



Data with Rough Attributes and Its Reduct Analysis

Prem Kumar Singh^{1,*}

¹Department of Computer Science and Engineering,
Gandhi Institute of Technology and Management-Visakhapatnam,
Andhra Pradesh 530045, India

* Correspondence: premsingh.csjm@gmail.com , premsingh.csjm@yahoo.com
ORCID: 0000-0003-1465-6572

Abstract: Recent time many researchers focused on dealing the uncertainty and its characterization. The precise approximation of uncertainty in many-valued data set is one of the major tasks. It becomes more difficult in case the given data sets are non-Euclidean. Hence the rough fuzzy set and its graphical visualization is introduced in this paper for knowledge processing tasks.

Keywords: Fuzzy Rough graph; Knowledge representation; Many-valued attributes; Non-Euclidean geometry; Rough Set; Rough graph

1. Introduction

The uncertainty and its approximation is considered as one of the major tasks for soft computing researchers [1-2]. It become more crucial while dealing the data with non-Euclidean [3-4] or cubic set [5]. To deal with this issue rough set and its properties is introduced by Pawlak [6-7]. The rough set given a way to approximate the given data sets based on its lower and upper approximation. Due to which the properties of rough set is applied in various fields for multi-decision process [8-11] as well as its graphical visualization [12-16]. This gave a way to characterize the uncertainty in three-way decision space [17-19]. In this process, a problem is addressed while dealing the data with rough attributes and its reduct. To solve this problem current paper focused on illustrating the data with rough attributes, its contextual representation and reducts.

The motive is to characterize them based on lower, upper and boundary regions as shown in Figure 1. The objective is to provide a basic understanding for new researchers for dealing the data with rough attributes.

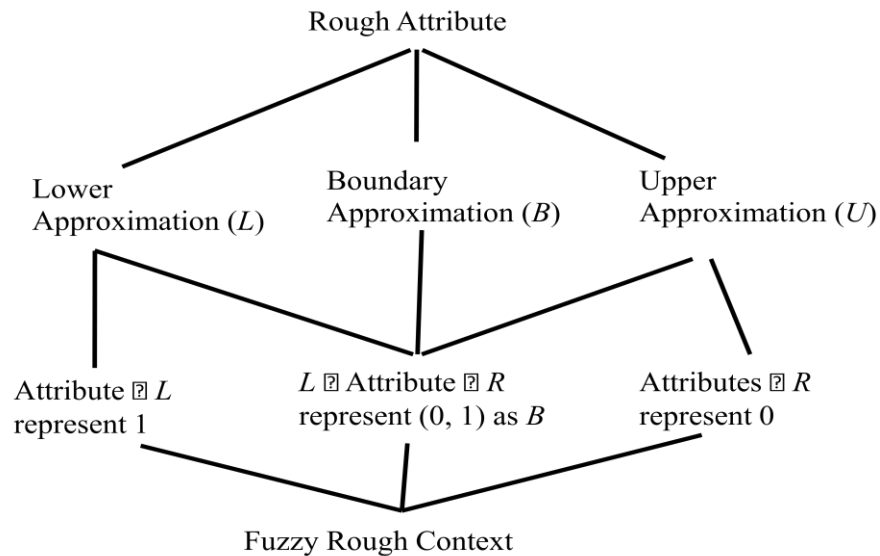


Figure 1: The motivation of this paper and its objective

Rest part of the paper is organized as follows: Section 2 provides background Fuzzy and Rough set. Section 3 contains the proposed method for characterization of rough context and its fuzzy membership-values with an illustrative example in Section 4. Section 5 contains conclusions followed by acknowledgements and references.

2. Background

This section provides the basic background to represent the data with rough attributes and its set approximation for decision making process.

2.1 Information System

The Table 1 represents the data with information system where row represents the a set of non-empty objects $\{O1, O2, \dots, O6\}$, the columns represents the attributes (A) with defined multi-valued information (R) in the given universe (U). In this way it provides an information system with tuple of 4-attributes $S = (U, R, V, f)$. It can be also represented as $S := (U, A)$, where A is non-empty set of attributes set such that for each $R^1 \subseteq A$ where $R = (C \cup D)$ i.e. subsets of conditional (C) and decision attributes (D). Table 1 represents following as conditional i.e. $C = \{A1, A2, A3\}$ and decision attributes i.e. $D = \{A4\}$, where V_i is the set of values of i^{th} attribute i.e. $A_1 := \{yes, yes, yes, no, no, no\}$, $f : R \rightarrow V$ is a description or information objective function. These data can be analyzed using the indiscernible relation and its set approximation.

Table 1: The data with Rough Attribute and its contextual representation

Objects	Attributes			Decision Flue (A_4)
	Temperature(A_1)	Headache(A_2)	Muscles pain(A_3)	
O_1	normal	yes	yes	no
O_2	high	yes	yes	yes
O_3	Very-high	yes	yes	yes
O_4	normal	no	yes	no
O_5	high	no	no	no
O_6	Very-high	no	yes	yes

2.2 Indiscernible Relation

The associated equivalence relation on universe (U) for a given nonempty subset of attribute set with any $R^1 \subseteq R$ is defined as $IND_S(R^1) := \{(x, y) \in U^2 \mid \forall_{r \in R^1} (r(x) = r(y))\}$, where $(x, y) \in IND_S(R^1)$ are defined as object x and y are inducible by attribute of from R^1 . The equivalence class of R^1 - indiscernible relations are denoted as $[x]_{R^1}$. The pair of $(U, IND_S(R^1))$, called estimated space. As for example: The set consists of nonempty subset of attributes “Headache” and “Muscle pain” i.e., A_1 and A_2 . $IND(A_1, A_2) := \{\{O_1, O_2, O_3\}, \{O_4, O_6\}, \{O_5\}\}$ containing three indiscernible sets also called elementary sets, one definable set $\{O_1, O_2, O_3, O_5\}$. Similarly, the other possible non-empty indiscernible subsets of C are as follows:

$$IND(A_1), IND(A_2), IND(A_3), IND(A_1, A_2, A_3), IND(A_1, A_3), IND(A_1, A_2), IND(A_2, A_3).$$

In this way the given information system can be defined based on approximating the set.

2.3 Set Approximations

It can be observed that the equivalence relations induce a partitioning of universe(U), can be used to create a new subset that are more often of interest have the same values for decision attribute(D). Let $R^1 \subseteq R$ be a desired subset of U . The description for R^1 is desired when we can determine the membership status of each object in U w.r.t R^1 , if the $[x]_{R^1}$ containing partial overlaps with any of the indiscernible defined for an object with an ambiguity. Such an object may not be distinguished, therefore the description of R^1 is defined in-terms of lower ($P_*(R^1)$), upper ($P^*(R^1)$) approximation sets respectively also called as positive (POS), negative (NEG) and boundary regions (BND) as follows:

$$P_*(R^1) = POS(R^1) = \{x \in U \mid [x] \subseteq R^1\}, \text{ where } [x] \text{ denotes the equivalence-class of } x. \dots\dots\dots(i)$$

$$P^*(R^1) = NEG(R^1) = \{x \in U \mid [x] \cap R^1 \neq \emptyset\} \dots\dots\dots(ii)$$

$$BND(R^1) = P^*(R^1) - P_*(R^1) \dots\dots\dots(iii)$$

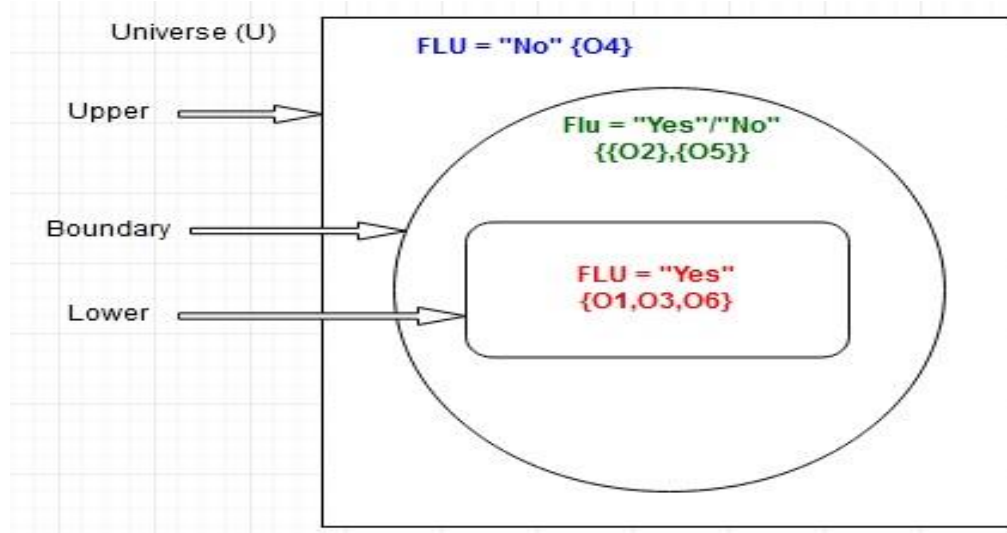


Figure 2: The rough set theory approximations of Table 1

A set R^1 for which $P_*(R^1) = P^*(R^1)$ is called as “exact set” otherwise rough-set w.r.t P . If an object $x \in P_*(R^1)$, then it belongs to target-set R^1 certainly. For any target or decision attribute subset $D \subset U$ and conditional attributes $C \subset R$, D is obtained as roughset when $P_*(Y) \neq P^*(Y)$. The roughness of set D w.r.t C is identified as follows: $P_C(Y) = 1 - \frac{|P_*(Y)|}{|P^*(Y)|}$, where $Y \neq \phi$ (if $Y = \phi$, then $P_C(Y) = 0$); $|\cdot|$ denote the cardinality essence of a set. Similarly, correctness is defined as $\alpha_c(Y) = \frac{|P_*(Y)|}{|P^*(Y)|}$, then apparently $0 \leq \alpha_c(Y) \leq 1$. If $\alpha_c(Y) = 1$, then Y is said to be CRISP w.r.t C , $\alpha_c(Y) < 1$ then it is “ROUGH”. If an object, $x \in P^*(R^1)$, it cannot be determined whether it belongs to the target or not. If an object, $x \in BND(R^1)$, then it does not belong to target-set R^1 certainly. A set is said to be “ROUGH”, if it's $BND(R^1) \neq \phi$, otherwise the set is “CRISP”. As for example, the objects O_2 and O_5 can not be distinguished (i.e indiscernible) from anyone of the attributes shown in Table 1. Hence, the objects present in $BND(R^1)$ region is $\{O_2, O_5\}$, which can not be classified properly based on knowledge O_2 and O_5 as shown in Figure 2. It shows that O_2 & O_5 are boundary line cases. The remaining objects in lower and upper regions as follows:

$$P_*(Flu = "yes") = \{O_1, O_3, O_6\}, P_*(Flu = "no") = \{O_4\}, BND(R^1) = \{O_2, O_5\} \dots \dots \dots (iv)$$

$$P^*(Flu = yes) = \{O_1, O_2, O_3, O_5, O_6\}, \dots \dots \dots (v)$$

$$P^*(Flu = "no") = \{O_2, O_4, O_5\}, BND(R^1) = \{O_2, O_5\} \dots \dots \dots (vi)$$

In this way, the set of approximation and its rough membership can be defined. To achieve this goal, a step by step method is illustrated in the next section.

3. The Rough Membership, Core and Reduct Analysis

It can be observed that, the data with rough attributes can be approximated based on lower, upper and boundary regions. The problem is how to characterize them in a membership function. To resolve this issue step by step demonstration is discussed in this section as given below:

3.1 Defining the Rough Membership Functions

The set approximations can be defined based on the degree of overlapping regions between the $\{X\}$ -set and the equivalence membership relation $R_{(X)}$, to which the object x belong to a set or not, it is defined using the membership function shown below:

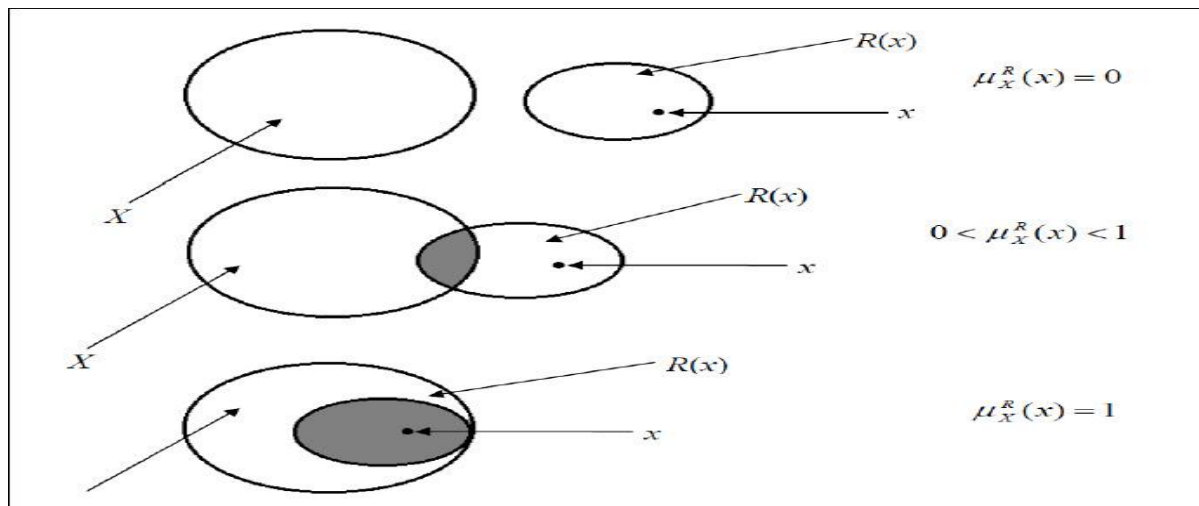


Figure 3: The characterization of rough-attributes as Membership Functions

$\mu_X^R : U \rightarrow \leq 0,1 >$, i.e., function accepts only the values 1 and 0 respectively, where $\mu_X^R(x) = \frac{|X \cap R(x)|}{|R(x)|}$ and $|\cdot|$ called the cardinality essence of an attribute (X). The meaning of rough -membership function indicates the assumptions and boundary regions of a set (X) is defined as below equations and its diagrammatic representations are shown in Figure 3.

$$R_*(X) = \{x \in U : \mu_X^R(x) = 1\} \quad \dots\dots\dots(vii)$$

$$R^*(X) = \{x \in U : \mu_X^R(x) \geq 0\} \quad \dots\dots\dots(viii)$$

$$R_*(X) = \{x \in U : 0 < \mu_X^R(x) < 1\} \quad \dots\dots\dots(ix)$$

3.2 Dependency of Decision System Attributes

The major issue with the Decision System is to identify same or indiscernible-objects that may appear several times, due to this the attributes of $(C \cup D)$ leads to superfluous for most of the Machine Learning Classifiers to design an

Doi : <https://doi.org/10.54216/JNFS.020104>

Received July 26, 2021 Accepted: Jan 08, 2022

effective Classification Model. Finding dependency and removal of such attributes may not degrade the performance of classification models. The decision system with A_i attributes totally depends on predicted attribute set D , and its relation called as $A_i \rightarrow D$, if all the values in attribute A_i are uniquely identified (classify) by the values of D . i.e., A_i depends on D , there exists functional dependency. In more general, the concept discusses about the partial dependencies of attributes i.e., only some set of A_i values are classifying the values of decision attribute (D). The RST introduce a degree of dependency measure to calculate dependency between two subset of attributes ($A_i, D \subseteq R$) is denoted as $\lambda_{A_i}(D)$. It is defined as shown below :

$$\lambda_{A_i}(D) = \frac{\text{card}(POS_{A_i}(D))}{\text{card}(U)}, \text{ where } POS_{A_i}(D) = \bigcup_{X \in U / IND(D)} A_{i*}(X) \dots\dots\dots(x)$$

The set $POS_{A_i}(D)$, positive region containing possible elements of U that can be uniquely distinguished from the partition $\frac{U}{IND(D)}$ by A_i . The objects of $\lambda_{A_i}(D)$ represents fraction of total no. of objects in the universe (U) that can be properly classified the elements of decision attribute D . If A_i totally depends on D then $\lambda_D(A_i) = 1$; else $\lambda_D(A_i) < 1$.

For better understanding the concept from above table (), the dependency of FLU (A_4) on Temperature (A_3), we observe that the values of (A_3) uniquely identifies some values of decision attribute (A_4), i.e., (A_3 , very high) $\Rightarrow$ (A_4 , yes), similarly (A_3 , normal) $\Rightarrow$ (A_4 , no), but (A_3 , high) $\neq$ (A_4 , yes), hence there exist partial dependence between A_3 and A_4 . To determine $\lambda_{A_3}(A_4)$ using above equation as shown below:

$$U = \{O_1, O_2, O_3, O_4, O_5, O_6\} \text{ and } \frac{U}{IND(A_4)} = \{\{O_1, O_2, O_3, O_6\}, \{O_4, O_5\}\}$$

$$POS_{A_3}(A_4) = \{O_3, O_6\} \cup \{O_4\} = \{O_3, O_4, O_6\},$$

$$\text{Thus } \lambda_{A_3}(A_4) = 3/6 = 0.5. \text{ Similarly, } \lambda_{A_1}(A_4) = 0 \text{ and } \lambda_{A_2}(A_4) = 0.$$

3.3 Accuracy Approximation

For a given real time decision system $S = (U, R, V, f)$, for any target variable subset $X \subseteq U$ and its attribute subset $A \subseteq R$, the roughness of set X w.r.t A about the classification model can be defined as below Eq. (3.8).

$$P_A(X) = 1 - \frac{|R_+(X)|}{|R_*(X)|}, \text{ obviously } 0 \leq P_A \leq 1, \text{ when } X \neq \phi; \text{ if } X = \phi, \text{ then } P_A(X) = 0; \text{ if } P_A(X) = 1, \text{ then } X \text{ is said to be "CRISP" w.r.t } A; \text{ similarly when } P_A(X) < 1, \text{ then } X \text{ is called "ROUGH" w.r.t } A.$$

3.4 Reducts

One often raises the question, how to remove irrelevant or redundant/superfluous attributes from a decision system by preserving its basic intrinsic properties including appropriate representation space for the learning system. RST allows identifying equivalence or in-discernible class relations, finds a minimal attribute subset that differentiate the entire classes of decision-attribute without deteriorating the performance of the classification model or towards

decision making applications. There are several such minimal attribute subsets called “REDUCTS” of the original set which retain the accuracy like the original set, and thus reduce the computational time.

3.5 Core

The set of conditional attributes (A_i) are unreliable in T , denoted as $CORE(A_i)$, such that $CORE(A_i) = \cap RED(A_i)$ i.e intersection of all $relative_{reducts}$ is termed as $relative_{core}$, each object of the core belongs to some reduct with an important minimal subset of attribute set, and further none of its objects could be excluded.

For example table 3.1, have two possible reducts i.e., $RED_1 = \{A_3, A_1\}$ and $RED_2 = \{A_3, A_2\}$ w.r.t decision attribute A_4 , the intersection (core) of the decision Table 1 is A_3 . Table 2 and Table 3 represents the minimal decision tables of RED_1 and RED_2 . In this way the rough provides a way to deal with multi-valued data for decision making process.

Table 2: The RED_1 for data with rough attributes shown in Table 1

Objects	Attributes			Decision Flue (A_4)
	Temperature(A_1)		Headache(A_3)	
O_1	normal		yes	no
O_2	high		yes	yes
O_3	Very-high		yes	yes
O_4	normal		no	no
O_5	high		no	no
O_6	Very-high		no	yes

Table 3: The RED_2 for data with rough attributes shown in Table 1

Objects	Attributes			Decision Flue (A_4)
	Temperature(A_1)		Muscles pain(A_3)	
O_1	normal		yes	no
O_2	high		yes	yes
O_3	Very-high		yes	yes
O_4	normal		yes	no
O_5	high		no	no
O_6	Very-high		yes	yes

In this way, the core and reduct of given rough context can be investigated. However, the characterization of rough attributes and its visualization is another issues. The author will try to focus on this issue in near future for knowledge processing tasks.

4. Conclusions

This paper introduces step by step method for dealing data with rough attributes, its approximation as well as rough membership function. The core reduct is also illustrated with an example. In near future the author will focus on defining the fuzzy rough membership and its graphical visualization for knowledge processing tasks.

Acknowledgements: Author thanks the editorial team for the valuable time.

Funding : Author declares that, there is no funding for this paper.

Conflicts of Interest: Author declares that, there is no conflict of interest.

Ethics approval: This article does not contain any studies with human or animals participants.

References

- [1] Singh P. K., “Three-way fuzzy concept lattice representation using neutrosophic set”, *International Journal of Machine Learning and Cybernetics*, Vol 8, Issue 1, pp. 69-79, 2017.
- [2] Singh PK, Ch. Aswani Kumar, “Concept lattice reduction using different subset of attributes as information granules”, *Granular Computing*, Vol. 2, Issue 3), pp. 159–173, 2017
- [3] Singh PK, “AntiGeometry and NeutroGeometry characterization of Non-Euclidean data sets”, *Journal of Neutrosophic and Fuzzy Systems*, Vol 1, Issue 1, pp. 24-33, DOI: <https://doi.org/10.54216/JNFS.0101012>
- [4] Singh PK, “Data with Non-Euclidean Geometry and its Characterization,” *Journal of Artificial Intelligence and Technology*, 2021, Vol. 2, Issue 1, pp-3-8., DOI: [10.37965/jait.2021.12001](https://doi.org/10.37965/jait.2021.12001)
- [5] Singh PK, “Cubic graph representation of concept lattice and its decomposition”, *Evolving System*, doi: 10.1007/s12530-021-09400-6
- [6] Pawlak Z, “Rough sets”, *Int. J. Comput. Inf. Sci. Vol.*, pp. 341–356, 1982
- [7] Pawlak Z, “Rough set theory and its applications to data analysis,” *Cybern Syst Vol. 29*, Issue 7, pp. 661–688, 1998
- [8] He T, Chan Y, Shi K, “Weighted rough graph and its application,” In: *Proceedings of IEEE Sixth Int Conf Intell Syst Des Appl* 1:486–492, 2006
- [9] He T, “Rough properties of rough graph,” *Appl Mech Mater Vol 157–158*, pp. 517–520, 2012
- [10] He T, “Representation form of rough graph,” *Appl Mech Mater, Vol. 157–158*, pp. 874–877, 2012
- [11] Liang M, Liang B, Wei L, Xu X, “Edge rough graph and its application,” In: *Proc. Of eighth International Conference on Fuzzy Systems and Knowledge Discovery 2011*, pp. 335–338
- [12] Wang S, Zhu Q, Zhu W, Min F, “Graph and matrix approaches to rough sets through matroids. *Information Sciences*, Vol. 288, pp. 1–11, 2014

Doi : <https://doi.org/10.54216/JNFS.020104>

Received July 26, 2021 Accepted: Jan 08, 2022

- [13] Li W, Huang Z, Jia X, Cai X, “Neighborhood based decision-theoretic rough set models,” *International Journal of Approximate Reasoning*, Vol. 69, pp. 1–17, 2016
- [14] Noor R, Irshad I, Javaid I, “Soft rough graphs”. arXiv preprint arXiv:1707.05837, 2017
- [15] Fariha Z, Akram M, “A novel decision-making method based on rough fuzzy information,” *Int J Fuzzy Syst* Vol. 20, Issue 3, pp. 1000–1014, 2018
- [16] Rehman N, Shah N, Ali MI, Park C, “Uncertainty measurement for neighborhood based soft covering rough graphs with applications,” *RACSAM*, Vol. 113, pp. 2515–2535, 2019
- [17] Mathew B, John SJ, Garg H, “Vertex rough graphs,” *Complex Intell. Syst.* Vol 6, pp. 347– 353, 2020
- [18] Yao YY, “Relational interpretations of neighborhood operators and rough set approximation operators,” *Inf. Sci.*, Vol. 101, pp. 239–259, 1998
- [19] Yao YY, “Three-way decisions with probabilistic rough sets,” *Inf. Sci.*, Vol. 180, pp. 341–353, 2010