, 4\}, \{2, 0, 4, 1, 6, 7\ 5\}, \{3, 1, 6, 7\}\} - \{2, 3, 5\} = \{0, 1, 4, 6, 7\}.$$

This means that objects 2, 3 and 5 can certainly be classified as belonging to a class attribute e, where considering attributes b and c. The rest of the objects cannot be classified as information that would make them discernible is absent.

D. Dependency of attributes

Another important issue in data analysis is discovering dependencies between attributes. Intuitively, a set of attributes Q depends totally on set of attributes P, denoted by $\Rightarrow Q$, if all values of attribute from Q are uniquely determined by values of attributes from P.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved

International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-3 Issue-5, November 2014

Formally, dependency can be defined in the following way. Let P and Q be subsets of A.

We will say that Q depends on P in a degree k ($0 \leq k \leq 1$), denoted$P \Rightarrow_k Q$, if

$$k = \gamma(P, Q) = \frac{|POS_P(Q)|}{|U|} \qquad (7)$$

Where

$$POS_P(Q) = \bigcup_{X \in U/Q} \underline{P}X$$

Called positive region of the partition *U/Q* with respect to *P*, is the set of all elements of *U* that can be uniquely classified to block of the partition *U/Q*, by means of *P*.

Obviously

$$\gamma(P, Q) = \sum_{x \in U/Q} \frac{|\underline{P}X|}{|U|} \qquad (8)$$

If *k=1* we say that *Q* depends totally on *P* and if *k <1*, we say that *Q* depends partially on *P*. Again

For example, if *P={a, b, c}* and *Q={e}* then

$$\gamma_{\{a,b,c\}}(\{e\}) = \frac{|\{2,3,5,6\}|}{8} = 4/8$$

$$\gamma_{\{a,b\}}(\{e\}) = \frac{|\{2,3,5,6\}|}{8} = 4/8$$

$$\gamma_{\{b,c\}}(\{e\}) = \frac{|\{2,3,5\}|}{8} = 3/8$$

$$\gamma_{\{a,c\}}(\{e\}) = \frac{|\{2,3,5,6\}|}{8} = 4/8$$

E. Reducts and Core

In several application problems, the information system is unnecessarily large due to existence of repeated objects or redundant features. One way to reduce the dimensionality is to search for a minimal representation of the original dataset. For this reason, concept of a reduct is introduced and defined as minimal subset R of the initial attribute set *C* such that for a given set of attributes *D*, $\gamma_R(D) = \gamma_C(D)$. *R* is a minimal subset if $\gamma_{R-\{a\}}(D) \neq \gamma_R(D)$ for all $a \in R$. This means that any attribute removed from the subset will affect the dependency degree. Hence a minimal subset by this definition may not be the *global* minimum (a reduct of smallest cardinality). A given dataset may have many reduct sets, and the collection of all reducts is denoted by

$$R_{all} = \{X | X \subseteq C, \gamma_X(D) = \gamma_C(D); \gamma_{X-\{a\}}(D) \neq \gamma_X(D), \forall a \in X\} \qquad (9)$$

The intersection of all the sets in R_{all} is called the core, denoted by CORE(C).

$$CORE(C) = \cap RED(C) \qquad (10)$$

Where RED(C) is the set of all reducts of C.

F. Discernibility matrix

Many applications of rough sets make use of discernibility matrices for finding rules or reducts. A discernibility matrix of a decision table $(U, C \cap D)$ is a symmetric $|U| \times |U|$ matrix with entries defined by

$$c_{ij} = \{a \in C | a(x_i) \neq a(x_j)\}, \quad i, j = 1, \ldots, |U| \qquad (11)$$

Each c_{ij} contains those attributes that differ between objects *i* and *j*.

For finding reduct, the decision-relative discernibility matrix is of more interest. This matrix considers only those object discernibilities that occur when the corresponding decision attributes differ [35]. The decision-relative discernibility matrix is produced as shown in Table II. For example, it can be seen from the table that objects 0 and 1 differ in each attribute. Although some attributes in objects 1 and 3 differ, their corresponding decisions are the same, so no entry appears in the decision-relative matrix. Grouping all entries containing single attributes forms the core of the dataset (those attributes appearing in every reduct). Here, the core of the dataset is {d}.

From this matrix, the concept of discernibility functions can be introduced. This is a concise notation of how each object within dataset may be distinguished from the others. A discernibility function f_D is a Boolean function of m Boolean variables $a_1^*, \ldots, a_m^*$ (corresponding to the membership of attributes $a_1 \ldots a_m$ to a given entry of the discernibility matrix), defined as follows:

Table. II
DECISION-RELATIVE DISCERNIBILITY MATRIX

x ∈ U	0	1	2	3	4	5	6	7
0								
1	a, b, c, d							
2	a, c, d	a, b, c						
3	b, c		a, b, d					
4	d	a, b, c, d		b, c, d				
5	a, b, c, d	a, b, c		a, b, d				
6	a, b, c, d		b, c		a, b, c, d	b, c		
7	a, b, c, d	d		a, c, d			a, d	

$$f_D(a_1^*, \ldots, a_m^*) = \wedge \{\vee c_{ij}^* | 1 \leq j \leq i \leq |U|, c_{ij} \neq \phi\} \quad (12)$$

Where $c_{ij}^* = \{a^* | a \in c_{il}\}$. The notation $\vee \{a, b, c, d\}$ and $\wedge \{a, b, c, d\}$ denote $a \vee b \vee c \vee d$ and $a \wedge b \wedge c \wedge d$, respectively. By finding the set of all prime implecants of the discernibility function, all the minimal reducts of a system may be determined. From Table II, the decision-relative discernibility function is (with duplicates removed)

$$f_D(a^*, b^*, c^*, d^*) = (a^* \vee b^* \vee c^* \vee d^*) \wedge (a^* \vee c^* \vee d^*) \wedge (b^* \vee c^*) \wedge (d^*) \wedge (a^* \vee b^* \vee c^*) \wedge (a^* \vee b^* \vee d^*) \wedge (b^* \vee c^* \vee d^*) \wedge (a^* \vee d^*)$$

Further simplification can be performed by removing those clauses that are subsumed by others:

$$f_D(a^*, b^*, c^*, d^*) = (b^* \vee c^*) \wedge (d^*)$$

The reducts of the dataset may be obtained by converting the expression above from conjunctive normal form to disjunctive normal form (without negation). Hence the minimal reducts are *{b, d}* and *{c, d}*.

After a brief introduction of rough sets, we are now ready to explore some of the research issues based on rough set theory. There have been several areas where intensive research is being carried out including following [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17].

Some of the research directions on Rough Sets are as follows-

- Classification
- Feature selection
- Dimensionality reduction
- Rough set based clustering
- Rough sets and noisy data
- Rough sets and relational databases
- Rough sets and incomplete information systems. In particular, missing value problems

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved

- Boolean reasoning and approximate Boolean reasoning strategies as the basis for efficient heuristics for rough set methods.
- Rough sets and inductive reasoning
- Rough set based approach based on neighbourhood (uncertainty) functions and inclusion relation. In particular, variable precision rough set model.

In this paper we present the state of art in the application of rough sets in feature selection and classification. Recently, researchers have focused their attention on reduct and classification algorithms based on rough sets [8], [36], [37], [38].

III. FEATURE SELECTION AND CLASSIFICATION USING ROUGH SET

Feature selection process refers to selecting the subsets of attributes (features) from the set of all attributes. The classification [39] is the process of separating the objects on the basis of some criteria. On many occasions, the class of each object is given in advance then it becomes easy to group the objects in to their classes. This type of classification is called supervised classification. On the other hand, many times there is no class attached to any object and we have to group them on the basis of some similarity based criteria like color, size or similar attributes. Such type of classification is called unsupervised. Clustering is an unsupervised classification. The purpose of the feature selection is to identify the significant features, eliminate the irrelevant or dispensable features. This will reduce the burden on learning models and as a result it will help in building better learning model. The benefits of feature selection are two folds: it considerably decreases the computation time of the induction algorithm and secondly increases the accuracy of the resulting mode. Feature selection has been studied intensively in the past one decade [3], [6], [16]. Nowadays, numerous successful implementations of feature selection to various applications are summarized in Table III and are discussed in this section.

Khoo et al. [3] proposed a novel approach for the classification and rule induction of inconsistent information systems. It was achieved by integrating rough set theory with a statistics-based inductive learning algorithm. The framework of a prototype rough set-based classification system (R-class) was also presented. This R-class technique was compared with the other rule techniques like ID3 and LERS. For each possible rule generated, R-class was able to provide an estimation of the expected classification reliability. This assisted users in deciding the appropriateness of the rules generated.

Swiniarski and Skowron [6] presented an application of rough set method for feature selection in pattern recognition. They proposed a new feature selection method to the result of principle component analysis (PCA [40], [41]) used for feature projection and reduction. Finally rough set methods had shown ability to reduce significantly the pattern dimensionality and had proven to be viable data mining techniques as a front end of neural network classifiers[42], [43].

Fen et al. [44] proposed new incremental rule-extraction algorithms to solve the dynamic database problem. When a new object is added-in the information system, it is unnecessary to re-compute rule sets from the very beginning. The proposed approach updates rule sets by partly modifying original rule sets. Therefore, the computation time is saved. This is especially useful while extracting rules in a large database.

Table. III
FEATURE SELECTION FOR DECISION SYSTEM BASED ON THE ROUGH SET THEORY APPROACH

S.No.	Authors	Proposal	Description
1.	Khoo et al. [3]	Classification and Rule Induction	Developed novel approach (R-Class) for classification and rule induction of inconsistent information.
2.	Swiniarski and Skowron [6]	Feature selection	Presented an application of rough set method for feature selection in pattern recognition.
3.	Meng and Shi [8]	Feature selection	Established reduction concepts specifically for IIDSs, mainly by extending related reduction concepts from other types of decision systems into IIDSs, and then derived their relationships and properties.
4.	Iquebal et al. [11]	Feature selection	Proposed a new variant of MTS feature selection method which explores anew measure of goodness-of-model in terms of conditional probability of system states on subset of variables.
5.	Parmarmar et al. [12]	Clustering	Proposed a new algorithm for clustering categorical data, termed Min–Min-Roughness (MMR), based on Rough Set Theory (RST), which has the ability to handle the uncertainty in the clustering process.
6.	Yu et al. [13]	Clustering	Proposed an efficient automatic method by extending the decision-theoretic rough set model to clustering.
7.	Park and Choi [14]	Clustering	Proposed information-theoretic dependency roughness (ITDR), an alternative technique for categorical data clustering.
8.	Xiang-wei and Yian-fang [36]	Classification	Proposed a novel effective pre-processing algorithm based on rough sets.
9.	Susmaga [37]	Feature selection	Proposed a model to generate reducts and constructs from rough set based on inter-class and intra-class information.
10.	Shu [38]	Feature selection	Proposed an incremental approach based on rough set for feature selection, which can accelerate the feature selection process in dynamic incomplete data.
11.	Fen et al. [44]	Rule Induction	Proposed an incremental rule-extraction algorithm based on the previous rule-extraction algorithm to resolve dynamic data set.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved

12.	Rady et al. [45]	Generalization of RST	Introduced a new method concerning the generalization and modification of the rough set theory
13.	Asharaf et al. [46]	Clustering	Proposed a novel incremental approach to clustering interval data.
14.	Peters [47]	Clustering	Proposed to modify a rough cluster algorithm and suggested some alternative solutions led to the introduction of a refined rough *k*-means.
15.	Li et al.[48]	Classification	Proposed customer classification prediction model to reduce the complexity of decision-makers.
16.	Trabelsi et al. [49]	Classification	Presented two new classification approaches based on rough sets called BRSC and BRSC-GDT under the belief function framework that are able to learn decision rules from uncertain data.
17.	Salamo and Lopez-Sanchez [50]	Feature selection	Investigated feature selection based on rough sets for dimensionality reduction in Case-Based Reasoning classifiers.
18.	Chakhar and Saad [51]	Classification	Proposed a methodology to support groups in multi-criteria classification problems.
19.	Ye et al. [52]	Classification	Proposed an approach based on rough set for measuring the data quality and guiding the process of anonymization operations.
20.	Hu [53]	Classification	Proposed a novel classification method by incorporating a preference index based on pairwise comparisons into a rough set approach.
21.	Lu et al. [55]	Feature selection	Propose a boundary region-based feature selection algorithm (BRFS), which has the ability to efficiently find a feature subset from a large incomplete decision system.
22.	Liu et al. [56]	Feature selection	Proposed a new approach to calculate the reduct SCHE in VRPS model which was focuses on calculating a β-distribution reduct while avoiding the anomaly problem in the VPRS model.
23.	Zheng et al.[57]	Feature Selection	Enhancement for heuristic attribute reduction (EHAR) in rough set is proposed and superior performance achieved.
24.	Boggia et al. [58]	Rural Sustainable Development potentialities	Developed a decision support system based on Dominance-based Rough Set Approach (DRSA), to assess the level of Rural Sustainable Development in specific areas.
25.	Min et al. [59]	Feature selection	Proposed a new feature selection problem concerning the test cost constraint.
26.	Liu et al. [60]	Classification	Proposed two integrated classification approaches, binary logistic regression and multinomial logistic regression, to combined logistic regression and DTRS together.
27.	Zhou [61]	Classification	Proposed a new model using decision-theoretic rough for an information table with more than two decision classes.
28.	Huang et al. [62]	Feature selection	Proposed a type of matroid called a nullity-based matroid in the context of rough sets and two types of matrices to characterize the metroid
29.	Liu et al. [63]	Classification	The rough set based incremental approaches were proposed to deal with the missing and incomplete information in real decision problems.
30.	Riza et al. [64]	R-Package (Feature selection , classification)	Proposed a Roughsets in an R package for implementing algorithms from rough set and fuzzy rough set theories.
31.	Kadzinski et al. [65]	Feature selection , classification	Proposed a new approach to multiple criteria sorting problems deriving from Dominance-based Rough Set Approach.
32.	Li et al. [93]	Clustering	Proposed an extended Rough c-means clustering algorithm based on the concept of decision-theoretic Rough Sets model.

Li et al. [48] proposed customer classification prediction model based on rough set. It classified the customers based on a few properties, and the analysis reduced the complexity of decision-makers. This model helped companies to predict in advance the new customer or potential costumer value level and make the more targeted client development strategy.

Meng and Shi [8] proposed a model of attribute reduction in inconsistent incomplete decision systems (IIDSs). The idea is that a missing attribute value may be replaced with any known value of a corresponding attribute (such as missing attribute value is called a “do not care” condition). Finally, they proposed approaches which were effective and suitable for handling both numerical and categorical attributes, but they had different application conditions and could provide a solution to the reduction problem for IIDSs.

Chakhar and Saad [51] proposed a two-phase methodology to support groups in multi criteria classification problems. The first phase, which relies on a dominance-based rough set approach (DRSA), takes a set of assignment examples as input and outputs a set of collective decision rules, representing a generalized description of the decision makers' preference information. The second phase then applies these collective decision rules to classify all decision objects. The methodology uses *if … then …* aggregation rules and coherently implements the majority principle and veto effect. The aggregation rules thus allow obtaining consensual decisions.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)
© Copyright: All rights reserved

Hu [53] presented a novel rough-set-based classification method (i.e., RSRC-P) by incorporating the pair wise-comparison-based tables into the RSRC provided by the well-known RSESlib [54]. RSRC-P is a variant of RSRC-O, which uses the pair wise comparisons using the preference relation as in the Preference Ranking Organization Methods for Enrichment Evaluations (PROMETHEE) methods, to gauge the intensity of preference for one pattern over another pattern on each criterion before classification. The Rough Set Based Rule Classifier (RSRC) provided by the well-known library for the Rough Set Exploration System (RSES) running under windows has been successfully used to generate decision rules by using the pairwise comparisons-based tables. Specially, parameters related to the preference function on each criterion have been determined using a genetic-algorithm-based approach.

Zheng et al. [57] proposed an enhancement for heuristic attribute reduction (EHAR) in rough set. With the application of EHAR, two representative heuristic attribute reduction algorithms (dependence based and consistency based algorithms) are improved. EHAR can significantly help both of the heuristic attribute reduction algorithms to achieve the optimal reduct under circumstances where two or more attributes have the same largest significance in some rounds.

Zhou [61] introduced a probabilistic rough set approximation for an information table with more than two decision classes. In order to emphasize the semantic interpretation of probabilistic rough sets with three-way decision; three pair-wise disjoint positive, boundary, and negative regions were used instead of pair of lower and upper approximations. This approach was considered as a straightforward generalization of the three-way classification in decision-theoretic rough set models and tackled the limitations of the previous related work and provided a cost-sensitive solution to multi-class decision making.

Huang et al. [62] proposed a type of matroid of rough sets based on the concept of nullity called a nullity-based matroid and defined the relation between nullities and matrices. Two types of matrices was represented: first is types of matriod and second is nullity of matroid. Matroids were applied to attribute reduction problems in information system and solved attribute reduction issues in information system using matrices.

Susmaga [37] worked on reducts and constructs, which are reduced subsets of attributes that represent filter-based approach of feature selection. The reducts were referred to as the inter-class reducts, originated from Classic Rough Set Approach (CRSA) and as such also adapted to the needs of Dominance-based Rough Set Approach (DRSA). Finally, the concept of the construct, which incorporated both inter-class as well as intra-class information, was introduced.

Kadzinski et al. [65] presented a new multiple criteria sorting method deriving from Dominance-based Rough Set Approach (DRSA). The preference information supplied by the Decision Maker (DM) is a set of possibly imprecise and inconsistent assignment examples on a subset of reference alternatives relatively well known to the DM. DRSA was used for structuring the data, and subsequently, represented the assignment examples by all minimal sets of rules compatible with the lower approximations of class unions. Such a minimal set of rules is one of the instances of the preference model compatible with DM's preference information and implemented the principle of Robust Ordinal Regression (ROR) to decision rule preference model.

From the above discussions, it is apparent that feature selection which is an important part in data processing especially classification is well recognized by rough sets. Feature selection and classification are closely associated to each other. The classifiers can be modelled for supervised or unsupervised databases. Several other approaches for building classifiers have been reported including ensemble[66], unsupervised using Sammon's function[67], Polynomial neural network PNN[68] based classifiers [69] etc. Rough sets have also proved to be significant in feature selection and classification.

IV. ROUGH SETS BASED HYBRIDIZATION AND APPLICATIONS

Like other theories such as neural networks, fuzzy sets, evolutionary techniques; rough sets are also used in hybridization with the established techniques. The reason for using hybridization is that, on many occasions, a single technique is not able to overcome some of its limitations, under such situation it is better to combine the technique with some other which is a proven technique to do well for that limitation. Some of the hybridizations of rough sets with other techniques are as follows:

- Rough sets with fuzzy sets: rough sets are closely associated to fuzzy sets [70] due to its capability to address vagueness in information. The hybridization of rough sets with fuzzy sets is used for improving the results in applications like data mining, feature reduction etc. [71], [72], [73], [74], [75], [76].

- Rough sets with Neural Networks: Neural networks [77], [78] are the mathematical simulation of human brain ad has extensive applications in various pattern recognition and machine learning processes. Rough sets have been used with neural networks on various applications for improving the results [42], [43].

- Rough sets with metaheuristic algorithm: A metaheuristic is a set of algorithmic concepts that can be used to define heuristic methods applicable to a wide set of different problems. Rough sets are used with genetic algorithm [79], ACO [80], PSO [81] and other optimization techniques for various applications [82], [83], [84].

Applications: Rough set theory has been successfully applied in almost all the fields. The major drawback of traditional rough set models in real life application is the inefficiency to compute reducts and generate cores attributes. To improve the efficiency of computing core attributes and reducts, many novel approaches have been developed [22], [25], [28], [29], [30], [32], [34], [85].

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved

Some more applications of rough sets in areas like medical images [16], breast cancer [86], texture classification [87] and [88], [89], [90], [91], [92] can be seen.

V. CONCLUSION

In several real life databases, the information collected to represent various decisions along with attributes contains vagueness. For few identical attributes, decisions made or the classes labelled are different for different patterns. Rough set theory has emerged as a powerful tool to handle such vagueness. This paper presents an overview of the rough set theory, terms used in the rough sets with examples. Rough sets can be applied to the important process of feature selection and classification. In this paper, the applications of rough sets to feature selection and classification are extensively discussed. The investigations and developments made in these areas are tabled and discussed in the paper. Rough sets can be combined with other techniques when they alone are not able to produce better results. Some of the hybridizations of rough sets with bench mark techniques like neural networks, fuzzy sets and evolutionary techniques are presented with the state of art therein. Further, applications of rough sets are numerous, some of the applications are summarised in the paper with references. The available literature in rough sets opens a promising gate towards future research directions in many other complex areas including big data, communications, computational intelligence etc.

REFERENCES

1. Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, System Theory, Kluwer, Dordrecht, 1991.
2. ZdzishawPawlak, AndrzejSkowron, "Rudiments of rough sets," Information Sciences, vol. 177, 2007, pp. 3-27.
3. L. P. Khoo, S. B. Tor and L. Y. Zhi, "A Rough-Set-Based Approach for Classification and Rule Induction," Int J AdvManufTechnol , vol.15,1999, pp. 438-444.
4. Daijin Kim, "Data classification based on tolerant rough set," Pattern Recognition , vol. 34, 2001,pp. 1613-1624.
5. Richard Jensen, Chris Cornelis, "Fuzzy-rough nearest neighbour classification and prediction," Theoretical Computer Science, vol. 412, 2011, pp. 5871-5884.
6. Roman W. Swiniarski, AndrzejSkowron, "Rough set methods in feature selection and recognition," Pattern Recognition Letter, vol. 24, 2003, pp. 833-849.
7. Rajen B. Bhatt, M. Gopal, "On fuzzy-rough sets approach to feature selection," Pattern Recognition Letters, vol. 26, 2005, pp. 965-975.
8. ZuqiangMeng, Zhongzhi Shi, "Extended rough set-based attribute reduction in inconsistent incomplete decision systems," Information Sciences, vol. 204, 2012, pp. 44-69.
9. Pramod Kumar P, Prahlad Vadakkepat, Loh Ai Poh, "Fuzzy- rough discriminative feature selection and classification algorithm, with application to microarray and image datasets," Applied Soft Computing, vol. 11, 2011, pp. 3429-3440.
10. Neil Mac Parthalain, Richard Jensen, "Unsupervised fuzzy-rough set-based dimensionality reduction," Information Sciences, vol. 229, 2013, pp. 106-121.
11. AsifSikanderIqueba1, Avishek Pal, DarekCeglarek, Manoj Kumar Tiwari, "Enhancement of Mahalanobis – Taguchi System via rough sets based feature selection," Expert Systems with Applications, In Press, 2014.
12. DarshitParmar, Teresa Wu, Jennifer Blackhurst, "MMR: An algorithm for clustering categorical data using Rough Set Theory," Data & Knowledge Engineering, vol. 63, 2007, pp. 879-893.
13. Hong Yu, Zhanguo Liu, Guoyin Wang, "An automatic method to determine the number of clusters using decision-theoretic rough set," International Journal of Approximate Reasoning , vol. 55, 2014, pp. 101-115.
14. In-Kyoo Park, Gyoo-Seok Choi, "Rough set approach for clustering categorical data using information-theoretic dependency measure, "Information Systems, In Press, 2014.
15. Yee Leung, Wei-Zhi Wu, Wwn-Xiu Zhang, "Knowledge acquisition in incomplete information systems: A rough set approach," European journal of Operation Research, vol. 168, 2006, pp. 164-180.
16. Shoji Hirano, ShusakuTsumoto, Rough representation of a region of interest in medical images , International Journal of Approximate Reasoning, vol. 40, 2005, pp. 23-34.
17. B. S. Ahn, S.S. Cho, C.Y. Kim, "The integrated methodology of rough set theory and artificial neural network for business failure prediction," Expert System with Applications, vol. 18, 2000, pp. 65-74.
18. Y. Y. Yao , "Constructive and algebraic methods of the theory of rough sets," Journal of Information Sciences, vol. 109, 1998, pp. 21-4.
19. Mohamed Quafafou, "α-RST: a generalization of rough set," Information Sciences, vol.124, 2000, pp. 301-316.
20. Huaguang Zhang, Hongli , Derong Liu, "Two new operators in rough set theory with applications to fuzzy sets," Information Sciences, vol. 166, 2004, pp. 147-165.
21. Van-Nam Huynh, YoshiteruNakamori, "A roughness measure for fuzzy sets," Information Sciences, vol. 173, 2005, pp. 255-275.
22. DominikSlezak, WajciechZiarko, "The investigation of the Bayesian rough set model," International Journal of Approximate Reasoning, vol. 40, 2005, pp. 81-91.
23. Daowu Pei, "On definable concepts of rough set models," Information sciences, vol. 177, 2007, pp. 4230-4239.
24. ZdzishawPawlak, AndrzejSkowron, "Rough sets: Some extensions," Information Sciences, vol. 177, 2007, pp. 28-40.
25. Yiyu Yao, "Probabilistic rough set approximations," International Journal of Approximate Reasoning, vol. 49, 2008, pp. 255-271.
26. Joseph P. Herbert, JingTao Yao, "Criteria for choosing a rough set model," Computer and Mathematics with Applications, vol. 57, , 2009, pp. 908-918.
27. Zhiming Zhang, "On characterization of generalized interval type-2 fuzzy rough sets," Information Sciences, vol. 219, 2013, pp. 124-150.
28. Degang Chen, Qiang He, Xizhao Wang, "FRSVMs : Fuzzy rough set based support vector machines," Fuzzy Sets and Systems, vol. 161, , 2010, pp. 596-607.
29. Junbo Zhang, Tianrui Li, Hongmei Chen, "Composite rough sets for dynamic data mining," Information Sciences, vol. 257, 2014, pp. 81-100.
30. Yuhua Qian, Hu Zhang, Yanli Sang, Jiye Liange, "Multigranulation decision-theoretic rough sets," International Journal of Approximate Reasoning, vol. 55, 2014, pp. 225-237.
31. Xiuyi Jia, Zhenmin Tang, Wenhe Liao, Lin Shang, "On an optimization representation of decision-theoretic rough set model," International Journal of Approximate Reasoning, vol. 55, 2014, pp. 156-166.
32. Yuhua Qian, Shunyong Li, Jiye Liang, Zhongzhi Shi, Feng Wang, "Pessimistic rough set based decisions: A multigranulation fusion strategy," Information Sciences, vol. 264, 2014, pp. 196-210.
33. Robert Susmaga, "Reducts and constructs in classic and dominance-based rough sets approach," Information Sciences, vol. 271, 2014, pp. 45-64.
34. Bingzhen Sun, Weimin Ma, Haiyan Zhao, "Decision-theoretic rough fuzzy set model and application," Information Sciences, In Press, 2014.
35. Rechard Jensen and QiangShen, Computational Intelligence and Feature Selection- Rough and Fuzzy Approaches, New Jersey: John Weley and Sons, 2008.
36. Li Xiang-wei, Qi Yian-fang, "A Data Processing Algorithms for Classification Model Based on Rough Sets," Physics Procedia, vol. 25, 2012, pp. 2025-2029.
37. Robert Susmaga, Roman Slowinski, "Generation of rough sets reducts and constructs based on inter-class and intra-class information," Fuzzy Sets and Systems, In Press, 2014.
38. WenhaoShu, Hong Shen, "Incremental Feature Selection Based on Rough Set in Dynamic Incomplete Data," Pattern Recognition, In Press, 2014.
39. Rechar O. Duda, Peter E. Hart, David G. Stork,Pattern Classification, 2nd ed, Wiley, 2001.
40. J. Han, M. Kamber and J. Pei, Data Mining Concepts and Techniques, 3rd ed., USA: Margan Kaufmann, 2012.
41. Jollife T., Principal Component Analysis, New York: Sringer-Verlag , 1986.

© Copyright: All rights reserved

42. Zhengyou He, Sheng Lin, Yujia Deng, Xiaopeng Li, QingquanQian, “A rough membership neural network approach for fault classification in transmission lines,” Electrical Power and Energy Systems, vol. 61, 2014, pp. 429-439.
43. Julio J. Valdes, Enrique Romero, Alan J. Barton, “Data and knowledge visualization with virtual reality spaces, neural networks and rough sets: Application to cancer and geophysical prospecting data,” Expert Systems with Applications, vol. 39, 2012, pp. 13193-13201.
44. Yu-Neng Fan, Tzu-Liang Tseng, Ching-Chin Chern, Chun-Che Huang, “Rule induction based on an incremental rough set,” Expert Systems with Applications, vol. 36, 2009, pp. 11429-11450.
45. E.A. Rady, A.M. Kozae, M.M.E. Abd El-Monsef, “Generalized rough sets,” Chaos Solitons and Fractals,vol. 21, 2004, pp. 49-53.
46. S. Asharaf, M. NarasimhaMurty, S. K. Shevade, “Rough set based incremental clustering of interval data,” Pattern Recognition Letters, vol. 27, 2006, pp. 515-519.
47. Georg Peters, “Some refinements of rough *k*-means clustering,” Pattern Recognition, vol. 39, 2006, pp. 1481-1491.
48. Ju Li, Xing Wang, Shan Xu, “Prediction of Customer Classification Based on Rough Set Theory,” Procedia Engineering, vol. 7, 2010, pp. 366-370.
49. Salsabil Trabelsi, Zied Elouedi, Pawan Lingras, “Classification systems based on rough set under the belief function framework,“ International Journal of Approximate Reasoning, vol. 52, 2011, pp.1409-1432.
50. Maria Salamo, Maite Lopez-Sanchez, “Rough set based approaches to feature selection for Case-Based Reasoning classifiers,” Pattern Recognition Letters, vol. 32, 2011, pp. 280-292.
51. Salem Chakhar, Ines Saad, “Dominance-based rough set approach for groups in multicriteria classification problems,” Decision Support Systems, vol. 54, 2012, pp. 372-380.
52. Mingquan Ye, Xindong Wu, Xuegang Hu, Donghui Hu, “Anonymizing classification data using rough set theory,” Knowledge-Based system, vol. 43, 2013, pp. 82-94.
53. Yi-Chung Hu, “Rough sets for pattern classification using pairwise-comparison-based tables,” Applied Mathematical Modelling, vol. 37, 2013, pp. 7330-7337.
54. A. Wojna, L. Kowalski, Rseslib: Programmer’s Guide, 2008. Available: http://rsproject.mimuw.edu.pl
55. Zhengcai Lu, Zheng Qin, Yongqiang Zhang, Jun Fang, “A fast feature selection approach based on rough set boundary regions,” Pattern Recognition Letters, vol. 36, 2014, pp. 81-88.
56. James N.K. Liu, Yanxing Hu, Yulin He, “A set covering based approach to find the reduct of variable precision rough set,” Information Sciences, vol. 275, 2014, pp. 83-100.
57. Kai Zheng, Jie Hu, Zhenfei Zhan, Jin Ma, Jin Qi, “An enhancement for heuristic attribute reduction algorithm in rough set,” Expert Systems With Applications, vol. 41, 2014, pp. 6748-6754.
58. Antonio Boggia, Lucia Rocchi, Luisa Paolotti, Francesco Musotti, Salvatore Greco, “Assessing Rural Sustainable Development potentialities using a Dominance-based Rough Set Approach,” Journal of Environment Management, vol. 144, 2014, pp. 160-167.
59. Fan Min, Qinghua Hu, Willium Zhu, “Feature selection with test cost constraint,” International Journal of Approximate Reasoning, vol. 55, 2014, pp. 167-179.
60. Dun Liu, Tianrui Li, Decui Liang, “Incorporating logistic regression to decision-theoretic rough sets for classifications,” International Journal of Approximate Reasoning, vol. 55, 2014, pp.197-210.
61. Bing Zhou, “Multi-class decision-theoretic rough sets,” International Journal of Approximate Reasoning, vol. 55, 2014, pp. 211-224.
62. Aiping Huang, Hong Zhao, Willium Zhu, “Nullity-based matroid of rough sets and its application to attribute reduction,” Information Sciences, vol. 263, 2014, pp. 153-165.
63. Dun Liu, Tianrui Li, Junbo Zhang, “A rough set-based incremental approach for learning knowledge in dynamic incomplete information systems,” International Journal of Approximate Reasoning, In Press, 2014.
64. Lala Septem Riza, Andrzej Janusz, Cristoph Bergmeir, Cris Cornelis Francisco Herrera, Dominik Slezak, Jose Manuel Benitez, “Implementing Algorithms of Rough Set Theory and Fuzzy Rough Set Theory in the R Package “Rough Sets",” Information Sciences, In Press, 2014.
65. Milosz Kadzinski, Salvatore Greco, Roman Slowinski, “Robust Ordinal Regression for Dominance-based Rough Set Approach to Multiple Criteria Sorting,” Information Sciences, In Press, 2014.
66. D. Parikh and R. Polikar, “An Ensembe-Based Increnetal Learning Approach to Data Fusion,” IEE Transactions on systems, Man and cybernetics-part B:cybernetics,vol. 37(2), 2007.
67. Amit Saxena, Nikhil R. Pal and Megha Vora, “Evolutionary methods for unsupervised feature selection using Sammon’s stress function,” Fuzzy Information and Engineering, vol. 2(3), 2010, pp. 229-247.
68. Ivakhnenko, A.G., “Polynomial theory of complex systems,” IEEE Trans. Syst., Man Cybern-I, 1971, pp. 364–378.
69. Amit Saxena, Dovendra Patre and Abhishek Dubey, “Polynomial neural networks architecture: analysis and design,” Computers & Electrical Engineering, vol. 29(6), 2003, pp. 703–725.
70. George J. Klir and Bo Yuan, Fuzzy sets and Fuzzy logic Theory and Applications, New Jersey : Prentice-Hall, 1995.
71. Zhengcai Lu, Zheng Qin, Yongqiang Zhang and Jun Fang, “A fast feature selection approach based on rough set boundary regions,” Pattern Recognition Letters, vol.36, 2014, pp. 81-88.
72. Saroj K. Meher, “Explicit rough-fuzzy pattern classification model,” Pattern Recognition Letters, vol. 36, 2014, pp. 54-61.
73. Xueping Gu, Yang Li and Jinghua Jia, “Feature selection for transient stability assessment based on kernelized fuzzy rough sets and memetic algorithm,” Electrical Power and Energy Systems, vol.64, 2015, pp. 664-670.
74. V. Murlidharan, V. Sugumaran, “Rough set based rule learning and fuzzy classification of wavelet features for fault diagnosis of monoblock centrifugal pump,” Measurement, vol. 46, 2013, pp. 3057-3063.
75. Yi Cheng, “Forward approximation and backward approximation in fuzzy rough sets,” Neurocomputing, In Press, 2014.
76. YuhuaQian, Qi Wang, Honghong Cheng, Jiye Liang, Chuangyin Dang, “Fuzzy-rough feature selection accelerator,” Fuzzy Sets and Systems, In Press, 2014.
77. Bishop C.M., Neural Network for pattern Recognition, Oxford: Oxford University Press, 1995.
78. Simon Haykins, Neural Networks-A Comprehensive Foundation, Singapore : Pearson Education, 1999.
79. Goldberg D. E., Genetic Algorithms in Search Optimization and Machine Learning,Oxford: Pearson Education, 2009.
80. Marco Dorigo and Thomas Stutzle, Ant colony optimization, A Bradford Book, 2004.
81. Maurice Clerc, Particle Swarn Optimization, ISTE Ltd, 2006.
82. H. Hannah Inbarani, Ahmad TaherAzar, G. Jothi, “Supervised hybrid feature selection based on PSO and rough sets for medical diagnosis,” Computer Methods and Programs in Biomedicine, vol. 113, 2014, pp. 175-185.
83. Jong-Hyun Lee, Javad Rahimipour Anaraki, Chang Wook Ahn and Jinung An, “Efficient classification system based on Fuzzy-Rough Feature Selection and Multitree Genetic Programming for intension pattern recognition using brain signal,” Expert Systems with Applications, vol. 42, 2015, pp. 1644-1651.
84. Shima Kashef and Nezamabadi-pour, “An Advanced ACO algorithm for feature subset selection,” Neurocomputing, vol. 147, 2015, pp. 271-279.
85. Ying-Chieh, Tasai, Ching-Hsue Cheng, Jing-Rong Chang, “Entropy-based fuzzy rough classification approach for extracting classification rules,” Expert Systems with Applications, vol. 31, 2006, pp. 436-443.
86. Aboul Ella Hassanien, “Fuzzy rough sets hybrid scheme for breast cancer detection,” Image and Vision Computing, vol. 25, 2007, pp. 172-183.
87. Dongbo Zhang, Yaonan Wang, Huixian Huang, “Rough neural network modelling based on fuzzy rough model and its application to texture classification,” Neurocomputing, vol. 72, 2009, pp. 2433-2443.
88. Shih-Hsun Chang, Shiuan Wan, “Discrete rough set analysis of two different soil-behavior-induced landslides in National Shei-Pa Park, Taiwan,” Geoscience Frontiers , In Press, 2014.
89. Piotr Sawicki, Jacek Zak, “The application of dominance-based rough sets theory to evaluation of transportation systems,” Procidia-Social and Behavioral Sciences, vol. 111, 2014, pp. 1142-1154.
90. Shuang An, Hong Shi, Qinghua Hu, Xiaoqi Li, Jianwu Dang, “Fuzzy rough regression with application to wind speed prediction,” Information Sciences, In Press, 2014.
91. Shan-Wen Zhang, De-Shuang Huang, Shu-Lin Wang, “A method of tumor classification based on wavelet packet transforms and neighbourhood rough set,” Computer in Biology and Medicine, vol. 40, 2010, pp. 430-437.
92. You-Shyang Chen, Ching-Hsue Cheng, “Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry,” Knowledge-Based Systems, vol. 39, 2013, pp. 224-239.
93. Fan Li, Mao Ye, Xudong Chen, “An extension to Rough C- means clustering based on decision-theoretic Rough Sets model,” International Journal of Approximate Reasoning, vol. 55, , 2014, pp. 116-129.

© Copyright: All rights reserved

AUTHOR PROFILE

Prof. Amit Saxena is a Professor in Computer Science at G G University Bilaspur India. He obtained PhD in Computer Science in 1998. He is a member of IEEE, CSI. He has publications in the area of Soft computing, pattern recognition, machine learning.

Leeladhar Kumar Gavel is a Ph.D. Scholar in computer science at G. G. University, Bilaspur India.

Madan Madhaw Shrivas is a Ph.D. Scholar in computer science at G. G. University, Bilaspur India.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved