

Rough Set Theory and its Applications: A recent Survey

¹Neha Saini, ²Dr. Sachin Patel

¹M Tech 4th (C.S.E.), ²Asso. Prof., C.S.E. Department S.I.R.T. Indore M.P. India.

¹nehasaini7893@gmail.com, ²drsachinpatel.sage@gmail.com

Abstract: Rough set theory is a mathematical method proposed by Pawlak . Rough set theory has been developed to manage uncertainties in information that presents missing and noises. Rough set theory is an expansion of the conventional set theory that supports approximations in decision making process. Fundamental of assumption of rough set theory is that with every object of the universe has some information associated it. Rough set theory is correlate two crisp sets, called lower and upper approximation. The lower approximation of a set consists of all elements that surely belong to the set, and the upper approximation of the set constitutes of all elements that possibly belong to the set. The boundary region of the set consists of all elements that cannot be classified uniquely as belonging to the set or as belonging to its complement, with respect to the available knowledge Rough sets are applied in several domains, such as, pattern recognition, medicine, finance, intelligent agents, telecommunication, control theory ,vibration analysis, conflict resolution, image analysis, process industry, marketing, banking risk assessment etc. This paper gives detail survey of rough set theory and its properties and various applications of rough set theory.

Keywords — Rough set, Lower, Upper, Approximation, Boundary, Region,

I. INTRODUCTION

Rough set theory is a mathematical method developed by Pawlak. This method is developed to handle uncertainties from information that contain missing or noises. Lower and upper approximations can be used by rough set when the available information is insufficient to determine. In decision making, rough set methods have a powerful essence in dealing with uncertainties. The rough set theory has been useful in several fields including image processing, data mining, pattern recognition, medical informatics, knowledge discovery and expert systems. In this survey several research works have been combined which used rough set theory with other field. The use of rough set theory to solve a specific complex problem has attracted world-wide attention r research and development. Rough set as a computationally efficient technique it presents a basic significance to many theoretical developments and practical applications. Rough set theory is also used in the areas of machine learning and data mining, decision analysis and intelligent.

II. BASIC CONCEPTS OF ROUGH SETS

Rough set theory proposes is a new mathematical approach to imperfect knowledge, i.e. to vagueness or imprecision. In this approach, boundary region is very important and it can expressed by a vagueness of a set. Rough set concept can be

defined as a means of topological operations, interior and closure, called approximations.

Let a finite set of objects U and a binary relation $R \subseteq U \times U$ be given. The sets U, R are called the universe and an indiscernibility relation, respectively.

The discernibility relation represents our lack of knowledge about elements of U . For simplicity, assume that R is an equivalence relation. A pair (U, R) is called an approximation space, where U is the universe and R is an equivalence relation on U .

Let X be a subset of U , i.e. $X \subseteq U$. The aim is to characterize the set X with respect to R . In order to do it, we need additional notation and basic concepts of rough set theory which are presented below.

By $R(x)$ we denote the equivalence class of R determined by element x . The indiscernibility relation R describes lack of knowledge about the universe U . Equivalence classes of the relation R , called granules, represent an elementary portion of knowledge we are able to perceive due to R . Using only the indiscernibility relation, in general, we are not able to observe individual objects from U but only the accessible granules of knowledge described by this relation.

The set of all objects which can be with certainty classified as members of X with respect to R is called the R -lower approximation of a set X with respect to R , and denoted by

$$\underline{R}(X) = \{x: R(x) \subseteq X\}$$

- The set of all objects which can be only classified as possible members of X with respect to R is called the R -upper approximation of a set X with respect to R , and denoted by

$$\overline{R}(X) = \{x: R(x) \cap X \neq \emptyset\}$$

- The set of all objects which can be decisively classified neither as members of X nor as members of \bar{X} with respect to R called the boundary region of a set X with respect to R , and denoted by

$$RN_R(X) = \overline{R}(X) - \underline{R}(X)$$

Now we are ready to formulate the definition of the rough set notion.

- A set X is called crisp (exact) with respect to R if and only if the boundary region of X is empty.
- A set X is called rough (inexact) with respect to R if and only if the boundary region of X is nonempty.

III. GRANULES OF KNOWLEDGE

The definitions of set approximations can be expressed in terms of granules of knowledge in the following way. The lower approximation of a set is union of all granules which are entirely included in the set; the upper approximation is union of all granules which have non-empty intersection with the set; the boundary region of a set is the difference between the upper and the lower approximation of the set.

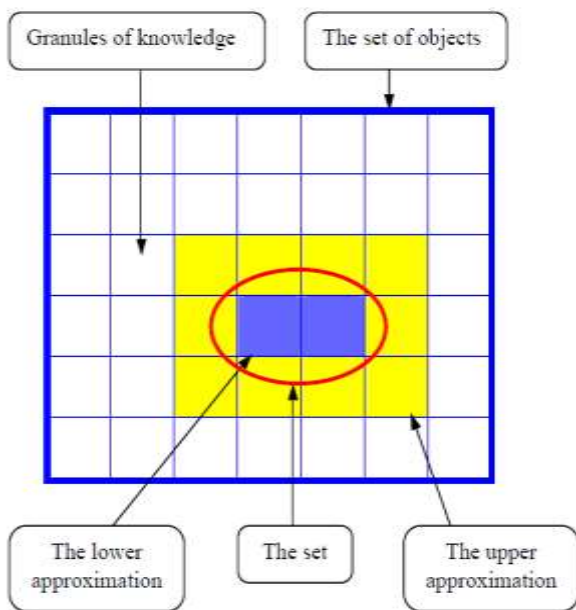


Figure 1.1 Granules of knowledge

IV. ROUGH SETS IN DATA ANALYSIS

A data set is represented as a table, where each row represents a case, an event, a patient, or simply an object. Every column represents an attribute that can be measured for each object; the attribute may be also supplied by a human expert or the user. Such table is called an information system. An information system is a pair $S = (U, A)$ where U is a non-empty finite set of objects called the 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 .

Let us consider a very simple information system shown in Table 1. The set of objects U consists of seven objects: $x_1, x_2, x_3, x_4, x_5, x_6, x_7$ and the set of attributes includes two attributes: *Age* and *LEMS*.

Table 1.1 Simple information systems

Objects	Age	LEMS
x_1	16-30	50
x_2	16-30	0
x_3	31-45	1-25
x_4	31-45	1-25
x_5	46-60	26-49
x_6	16-30	26-49
x_7	46-60	26-49

Table 1.2 Simple information systems with decision

Objects	Age	LEMS	Walk
x_1	16-30	50	Yes
x_2	16-30	0	No
x_3	31-45	1-25	No
x_4	31-45	1-25	Yes
x_5	46-60	26-49	No
x_6	16-30	26-49	Yes
x_7	46-60	26-49	No

Let $X = \{x: \text{Walk}(x) = \text{Yes}\}$, as given by Table 2. In fact, the set X consists of three objects: x_1, x_4, x_6 . Now, we want to describe this set in terms of the set of conditional attributes $A = \{\text{Age}, \text{LEMS}\}$. Using the above definitions, we obtain the following approximations:

The A-lower approximation $AX = \{x_1, x_6\}$

The A-upper approximation $AX = \{x_1, x_3, x_4, x_6\}$

A-boundary region $BN S(X) = \{x_3, x_4\}$

The A-outside region $U - AX = \{x_2, x_5, x_7\}$

V. LITERATURE SURVEY

In 2002 Zdzislaw Pawlak proposed “**Rough set theory and its applications**”. They outlined basic concepts of the theory and illustrated by a simple tutorial example, concerning churn modeling in telecommunications. They used real life applications and used more advanced extensions of the theory and discuss these extensions here. They showed that rough set theory has an overlap with many other theories dealing with imperfect knowledge, e.g., evidence theory, fuzzy sets, Bayesian inference and others. Never the less, the theory can be regarded as an independent, complementary, not competing, discipline in its own rights[1].

In 2004 Zbigniew Suraj proposed “**An Introduction to Rough Set Theory and Its Applications**”. They described data model of concurrent systems discovered from a given information system allows to understand better the structure and behavior of the modeled. They showed that it is possible to represent the dependencies between the processes in information system and their dynamic interactions in graphical way. The presented approach can be treated as a kind of decomposition of a given information system[2].

In 2006 Xin Li San Jose Proposed “**Finding Optimal Reduct for Rough Sets by Using a Decision Tree Learning Algorithm**”. They gives the details about Rough Set theory is a mathematical theory for classification based on structural analysis of relational data. The proposed approach uses the decision tree generated by C4.5 to find the optimal reduct for a relational table. This method does not guarantee finding a minimal reduct, but test results show that the optimal reduct generated by this approach is equivalent or very close to the minimal reduct[3].

In 2007 E. A. Rady et al “**A Modified Rough Set Approach to Incomplete Information Systems**”. They remarked that a rough set theory can be an effective tool to deal with the incomplete information system. It can clarify a very important notation in statistics, which is the scaling. They showed e importance of rough set theory in case of the incomplete information systems and solving some problems in scaling in statistics. They also, reduce the decision rules and the decisions can be introduced in a general form. This point of view causes many problems in the analysis[4].

In 2008 Puntip Pattaraintakorna, Nick Cercone Proposed “**Integrating rough set theory and medical applications**”. They described Medical science is not an exact science in which processes can be easily analyzed and modeled. They present a short survey of ongoing research and a case study on integrating rough set theory and medical application. Issues in the current state of rough sets in advancing

medical technology and some of its challenges are also highlighted [5].

In 2010 Yiyu Yao Proposed “**Three-way decisions with probabilistic rough sets**”. They showed approximates a concept by three regions, namely, the positive, boundary and negative regions. Rules constructed from the three regions are associated with different actions and decisions, which immediately leads to the notion of three-way decision rules[6].

In 2011 Yiyu Yao_ and Xiaofei Deng proposed “**Sequential Three-way Decisions with Probabilistic Rough Sets**”. They generalize three-way decisions with probabilistic rough sets into a sequential three-way framework. This enables us to consider both the cost of various misclassifications and the cost of obtaining the necessary evidence for making a classification decision. They reports some preliminary results on the topic. The exploration of multiple representations of objects for decision-making is a useful direction in granular computing. Typically, decisions made at a higher level of granularity or abstraction may be less accurate or reliable but with a lower cost of resources[7].

In 2012 M. E. Abd El-Monsef et al “**A Comprehensive Study of Rough Sets and Rough Fuzzy Sets on Two Universes**”. They present a framework for the study of rough sets and rough fuzzy sets on two universes of discourse. The proposed model will be more natural in the sense that rough sets approximated by sets on the same universe. They presented a new definition of the lower approximation and upper approximation on two universes through the use of the intersection of right neighborhoods. The properties of rough sets have been simulated by employing this notion. It has been revealed that proposed model that the properties of Pawlak approximation space are attainable if the relation is strong inverse serial[8].

In 2013 Thabet Slimani Proposed “**Application of Rough Set Theory in Data Mining**”. Rough set theory is a new method that deals with vagueness and uncertainty emphasized in decision making. Data mining is a discipline that has an important contribution to data analysis, discovery of new meaningful knowledge, and autonomous decision making. They introduced the fundamental concepts of rough set theory and other aspects of data mining, a discussion of data representation with rough set theory including pairs of attribute-value blocks, information tables reducts, indiscernibility relation and decision tables[9].

In 2014 Y.H.Qian et al Proposed “**An efficient accelerator for attribute reduction from incomplete data in rough set framework**”. They used attribute reduction from large-scale incomplete data is a challenging problem in areas such as pattern recognition, machine learning and data mining. In rough set theory, feature selection from incomplete data

aims to retain the discriminatory power of original features. To address this issue, many feature selection algorithms have been proposed, however, these algorithms are often computationally time-consuming. By integrating the accelerator into a heuristic algorithm, they obtain several modified representative heuristic feature selection algorithms in rough set theory[10].

In 2015 M. Pushpalatha et al proposed “**A Survey: Rough Set Theory in Incomplete Information Systems**”. Rough Set theory has been conceived as a tool to conceptualize, This survey presents an overview of the rough set theory, terms used in the rough sets. Rough sets can be applied to the important process of feature selection and learning. The investigations and developments made in these areas are tabled and discussed. Rough sets can be combined with other techniques when they alone are not able to produce better results. Further, applications of rough sets are numerous; some of the applications is summarized[11].

In 2016 Qinghua Zhang et al proposed “**A survey on rough set theory and its applications**”. After probability theory, fuzzy set theory and evidence theory, rough set theory is a new mathematical tool for dealing with vague, imprecise, inconsistent and uncertain knowledge. They discussed the basic concepts, operations and characteristics on the rough set theory are introduced firstly, and then the extensions of rough set model, the situation of their applications, some application software and the key problems in applied research for the rough set theory are presented[12].

In 2017 B.S.Panda et al proposed “**Retrieving the Missing Information from Information Systems Using Rough Set, Covering Based Rough Set and Soft Set**”. They proposed a study the various aspects of rough set theory, covering-based rough set and soft set theory to handle the missing information systems. Covering based rough set is an extension of the basic rough set. Soft set theory can be applied to problems that contain uncertainties in decision making problems. These concepts are applied to a real life malaria disease dataset to replace missing information in the malaria disease information table and compared the results. They summarized the basic concepts of rough set, covering based set and soft set the manner in which rough set.[13].

In 2018 Andrzej Skowron et al proposed “**Rough sets: past, present, and future**”. They outline some selected past and present research directions of rough sets. They emphasize the importance of searching strategies for relevant approximation spaces as the basic tools in achieving computational building blocks required for approximation of complex vague concepts. They have discussed some issues related to the development of rough sets over 35 years, together with some challenges for the rough set approach, especially in the environment where computations are progressing due to interactions between

physical and abstract granules, and where they can be controlled by performing actions activated on the basis of satisfiability (to a degree) of complex vague concepts, modeled by approximations[14].

In 2019 Yonca Yazirli et al proposed “**Comparison of Algorithms Based on Rough Set Theory for A 3-Class Classification**”. The performance of algorithms is measured based on accuracy, AUC and standard error for a 3-class classification problem on training on test data sets. Based on the test data, the results showed that genetic algorithm over performed the others. Reduction algorithms based on rough set theory for efficient classification with a minimum set of attributes for real estate in Istanbul have been examined. The reduction algorithms were evaluated by using the same classifier: the voting method. The housing unit prices of real estate for sale in different districts of Istanbul obtained from an online web source and 250 real estate were investigated. In the process of determining the best reduction algorithm based on rough set theory, the classification performance of the test data was taken into consideration and the genetic algorithm was chosen as the most successful reduction algorithm[15].

VI. ADVANTAGE OF ROUGH SET

Rough set has following advantages

- It does not need any additional information about data – like probability in statistics, grade of membership in the fuzzy set theory.
- It provides efficient methods and tools for finding hidden patterns in data.
- It allows reducing original data and find minimal sets of data with the same knowledge as in the original data.
- It allows evaluating the significance of data.
- It allows generating in automatic way the sets of decision rules from data.
- It is easy to understand.
- It offers straightforward interpretation of obtained results.
- It is suited for concurrent (parallel/distributed) processing.
- It is easy internet access to the rich literature about the rough set theory, its
- extensions as well as interesting applications,

VII. CONCLUSION

In this paper we proposed basic terminology used in rough set and application of rough set. We gives the details study of rough set based application by various literature survey. In This paper we give the details study that how rough set terminology help to solve various problems and uncertainty for information.

REFERENCES

- [1] Zdzisław Pawlak *Journal of telecommunication and information telecommunication 2002* Institute of Theoretical and Applied Informatics Polish Academy of Sciences Bałtycka 5 44-000 Gliwice, Poland.
- [2] Zbigniew Suraj Zdzisław Pawlak *Polish mathematician and computer scientist*, Polish Academy of Sciences and Warsaw University of Technology. Cairo University in December 2004.
- [3] Xin Li Finding Optimal Reduct for Rough Sets by Using a Decision Tree Learning Algorithm” (2006).
- [4] E. A. Rady, *Corporation Journal of Applied Mathematics and Decision Sciences* Volume 2007, Article ID 58248, 13 pages doi:10.1155/2007/58248.
- [5] PuntipPattaraintakorna, Nick Cercone *Integrating rough set theory and medical applications*, Volume 21, Issue 4, Pages 400–403, April 2008.
- [6] Yiyu Yao *Information Sciences*, Vol. 180, No. 3, pp. 341-353, 2010. Department of Computer Science, University of Regina, Regina, Saskatchewan, Canada S4S 0A2.
- [7] Yiyu Yao_ and Xiaofei Deng *Department of Computer Science*, University of Regina Regina, Saskatchewan, Canada S4S 0A2.
- [8] M. E. Abd El-Monsef, A. M. Kozae, A. S. Salama, R. M. Aqeel *Journal Of Computing*, Volume 4, Issue 3, March 2012, ISSN 2151-9617.
- [9] Thabet Slimani *College of Computer Science and Information Technology*, Taif University 2013.
- [10] Y. H. Qian, J. Y. Liang, *Pattern Recognit.*, vol. 44, no. 8, pp. 1658–1670, Aug. 2014.
- [11] M. Pushpalatha, Dr. V. Anuratha *International Journal of Innovative Research in Computer and Communication Engineering* (An ISO 3297: 2007 Certified Organization) Vol. 3, Issue 8, August 2015.
- [12] Qinghua Zhang , Qin Xie www.sciencedirect.com *Science Direct CAAI Transactions on Intelligence Technology* (2016) .
- [13] B.S.Panda, S.S.Gantayat, Covering Based Rough Set and Soft Set B.S.Panda et al, / (IJCSIT) *International Journal of Computer Science and Information Technologies*, Vol. 7 (3) , 2016, 1403-1407.
- [14] Andrzej Skowron Rough sets: past, present, and future *Natural Computing* (2018) 17:855–876 Published online: 25 July 2018.
- [15] Yonca Yazirli , Betül Kan-Kilinç 3-Class Classification [Kilinç et. al., Vol.7 (Iss.8): August 2019] ISSN- 2350-0530(O), ISSN- 2394-3629(P) DOI: 10.5281/zenodo.3401362 International Journal of Research - Granthaalayah