Big Data and Computing Visions

www.bidacv.com

Big. Data. Comp. Vis. Vol. 1, No. 2 (2021) 96–100.

Paper Type: Original Article

Rough Sets Theory and Its Extensions for Attribute Reduction: A Review

Sadegh Eskandari

Department of Computer Science, University of Guilan, Rasht, Iran; eskandari@guilan.ac.ir.

Citation:

Eskandari, S. (2021). Rough sets theory and its extensions for attribute reduction: a review. *Big data and computing visions, 1* (2), 96-100.

Received: 16/02/2021 Reviewed: 27/03/2021 Revised: 13/04/2021 Accept: 05/05/2021

Abstract

The rough sets theory is a mathematical tool to express vagueness by means of boundary region of a set. The main advantage of this implementation of vagueness is that it requires no human input or domain knowledge other than the given data set. Several efforts have been made to make close the rough sets theory and machine learning tasks. In this regard several extensions and modifications of the original theory are proposed. This paper provides the basic concepts of the theory as well as its well-known extensions and modifications.

Keywords: Rough set theory, Data science, Data set.

1 | Introduction

The rough sets theory is introduced by Pawlak [5] to express vagueness by means of boundary region of a set. The main advantage of this implementation of vagueness is that it requires no human input or domain knowledge other than the given data set [8] and [4]. This section describes the fundamentals of the theory.

1.1 | Information System and Indiscernibility

An information system is a pair $IS = (U, F)$, where U is a non-empty finite set of objects called universe and F is a non-empty finite set of features such that $f: U \to V_f$, for every $f \in F$. The set V_f is called the value set or domain of f. Information system in rough sets theory is analogous with data set in unsupervised machine learning and classification tasks. A decision system is an information system of the form $IS = (U, F, d)$, where d is called the decision feature. data set in a supervised classification and learning can be seen as a decision system, where instances are the objects of universe, features are the elements of F and labels represent decision feature values.

For any set $B \subseteq F \cup \{d\}$, we define the B-indiscernibility relation as:

Corresponding Author: eskandari@guilan.ac.ir

https://doi.org/10.22105/bdcv.2021.142085

$$IND_{IS}(B) = \{(x, y) \in U \times U | \forall f \in B, f(x) = f(y)\}. \quad (1)$$

If (x, y) belongs to $IND_{IS}(B)$, x and y are indiscernible according to the feature subset B. Equivalence classes of the relation $IND_{IS}(B)$ are denoted $[x]_B$ and referred to as B-elementary sets. The partitioning of U to B-elementary subsets is denoted $U/IND_{IS}(B)$ or simply U/B. Generating such a partitioning is a common computational routine, that effects the performance of any rough set based operation.

1.2 | Lower and Upper Approximations

The fundamental notions of rough sets are lower and upper approximations of sets. let $B \subseteq F$ and $X \subseteq U$, the B −lower and B −upper approximations of X are defined as follow:

$$\underline{B}X = \{x | [x]_B \subseteq X\}. \quad (2)$$

$$\overline{B}X = \{x | [x]_B \cap X \neq 0\}. \quad (3)$$

The $\underline{B}X$ and $\overline{B}X$ approximations define information contained in B [4]. If $x \in \underline{B}X$, it is certain that it belongs to X and if $x \in \overline{B}X$, we can only say that x may belong to X.

By the definition of $\underline{B}X$ and $\overline{B}X$,the objects in U can be partition into three regions which are the positive, boundary and negative regions.

$$POS_B(X) = \underline{B}X. \quad (4)$$

$$BND_B(X) = \overline{B}X - \underline{B}X. \quad (5)$$

$$NEG_B(X) = U - \overline{B}X. \quad (6)$$

1.3 | Dependency

Discovering dependencies between attributes is an important issue in data analysis. Let D and C be subsets of $F \cup \{d\}$. It is said that D depends on C in a degree k $(0 \leq k \leq 1)$, denoted $C \Rightarrow_k D$, if

$$k = \gamma(C, D) = \frac{|POS_C(D)|}{|U|}. \quad (7)$$

Where

$$POS_C(D) = \bigcup_{X \in U/D} \underline{C}X.$$

Called a positive region of the partition U/D with respect to C. This region is the set of all elements of U that can be uniquely classified to blocks of the partition U/D, by means of C [7]. Functional dependency of D and C denoted $C \Rightarrow D$ is an special case of dependency where $\gamma(C, D) = 1$. In this case we say that all values of attributes from D are uniquely determined by values of attributes from C.

2 | Rough Set Extentions

Several efforts has been made to make close rough sets theory and machine learning tasks. However traditional rough sets based which make it ineffective in real world applications [2], [3] and [4]; First, it only operates effectively with datasets containing discrete values and therefore it is necessary to perform a discretization step for real-valued attributes. Second, rough set is highly sensitive to noisy data. Finally, rough set methods examine only the information contained within the lower approximation of a set and ignore the information contained in the boundary region.

Therefore several extensions to the original theory have been proposed to overcome such shortcomings. Four notable extensions are Variable Precision Rough Sets (VPRS) [8], Tolerance Rough Set Model (TRSM) [6], Fuzzy Rough Sets (FRS) [2] and [1] and an extension to dependency measure proposed in [4].

VPRS attempts to overcome the traditional rough sets shortcomings by generalizing the standard set inclusion relation ($\subseteq$) [8]. In the generalized inclusion relation, a set X is considered to be a subset of Y if the proportion of elements in X which are not in Y is less than a predefined threshold. However, the introduction of a suitable threshold requires more information than contained within the data itself. This is contrary to the rough sets theory and OSF consideration of operating with no domain knowledge.

TRSM uses a similarity relation instead of indiscernibility relation to relax the crisp manner of classical rough sets theory [6]. As equivalence classes (elementary sets) in classical rough sets, tolerance classes are generated using similarity relation in TRSM, which are used to define lower and upper approximations. TRSM has two deficiencies which are contrary two our OSF considerations; First, it needs a tolerance threshold to generate tolerance classes, which like VPRS this threshold is human defined. Second, the time complexity of generating all tolerance classes, using attribute subset B, is $\Theta(|B||U|^2)$, which is equal to worst-case time complexity of PARTITION algorithm.

FRS uses fuzzy equivalence classes generated by a fuzzy similarity relation to represent vagueness in data [1] and [2]. Fuzzy Lower and upper approximations are generated based on fuzzy equivalence classes. These approximations are extended versions of their crisp notions in classical rough sets, except that in the fuzzy approximations, elements may have membership degree in the range [0,1]. FRS needs no extra knowledge to define operations on a given dataset, however as tolerance classes in TRSM, generating fuzzy equivalence classes in FRS is an expensive routine ($\Theta(|B||U|^2)$).

3 | Rough Set Modifications

In addition to rough sets extensions, there are also some modifications, which does not change classical rough sets principals. Parthaláin et al. [4] redefines the dependency notion in classical rough sets to deal with useful information that may be contained in the boundary region. Unlike the other three extensions, this extension does not redefine the lower and upper approximations in classical rough sets, therefore it needs no human input knowledge to deal with available data.

3.1 | Useful Information in Boundary Region

Almost all the classical rough set based attribute reduction methods use only the information contained in the positive region. However the boundary region may also contain useful information that are ignored in this methods [4]. Such scenario is common in real-valued datasets, where some adjacent values may placed in different regions because of crisp manner of classical rough sets. Measuring the proximity of objects in the boundary region to the objects in positive region could help to qualify the information contained in boundary region. The method proposed uses a distance metric to calculate such proximities [4].

Let X be a set of objects and B a subset of attributes. The mean positive region, which is the mean of all object attribute values in $POS_B(X)$, is defined as

$$m = \left\{ \frac{\sum_{x \in \underline{B}X} f(x)}{|POS_B(X)} : \forall f \in B \right\}. \tag{8}$$

The proximity of any object $y \in BND_B(X)$ from the mean positive region is defined as

Eskandari | Big. Data. Comp. Vis. 1(2) (2021) 96-100

$$\delta_B(m, y) = \begin{cases} d(m, y) & \text{if} |POS_B(X)| \neq 0, \\ 0 & \text{if} |POS_B(X)| = 0. \end{cases} \quad (9)$$

Where d can be any distance function such as euclidean distance metric.

The proximity of the boundary region to the positive region is defined as

$$\omega(C, D) = \begin{cases} \psi_B^{-1} & \text{if} |BND_C(D)| \neq 0, \\ 1 & \text{if} |BND_C(D)| = 0. \end{cases} \quad (10)$$

Where

$$\psi_B = \sum_{y \in BND_B(X)} \delta_B(m, y). \quad (11)$$

This proximity measure combined with rough-set dependency value create a new evaluation measure M as

$$M(C, D) = \frac{\gamma(C, D) + \omega(C, D)}{2}. \quad (12)$$

3.2 | Impurity Rate and Noise Resistant Measure

The noise resistant measure attempts to qualify the information that may be unseen duo to the crisp manner of the inclusion relation in defining lower approximations [10]. This measure uses an impurity rate value to calculate the noisy portion of a set. Let A and B be two sets. The impurity rate of A with respect to B can be defined as follow:

$$c(A, B) = \frac{|A - B|}{|A|}. \quad (13)$$

This value calculates the portion of the elements that should be eliminated from A to make it totally included in B. It is important to note that if $c(A, B) > 0.5$, the impurity of A with respect to B is more than its impurity with respect to $\overline{B}$. In this case, A could be supposed as a noisy version of $\overline{B}$ and all elements in $A \cap B$ will constitute the noisy portion of A. Therefore, the B-related information that could be retrieved after removing impurities from A can be formulated as

$$\xi(A, B) = \begin{cases} 1 - c(A, B) & \text{if} c(A, B) \leq 0.5, \\ 0 & \text{if} c(A, B) > 0.5. \end{cases} \quad (14)$$

This formulation can be applied to elementary sets to extract information that may be unseen in calculating lower approximations. To do this, a noise measure function, ϕ, is defined as

$$\phi_B(X) = \frac{\sum_{Y \in U/B} \xi(Y, X) \quad [\xi(Y, X) \neq 1]}{|U/B|}. \quad (15)$$

This function quantifies the possibility of transferring some objects from boundary to the positive region of a set, if the noisy elements could be removed.

Let C and D be two attribute sets. The noisy dependency of D on C can be defined as follow:

$$\nu(C, D) = \sum_{Y \in U/D} \phi_C(X). \quad (16)$$

The noisy dependency operates on boundary region as proximity measure [10]. However the proximity measure considers each point in the boundary region separately and calculates its distance from the positive region, while the noisy dependency considers subsets of objects to measure their transmission possibility to the positive region. Therefore the two values are combined to create a new measure for evaluating boundary region as

$$\tau(C,D) = \omega(C,D) + \nu(C,D). \tag{17}$$

This new measure can be used alongside the classical dependency. As one measure only operates on the objects in boundary region and the other only on the objects in positive region, the two operators are combined to create a noise resistant evaluation measure ρ:

$$\rho(C,D) = \frac{\tau(C,D) + \gamma(C,D)}{2}. \tag{18}$$

4 | Conclusion

This paper considered the rough sets theory as a mathematical tool to express vagueness by means of boundary region of a set. The main advantage of this implementation of vagueness is that it requires no human input or domain knowledge other than the given data set, therefore, several efforts has been made to make close the rough sets theory and machine learning tasks. In this regard many modifications and extensions to the original theory is proposed in the literature. The paper provided a review to three extentions, VPRS, TRSM and FRS as well as two modifications proposed in [4], [9] and [10].

References

[1] Dubois, D., & Prade, H. (1992). Putting rough sets and fuzzy sets together. In *intelligent decision support* (pp. 203-232). Springer, Dordrecht.

[2] Jensen, R., & Shen, Q. (2004). Semantics-preserving dimensionality reduction: rough and fuzzy-rough-based approaches. *IEEE Transactions on knowledge and data engineering*, *16*(12), 1457-1471.

[3] Jensen, R., Tuson, A., & Shen, Q. (2014). Finding rough and fuzzy-rough set reducts with SAT. *Information sciences*, *255*, 100-120.

[4] Parthaláin, N., Shen, Q., & Jensen, R. (2009). A distance measure approach to exploring the rough set boundary region for attribute reduction. *IEEE transactions on knowledge and data engineering*, *22*(3), 305-317.

[5] Pawlak, Z. (1982). Rough sets. *International journal of computer & information sciences*, *11*(5), 341-356.

[6] Skowron, A., & Stepaniuk, J. (1996). Tolerance approximation spaces. *Fundamenta informaticae*, *27*(2, 3), 245-253.

[7] Swiniarski, R. W., & Skowron, A. (2003). Rough set methods in feature selection and recognition. *Pattern recognition letters*, *24*(6), 833-849.

[8] Walczak, B., & Massart, D. L. (1999). Rough sets theory. *Chemometrics and intelligent laboratory systems*, *47*(1), 1-16.

[9] Ziarko, W. (1993). Variable precision rough set model. *Journal of computer and system sciences*, *46*(1), 39-59.

[10] Javidi, M. M., & Eskandari, S. (2017). A noise resistant dependency measure for rough set-based feature selection. *Journal of intelligent & fuzzy systems*, *33*(3), 1613-1626.