ATLANTIS PRESS

Advances in Computer Science Research, (ACSR), volume 52

2016 International Conference on Computer Engineering and Information Systems (CEIS-16)

Inconsistent Neighborhoods and Relevant Properties In Neighborhood Rough Set Models

Shu-Jiao Liao, Qing-Xin Zhu, Rui Liang

School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

E-mail: sjliao2011@163.com, qxzhu@uestc.edu.cn, harry8384@163.com

Shu-Jiao Liao

School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, China

E-mail: sjliao2011@163.com

Xin-Zheng Niu

School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

E-mail: xinzhengniu@uestc.edu.cn

***Abstract*-Rough set theory is an important branch of data mining and machine learning, among which neighborhood rough set is presented to deal with numerical data and hybrid data. In this paper, we propose a new concept called inconsistent neighborhood, and explore the relations between it and the existing notions in neighborhood rough set models. Some interesting properties are obtained accordingly. These properties can generate some new solutions to compute the quantities in neighborhood rough set, which are often more direct and more quick to obtain the results than the previous solutions.**

Keywords*-*Inconsistent neighborhood, rough set, lower and upper approximations, positive region, reduct.

I. INTRODUCTION

Rough set theory, which was proposed by Pawlak [1], is an effective tool to address uncertain data and reduce data dimensionality. Up to now, except classical rough set [1, 2], a series of extended models of rough set have been developed, such as decision-theoretic rough set [3, 4], dominance-based rough set [5], multi-granulation rough set [6] and so on. Information granulation, set approximation and attribute reduction are three key issues in rough set methodology. Information granulation refers to segmenting the universe, namely the whole set of objects, into different subsets with some certain criteria [7]. The generated subsets are often called information granules or elemental concepts. While set approximation is to approximately describe arbitrary subset of the universe with the elemental concepts [8]. As for attribute reduction, it means selecting a suitable attribute subset (also called a reduct) to reduce the data dimensionality and meanwhile keep the ability of original decision system [4, 9]. Although each kind of rough set models has its own characteristics, there are some common concepts in these models, such as lower and upper approximations, positive region, reduct and so on.

Most of rough set models are only applicable for nominal data, whereas numerical data and hybrid data exist widely in real applications. To deal with the two kinds of data, some discretizing methods are employed in data preprocessing to transform numerical attributes into nominal attributes [10], but information loss may happen in the discretization process. To address this issue, neighborhood rough set was proposed [11], which has been verified to be a powerful mechanism to handle numerical data and hybrid data. Neighborhoods play a crucial role in neighborhood rough set models. According to a certain criterion (usually a specific distance function), a family of neighborhood granules can be generated from the universe, and then these neighborhoods can be used to approximate any object subsets or decision classes. For an object, its neighborhoods often contain not only the objects with the same class as it but also those with different classes.

In this paper, we give a new concept called inconsistent neighborhood. For a given object, its inconsistent neighborhoods include only the objects having different classes with it. We explore the relations between inconsistent neighborhood and the existing fundamental notions in neighborhood rough set models. Relevant properties are discussed, which reveal that the introduction of inconsistent neighborhood gives new characterizations for existing concepts in neighborhood rough set. Moreover, some new solutions are provided for computing the quantities in neighborhood rough set, which are usually more direct and more quick to obtain the results than the previous solutions. The rest of the paper is organized as follows. In Section 2, we mainly review some fundamental concepts and properties of neighborhood rough set. In Section 3, we introduce inconsistent neighborhoods and discuss relevant properties. Finally, we conclude the paper in Section 4.

II. KEY CONCEPTS AND PROPERTIES IN NEIGHBORHOOD ROUGH SET MODELS

In this section, we mainly review the key concepts and properties in rough sets, especially neighborhood rough set. The concepts and properties can be found in [2, 11, 12].

We start from decision system, which is a fundamental concept in data mining domain including rough sets.

Definition 1. A decision system (DS) S is the 5-tuple:

$$S = (U, C, D, V = \{V_a \mid a \in C \cup D\}, I = \{I_a \mid a \in C \cup D\}),$$

where U is a finite set of objects called the universe, C is

the set of conditional attributes, D is the set of decision attributes with only discrete values, V_a is the set of values for each $a \in C \cup D$, $I_a : U \to V_a$ is an information function for each $a \in C \cup D$.

For convenience, the above $\{V_a \mid a \in C \cup D\}$ and $\{I_a \mid a \in C \cup D\}$ are often denoted as V and I respectively. In most applications, $D = \{d\}$, that is, we are given only one decision attribute called the class. If $|D| > 1$, we can construct $|D|$ decision systems, with each having only one class. In neighborhood rough set, the decision system is also called neighborhood decision system (NDS), and the attribute values of numerical conditional attributes are often normalized to facilitate data processing. The normalization approach is to employ the linear function $y = (x - min)/(\max - \min)$, where x is the initial value, y is the normalized value, min and $\max$ are the minimal and maximal value of the attribute domain, respectively.

Neighborhood, namely neighborhood granule, which plays an important role in neighborhood rough set, has been defined as follows.

Definition 2. Given $x_i \in U$, $B \subseteq C$ and $\delta > 0$, the neighborhood of x_i with respect to attributes B and radius δ is defined as:

$$\delta_B(x_i) = \{x_j \in U \mid \Delta_{B_o}(x_i, x_j) = 0 \wedge \Delta_{B_u}(x_i, x_j) \le \delta\} \quad (1)$$

where Δ is a distance function, $\wedge$ is "and" operator, B_o and B_u are the subsets of B which contain only nominal attributes and numerical attributes respectively, namely $B_o \cup B_u = B$ and $B_o \cap B_u = \phi$.

We use the frequently-used metric Euclidean distance in the paper. Concretely, assuming that $x_1, x_2 \in U$, $B = \{a_1, a_2, \cdots, a_K\}$, and $v(x, a_i)$ denotes the value of sample x on attribute a_i, then Euclidean distance is

$$\Delta_B(x_1, x_2) = \sqrt{\sum_{i=1}^{K} |v(x_1, a_i) - v(x_2, a_i)|^2} \quad (2)$$

Other distance functions can be found in [13].

Two types of monotonicity have been discussed for the neighborhood granules in the two following propositions .

Proposition 1. Let $S = (U, C, D, V, I)$ be a neighborhood decision system (NDS), $B_1 \subseteq B_2 \subseteq C$. We have $\forall x \in U, \delta_{B_1}(x) \supseteq \delta_{B_2}(x)$.

Proposition 2. Let $S = (U, C, D, V, I)$ be a NDS, $B \subseteq C$, $\delta_1 \le \delta_2$. We have $\forall x \in U, \delta_1(x) \subseteq \delta_2(x)$.

Lower and upper approximations, positive region and boundary region are fundamental issues in rough set theory. Their definitions in neighborhood rough set have been given as follows.

Definition 3. Let $S = (U, C, D, V, I)$ be a NDS, and $X_1, X_2, \cdots, X_K$ be the object subsets with decisions 1 through K . The lower and upper approximations of decision D with respect to $B \subseteq C$ are defined as

$$\underline{N_B} D = \bigcup_{i=1}^{K} \underline{N_B}(X_i), \overline{N_B}(D) = \bigcup_{i=1}^{K} \overline{N_B}(X_i), \quad (3)$$

where

$$\begin{aligned} \underline{N_B}(X) &= \{x \in U \mid \delta_B(x) \subseteq X\} \\ \overline{N_B}(X) &= \{x \in U \mid \delta_B(x) \cap X \neq \phi\} \end{aligned} \quad (4)$$

are the lower and upper approximations of object subset X . The decision boundary region of D with respect to attributes B is defined as

$$BN_B(D) = \overline{N_B}(D) - \underline{N_B} D.$$

The lower approximation $\underline{N_B} D$ is also called positive region and denoted by $POS_B(D)$ .The relations of above concepts have been given, which are

(1) $\overline{N_B}(D) = U$;

(2) $POS_B(D) \cap BN_B(D) = \phi$;

(3) $POS_B(D) \cup BN_B(D) = U$.

From the relations, it is known that

$$BN_B(D) = U - POS_B(D) \quad (5)$$

Reduct, which is a subset of attributes that has the same approximating power as the whole set of attributes, is an important concept in rough sets.

Proposition 3. Let $S = (U, C, D, V, I)$ be a decision system. Any $R \subseteq C$ is a decision-relative reduct if and only if:

(1) $POS_R(D) = POS_C(D)$;

(2) $\forall a \in R, POS_{R-\{a\}}(D) \subset POS_R(D)$.

III. INCONSISTENT NEIGHBORHOODS AND RELEVANT PROPERTIES

In this section, we introduce the concept of inconsistent neighborhood, and use it to describe the existing notions in

neighborhood rough set. Some interesting properties are obtained accordingly.

For an object, each of its neighborhoods may exist some objects with different decision classes from the object, namely inconsistent objects. We call the set of inconsistent objects as inconsistent neighborhood, whose definition is given as follows.

Definition 4. Let $S=(U,C,D,V,I)$ be a NDS. Given $x_i \in U$, $B \subseteq C$ and $\delta > 0$, the inconsistent neighborhood of x_i with respect to attributes B and radius δ is defined as

$$in_B(x_i)=\{x_j \in U \mid x_j \in \delta_B(x_i), D(x_j) \neq D(x_i)\} \quad (6)$$

Naturally, according to Equation (1), Equation (6) is equivalent to

$$in_B(x_i)=\{x_j \in U \mid \triangle_{B_o}(x_i,x_j)=0 \wedge \triangle_{B_u}(x_i,x_j) \leq \delta, D(x_j) \neq D(x_i)\} \quad (7)$$

where B_o and B_u are the nominal-attribute subset and the numerical-attribute subset of B, respectively. It means that the inconsistent neighborhoods can be obtained based on neighborhoods or by using Equation (7) directly.

The following properties can be obtained for inconsistent neighborhoods.

Proposition 4. Let $S=(U,C,D,V,I)$ be a NDS. Given any $x, x_i, x_j \in U$ and $B \subseteq C$, we have

(1) $in_B(x_i) \subset \delta_B(x_i)$;

(2) $x_j \in in_B(x_i) \Leftrightarrow x_i \in in_B(x_j)$;

(3) $x \in POS_B(D) \Leftrightarrow in_B(x)=\phi$.

Proof. (1) and (2) can be known immediately from Definition 4.

(3) Let $X_1, X_2, \cdots, X_K$ be the object subsets with decisions 1 through K. $x \in POS_B(D) \Leftrightarrow \exists X_i (1 \leq i \leq K), \delta_B(x) \subseteq X_i \Leftrightarrow \forall y \in \delta_B(x), D(y)=D(x) \Leftrightarrow in_B(x)=\phi$.

Note that, according to the essence of lower and upper approximations of object subsets, we can rewrite Equation (4) as follows:

$$\begin{aligned} \underline{N_B}(X) &= \{x \in X \mid \delta_B(x) \subseteq X\}, \\ \overline{N_B}(X) &= X \cup \{x \notin X \mid \delta_B(x) \cap X \neq \phi\}. \end{aligned} \quad (8)$$

Compared with Equation (4), Equation (8) is more explicit and the computational efficiency of lower and upper approximations can be improved by using it. Further, based on Definition 4 and Proposition 4, Equation (8) can be rewritten in a new form by using inconsistent neighborhoods, namely, we have

Proposition 5. Let $S=(U,C,D,V,I)$ be a NDS and $B \subseteq C$. For any $X \subseteq U$, we have

$$\begin{aligned} \underline{N_B}(X) &= \{x \in X \mid in_B(x)=\phi\}, \\ \overline{N_B}(X) &= X \cup \{x \notin X \mid in_B(x) \cap X \neq \phi\}. \end{aligned} \quad (9)$$

Moreover, according to Proposition 4 and Equation (5), we can obtain a new formulation for the positive region and the boundary region.

Proposition 6. Let $S=(U,C,D,V,I)$ be a NDS and $B \subseteq C$. We have

$$\begin{aligned} POS_B(D) &= \{x \in U \mid in_B(x)=\phi\}, \\ BN_B(D) &= \{x \in U \mid in_B(x) \neq \phi\}. \end{aligned} \quad (10)$$

In general, Propositions 5-6 describe the lower and upper approximations, positive region and boundary region from a new perspective.

Similarly with neighborhoods, there are two types of monotonicity for inconsistent neighborhoods according to Propositions 1-2 and Definition 4.

Proposition 7. Let $S=(U,C,D,V,I)$ be a NDS, $B_1 \subseteq B_2 \subseteq C$. We have $\forall x \in U, in_{B_1}(x) \supseteq in_{B_2}(x)$.

Proposition 8. Let $S=(U,C,D,V,I)$ be a NDS, $B \subseteq C$, $\delta_1 \leq \delta_2$. We have $\forall x \in U, in_1(x) \subseteq in_2(x)$.

Furthermore, we can obtain the following proposition.

Proposition 9. Let $S=(U,C,D,V,I)$ be a NDS, $C=\{a_1,a_2,\cdots,a_N\}$ and $B \subseteq C$. Assuming that $\{B_i\}_{1 \leq i \leq L} (2 \leq L \leq N)$ are disjoint subsets of B which satisfy $\bigcup_{i=1}^{L} B_i = B$, then for any $x \in U$, we have

$$in_B(x) \subseteq \bigcup_{i=1}^{L} in_{B_i}(x), \quad (11)$$

where "=" holds when at most one of B_i contains numerical attributes.

Proof. Based on Proposition 7, For any $x \in U$ and any $B_i (1 \leq i \leq L, 2 \leq L \leq N)$, $in_B(x) \subseteq in_{B_i}(x)$, so $in_B(x) \subseteq \bigcup_{i=1}^{L} in_{B_i}(x)$. "=" holds when at most one of B_i contains numerical attributes according to Equation (7).

According to Propositions 3 and 4, we obtain the following proposition, which can be used as an alternative definition of a reduct.

Proposition 10. Let $S=(U,C,D,V,I)$ be a NDS. Any $R\subseteq C$ is a reduct if and only if:

(1) $\forall x\in POS_C(D), in_R(x)=\phi$;

(2) $\forall a\in R, \exists x\in POS_C(D), s.t. in_{R-\{a\}}(x)\neq\phi$.

In general, the introduction of inconsistent neighborhoods gives a new characterization for the key concepts in neighborhood rough set models. Some interesting properties are generated, which can provide some new solutions to compute the quantities in neighborhood rough set. The new solutions are analyzed as follows:

- In previous methods, the reducts cannot be known until the positive regions or related values such as the dependency degrees have been computed [11]. Now the reducts can be captured according to the case of inconsistent neighborhoods directly, which will speed up the process of attribute reduction.
- In existing work, the positive regions and boundary regions cannot be obtained until the lower and upper approximations of object subsets have been calculated. Now they can be gained immediately by using the inconsistent neighborhoods.
- The introduction of inconsistent neighborhoods induces that there are often two or more solutions for the quantities in neighborhood rough set. For example, the reducts, positive regions and boundary regions can be found through using the traditional methods, or using our methods mentioned in (1) and (2). Besides, we can compute the lower and upper approximations of object subsets by using neighborhoods according to Equation (4) or (8), or using inconsistent neighborhoods according to Equation (9).

IV. CONCLUSIONS

Traditional neighborhood rough set models employ neighborhoods to construct the theoretical framework. In this paper, we introduced a new concept called inconsistent neighborhood and discussed the relations between it and the existing concepts in neighborhood rough set. A number of interesting properties were obtained, by which some new solutions were provided to compute the quantities in neighborhood rough set. In fact, the inconsistent neighborhood concept is suitable for not only neighborhood rough set, but also some other types of rough set models. We will further study the inconsistent neighborhoods and relevant properties in other kinds of rough set models.

ACKNOWLEDGEMENTS

This work is supported in part by the National Natural Science Foundation of China under Grant No. 61300192, the Education Department of Fujian Province, China under Grant No. JAT160291, and the Fundamental Scientific Research Project for the Central Universities under Grant No. ZYGX2014J052.

REFERENCES

[1] Z. Pawlak, “Rough sets”, International Journal of Computer and Information Sciences, vol. 11, pp. 341–356, 1982.

[2] Z. Pawlak and A. Skowron, “Rudiments of rough sets”, Information Sciences, vol. 177, pp. 3–27, 2007.

[3] Y.Y. Yao and Y. Zhao, “Attribute reduction in decision-theoretic rough set models”, Information Sciences, vol. 178, pp. 3356–3373, 2008.

[4] S.J. Liao, Q.X. Zhu and F. Min, “Cost-sensitive attribute reduction in decision-theoretic rough set models”, Mathematical Problems in Engineering, vol. 2014, Article ID 875918, 9 pages, 2014.

[5] S. Greco, M. Inuiguchi and R. Slowinski, “Dominance-based rough set approach using possibility and necessity measures”, In: Proceedings of Rough Sets and Current Trends in Computing, Volume 2475 of LNCS, pp. 85–92, 2002.

[6] Y.H. Qian, J.Y. Liang, Y.Y. Yao and C.Y. Dang, “Mgrs: a multi-granulation rough set”, Information Sciences, vol.180, pp. 949–970, 2010.

[7] L. Zadeh, “Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic”, Fuzzy Sets and Systems, vol. 19, pp.111–127, 1997.

[8] S. Wang, T.R. Li, C. Luo, and H. Fujita, “Efficient updating rough approximations with multi-dimensional variation of ordered data”, Information Sciences, vol. 372, pp. 690–708, 2016.

[9] H.M. Chen, T.R. Li, Y. Cai, C. Luo, and H. Fujita, “Parallel attribute reduction in dominance-based neighborhood rough set”, Information Sciences, vol. 373, pp. 351–368, 2016.

[10] C. Su and J. Hsu, “An extended chi2 algorithm for discretization of real value attributes”, IEEE Transactions on Knowledge and Data Engineering, vol. 17, pp. 437–441, 2005.

[11] Q. Hu, D. Yu, J. Liu, and C. Wu, “Neighborhood rough set based heterogeneous feature subset selection”, Information Sciences, vol. 178, pp. 3577–3594, 2008.

[12] Y.Y. Yao, “A partition model of granular computing”, Lecture Notes in Computer Science, vol. 3100, pp. 232–253, 2004.

[13] D.R. Wilson and T.R. Martinez, “Improved heterogeneous distance functions”, Artificial Intelligence Research, vol. 6, pp. 1–34, 1997.