

Fuzzy Rough Sets and Its Application in Data Mining Field

Megha Kumar¹ and Nidhika Yadav²

¹CSE & IT Dept. ITM University Gurgaon, India

²CSE & IT Dept. ITM University Gurgaon, India

E-mail: ¹meghawonders02@gmail.com, ²nidhikayadav@itmindia.edu

Abstract—Rough set theory is a new method that deals with vagueness and uncertainty emphasized in decision making. The theory provides a practical approach for extraction of valid rules from data. This paper discusses about rough sets and fuzzy rough sets with its applications in data mining that can handle uncertain and vague data so as to reach at meaningful conclusions.

Keywords: Fuzzy logic, Rough Sets, Fuzzy Rough Sets.

1. INTRODUCTION

Fuzzy logic is based on computing "degrees of truth" or "degrees of membership" rather than "true or false" (1 or 0) concept of Boolean logic on which the modern computer is based [1].

Data mining means "knowledge discovery" which refers to extracting meaningful data or hidden patterns from the knowledge base so as to make decisions.

Nowadays data contains some uncertainty or ambiguity. Some of the techniques like rough sets with fuzzy logic are implemented to handle vague and uncertain data.

The remaining of the paper is organized as follows. Section 2 gives the literature review. Section 3 describes fuzzy rough sets. Section 4 outlines the conclusion and future work.

2. LITERATURE REVIEW

Fuzzy logic

Sometimes a Boolean logic based "true or false" values are not sufficient in human reasoning. Fuzzy logic uses the values ranging within the interval between 0 and 1 to describe human reasoning. [1].

The fuzzy membership function is given as [2]:

$$\mu_x(x) \in \langle 0, 1 \rangle$$

Where, X is the set and x represent the element.

Fuzzy membership function has the following properties:

a) $\mu_{U-X}(x) = 1 - \mu_X(x)$ for any $x \in U$

b) $\mu_{X \cup Y}(x) = \max(\mu_X(x), \mu_Y(x))$ for any $x \in U$

c) $\mu_{X \cap Y}(x) = \min(\mu_X(x), \mu_Y(x))$ for any $x \in U$

3. ROUGH SETS

Rough set theory is a new mathematical approach to uncertain knowledge. The problem of imperfect or uncertain knowledge has been tackled by philosophers, mathematicians and logicians. Recently, in the area of artificial intelligence it too became a critical issue for computer scientists. There are many approaches to handle and manipulate uncertain knowledge, of which the most successful one is the fuzzy set concept. The main advantage of using rough sets is that it does not need any additional or prior information about data [2].

Rough set theory has many interesting applications in the areas of machine learning, AI and cognitive sciences, knowledge acquisition, knowledge discovery from databases, decision analysis, expert systems, pattern recognition and inductive reasoning.

Some of the basic concepts of rough set theory are presented as:

Information System

The information system is described as,

$$IS = (U, A)$$

Where, U is the universe (finite set of objects, $U = \{x_1, x_2, \dots, x_n\}$) and A is the set of attributes

(variables, features). Each attribute $a \in A$ defines an information function $f_a: U \rightarrow V_a$, where V_a is the set of values of attribute a , called as domain of attribute a [2].

Decision Table

A decision table describes information systems as:

$$S = (U, A = C \cup \{d\}),$$

Where, attributes in C are called as condition attributes and d is the decision attribute [2].

Approximation of sets

Let $S = (U, R)$ represents an approximation space and X be a subset of U .

The lower approximation of X represents those elements which doubtlessly belong to set X .

The lower approximation of X by R in S is defined as

$$\underline{RX} = \{e \in U \mid [e] \subseteq X\} \text{ and}$$

The upper approximation of X represents those elements which possibly belong to set X .

The upper approximation of X by R in S is defined as

$$\overline{RX} = \{e \in U \mid [e] \cap X \neq \emptyset\}$$

Where, $[e]$ denotes the equivalence class containing e .

The boundary set $BNR(X)$ is defined as $\overline{RX} - \underline{RX}$

A set X is rough in S if its boundary set is nonempty.

The approximation of sets is shown in fig.1.

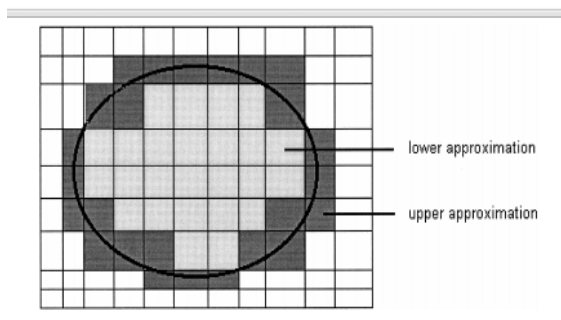


Fig. 1: Approximation of sets in rough set theory

Discernibility and indiscernibility relations

Suppose a finite set of attributes are used to define a finite set of objects. By considering any subset of attributes, discernibility and indiscernibility relations can be defined.

Two objects are indiscernible or equivalent if and only if they have the same values for all attributes in the set. In other words, in terms of the given set of attributes, it is impossible to differentiate the two objects.

Two objects are discernible if and only if for at least one attribute, they have different values. Since the indiscernibility and discernibility relations are defined with respect to at least one attribute or the set of all attributes, respectively, they may be viewed as strong indiscernibility and weak discernibility [3].

Core and reduct of attributes

The concepts of core and reduct are two fundamental concepts of the rough sets theory. The reduct is the essential part of an IS, which can discern all objects discernible by the original IS. A set of reducts are all possible minimal subsets of attributes, which lead to the same number of elementary sets as the whole set of attributes [2,4].

The core is the common part of all reducts or the intersection of different reducts gives the core of the attributes [2,4].

4. ROUGH SETS IN DATA MINING

- Rough set theory in materials science: Rough sets provide an algorithmic approach for understanding the properties of the materials, which further helps in designing new products [5].
- LERS Software: The LERS software is used to generate decision rules from data. The rules extracted are used in classification of new cases. In LERS software, the rule generation starts from uncertain or imperfect data (e.g., data characterized by missing attribute values or inconsistent cases). Data discretization is used to deal with numerical attribute. LERS uses lots of methods which help in handling of missing attribute values. For inconsistent data (data characterized by same values of all attributes belonging to two different targets), LERS calculate lower and upper approximations of all sets.

LERS systems are used in the medical field, where they can be used to diagnose preterm birth by comparing the effects of warming devices for postoperative patients, even used in diagnosis of melanoma [5].

- Other applications: Other applications of rough set theory can be found in music fragment classification, medical diagnosis and control, pattern recognition, including speech recognition, and handwriting recognition [5].

5. FUZZY ROUGH SETS

A fuzzy-rough set is a generalisation of a rough set, derived from the approximation of a fuzzy set in a crisp approximation space. This corresponds to the case where the values of

conditional attribute are crisp and the decision attribute values are fuzzy.

The main focus of fuzzy-rough sets is to define lower and upper approximation of the set when universe of fuzzy set becomes rough because of equivalence relation, or transforming the equivalence relation to similar fuzzy relation.

Rough sets can be expressed by a fuzzy membership function $\mu \rightarrow \{0, 0.5, 1\}$ to represent the negative, boundary, and positive regions. In this model, the elements belonging to lower approximation or positive region have a membership value of one, those belonging to boundary region have a membership value of 0.5 and those elements belonging to upper approximation or negative region have a membership value of 0 [6].

Fuzziness is integrated into rough sets which use fuzzy membership values to qualify levels of roughness in boundary region. Therefore, the membership values of boundary region objects can range from 0 to 1, instead of only having membership value of 0.5 [6].

Suppose, R is an equivalence relation, which is imposed on the universe U . The equivalence class is expressed as fuzzy sets $F = \{F_1, F_2, \dots, F_H\}$, when the classes to which the elements attribute are ambiguous. F_j is a fuzzy set, $j \in \{1, 2, \dots, H\}$. The fuzzy lower and upper approximations are given as [7,8]:

$$\mu_{PX}(F_i) = \inf_x \max\{1 - \mu_{F_i}(x), \mu_X(x)\} \quad \forall i$$

$$\mu_{PX}(F_i) = \sup_x \min\{\mu_{F_i}(x), \mu_X(x)\} \quad \forall i$$

They show the degree of possibility and inevitability of fuzzy set F .

Fuzzy Rough Sets in prediction of k-nearest neighbour

There are some uncertainties which appear in moving objects and k-nearest neighbors' prediction. The space-uncertainties arising in moving object's future direction is considered [9]. The uncertainty in predicted position is represented using fuzzy membership degree that actual position locates around predicted position. Fuzzy-rough set is used to analyze fuzzy membership degree of moving objects' predicted position and its $k + m$ nearest neighbor, so as to get more accurate k-nearest neighbor [9].

6. FUZZY ROUGH SETS IN CLUSTERING

Cluster analysis is the task of grouping a set of objects in such a way that objects in the same cluster are more similar to each other than to those in other clusters. Cluster analysis is an important function in data mining. A good clustering algorithm should possess expansibility, fast process high dimensions of data, and is insensitive to noise, so fuzzy rough sets are used to handle this [10].

7. FUZZY ROUGH SETS IN ATTRIBUTE REDUCTION

Here, a method is used to compute reducts for fuzzy rough sets, where only the minimal elements in the discernibility matrix are considered. First, relative discernibility relations of conditional attribute are defined and relative discernibility relations are used to characterize minimal elements in the discernibility matrix. Then, an algorithm to compute the minimal elements is developed. Finally, novel algorithms to find proper reducts with the minimal elements are designed [11].

8. FUZZY ROUGH SETS IN CLASSIFICATION

A hybrid scheme that combines the advantages of fuzzy sets and rough sets is used in classifying the objects to the respective classes.

An application of breast cancer imaging has been chosen and this hybridization scheme has been applied to test their ability and accuracy to classify the breast cancer images into two classes: cancer or non-cancer. The introduced scheme starts with fuzzy image processing as pre-processing techniques to enhance the contrast of the whole image; to extract and the regions of interest and then to enhance the edges surrounding the region of interest. A subsequently extract features from the segmented regions of the interested regions using the gray-level co-occurrence matrix is presented. Rough sets approach for generation of all reducts that contains minimal number of attributes and rules are introduced. Finally, these rules are passed to a classifier for discrimination for different regions of interest to test whether they are cancer or non-cancer. A new rough set distance function is presented to measure the similarity. The experimental results show that the hybrid scheme applied in this study performs well reaching over 98% in overall accuracy with minimal number of generated rules [12].

9. CONCLUSION & FUTURE WORK

Fuzzy rough sets have various applications in the data mining field that are used to handle uncertainty and vagueness present in the data. These are not capable of handling indeterminate relations that exist in the data.

As an extension of our work, we project to develop a hybrid model of rough sets and neutrosophic logic that would be able to handle indeterminacy and give more realistic results compared to fuzzy rough sets.

REFERENCES

- [1] Fuzzy logic. (2015, January 6). In Wikipedia, The Free Encyclopedia. Retrieved 13:19, January 14, 2015, from http://en.wikipedia.org/w/index.php?title=Fuzzy_logic&oldid=641189015
- [2] Pawlak, Z. (1982). Rough sets. *International Journal of Computer & Information Sciences*, 11(5), 341-356.
- [3] Zhao, Y., Yao, Y., & Luo, F. (2007). Data analysis based on discernibility and indiscernibility. *Information Sciences*, 177(22), 4959-4976.
- [4] Walczak, B., & Massart, D. L. (1999). Rough sets theory. *Chemometrics and intelligent laboratory systems*, 47(1), 1-16.
- [5] Slimani, T. (2013). Application of Rough Set Theory in Data Mining. arXiv preprint arXiv:1311.4121.
- [6] Kasemsiri, W., & Kimpan, C. (2001). Printed thai character recognition using fuzzy-rough sets. In *TENCON 2001. Proceedings of IEEE Region 10 International Conference on Electrical and Electronic Technology* (Vol. 1, pp. 326-330). IEEE.
- [7] Richard Jensen "Combining rough and fuzzy sets for feature selection" Book published in 2005.
- [8] Tian, H., Kang, X. Y., Zhang, J. N., & Han, S. S. (2012, June). Application of fuzzy rough sets in patterns recognition of bearing. In *Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR2MSE), 2012 International Conference on* (pp. 731-734). IEEE.
- [9] Prediction of Moving Objects' K-Nearest Neighbor Based on Fuzzy-Rough Sets Theory. In *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on* (Vol. 1, pp. 407-411). IEEE.
- [10] Jiangping, L., Baochang, P., & Yuke, W. (2009, May). The Fuzzy Rough Sets and Algorithm of Fuzzy Rough Clustering Based on Grid. In *Software Engineering, 2009. WCSE'09. WRI World Congress on* (Vol. 4, pp. 538-542). IEEE
- [11] Chen, D., Zhang, L., Zhao, S., Hu, Q., & Zhu, P. (2012). A novel algorithm for finding reducts with fuzzy rough sets. *Fuzzy Systems, IEEE Transactions on*, 20(2), 385-389.
- [12] Hassaniien, A. (2007). Fuzzy rough sets hybrid scheme for breast cancer detection. *Image and vision computing*, 25(2), 172-183.