

FUZZY AND ROUGH SET

Combining Fuzzy Set and Rough Set for Inductive Learning

Hong Jing¹, Lu Jingui¹, Shi Feng²

¹*Department of Computer Science and Engineering, Nanjing University of Technology, Nanjing, 210009;* ²*National Die & Mold CAD Engineering Research Center, Shanghai Jiaotong University, Shanghai, 200030*

Abstract: A fuzzy-rough set model is presented based on the extension of the classical rough set theory. The continuous attributes are fuzzified. The indiscernibility relation in classical rough set is extended to the fuzzy similarity relation. Then an inductive learning algorithm based on fuzzy-rough set model (FRILA) is proposed. Finally, with comparison to the decision tree algorithms, the effectiveness of the proposed method is verified by an example.

Key words: Fuzzy set, Rough set, Fuzzy similarity relation, Inductive learning

1. INTRODUCTION

In the early 1980s, Pawlak Z proposed rough set theory, which combines knowledge with classification and provides a new approach to vague and uncertain data analysis [1]. However, there are some limitations with the classical rough set. The original rough set cannot deal with the continuous attributes well. And it is based on the indiscernibility relation. Pawlak proposed that both fuzzy set and rough set were not competitive but complementary [2]. Dubois and Prade also proposed that they were related but distinct and complementary theories [3]. Hence, it is possible to combine the two theories.

2. FUZZY-ROUGH SET MODEL

2.1 Fuzzifying the continuous attributes

Practically, there are many continuous or numerical attributes in the decision table. Each attribute a is fuzzified into k linguistic values T_i , $\forall i=1, \dots, k$. The slop of triangular membership functions are selected in the way that adjacent functions cross at the membership value 0.5, so the only parameters to be determined are the set of k centers $M=\{m_i, i=1, 2, \dots, k\}$. The center m_i can be calculated through Kohonen's feature-map algorithm.

2.2 New definitions based on fuzzy similarity relation

The classical lower and upper approximations are originally introduced with reference to an indiscernibility relation (reflexive, symmetric, and transitive). Practically, it can be extended to fuzzy similarity relation.

In order to obtain the partition of U given the fuzzy similarity relation \tilde{R} , an algorithm is designed as follows: *Algorithm 1*:

Input: fuzzy similarity matrix \tilde{R} and level value λ

Output: $U/IND(\tilde{R}_\lambda)$

1) Calculate normal similarity relation matrix \tilde{R}_λ ; 2) $x_i \in U, X \leftarrow \phi, Y \leftarrow \phi$; 3) $j \leftarrow 0$; 4) If $r_{ij}=1$ and $x_j \notin X$, then $X \leftarrow X \cup \{x_j\}, Y \leftarrow Y \cup \{x_j\}$; 5) $j \leftarrow j+1$; 6) If $j < n$, then GOTO 4; otherwise, GOTO next step; 7) If $card(Y) > 1$, then select $x_i \in Y$ and $Y \leftarrow Y - \{x_i\}$, GOTO 3; otherwise, GOTO next step; 8) Output the set X and let $U \leftarrow U - X$; 9) If $U = \phi$, then end; otherwise, GOTO 2. Where $card(Y)$ denotes the cardinality of set Y .

Considering a subset $X \subseteq U$ and a fuzzy similarity relation \tilde{R}_λ^A defined on U , the lower approximation of X , $\tilde{R}_\lambda^A_-(X)$, and upper approximation of X , $\tilde{R}_\lambda^A_+(X)$, are respectively defined as follows: $\tilde{R}_\lambda^A_-(X) = \cup \{Y: Y \in U/IND(\tilde{R}_\lambda^A), Y \subseteq X\}$; $\tilde{R}_\lambda^A_+(X) = \cup \{Y: Y \in U/IND(\tilde{R}_\lambda^A), Y \cap X \neq \emptyset\}$. Assuming $U/IND(\tilde{R}_\lambda^C)$ and Y are two partitions on U , where $U/IND(\tilde{R}_\lambda^C) = \{X_1, X_2, \dots, X_k\}$ and $Y = \{Y_1, Y_2, \dots, Y_r\}$, the positive region $POS_\lambda^C(Y)$ is defined as follows: $POS_\lambda^C(Y) = \cup \{\tilde{R}_\lambda^C_-(Y_i) : Y_i \in Y\}$.

2.3 Fuzzy similarity relation based attribute reduction

Assuming a condition attribute set C and a decision attribute set D , the degree of dependency of C on D , denoted by $\gamma(C, D)$, is defined as: $\gamma(C, D) = card(POS_C(D)) / card(U)$, where $card(X)$ denotes the cardinality of set X and $0 \leq \gamma(C, D) \leq 1$. According to the definition of the

degree of dependency, the attribute significance for every attribute $a \in C-R$ can be defined as follows: $SIG(a, R, D) = \gamma(R \cup \{a\}, D) - \gamma(R, D)$.

In order to obtain the minimal reduction, a hierarchy attribute reduction algorithm is constructed as follows: *Algorithm 2*:

Input: decision table $T = \langle U, C \cup D, V, f \rangle$

Output: the minimal attribute reduction set R .

1) Let $R \leftarrow \phi$; 2) Compute $SIG(x, R, D)$ for every attribute $x \in C-R$; 3) Select attribute x with maximum $SIG(x, R, D)$ and let $R \leftarrow R \cup \{x\}$; 4) If $\gamma(R, D) = \gamma(C, D)$, then GOTO 5; otherwise, GOTO 2; 5) Return R .

3. DESCRIPTION OF FRILA

Based on the fuzzy-rough set model, FRILA can be described as follows.

1) Calculate the center m_i and fuzzify the continuous attributes. 2) Calculate the fuzzy similarity matrix for every attribute; 3) Calculate fuzzy partition $U/IND(\tilde{R}_\lambda^{(a)})$ given the fuzzy similarity relation $\tilde{R}_\lambda^{(a)}$ with the value set λ ; 4) Calculate the minimal attribute reduction; 5) Calculate the attribute core of the condition attribute with respect to the decision attribute and obtain the minimal reduction of the condition attribute, then delete the redundant objects; 6) For every object, calculate the value core of the condition attribute, and then delete the redundant attribute values and objects; 7) Delete the same objects in decision table and translate the decision rules.

4. A CASE STUDY

In Reference [4], there is a simple relational database system shown as Table 1.

Tab. 1 A relational database

ID	Degree	Experience	Salary	ID	Degree	Experience	Salary
1	Ph. D.	7.2	63,000	12	Master	3.6	41,000
2	Master	2.0	37,000	13	Master	10	68,000
3	Bachelor	7.0	40,000	14	Ph. D.	5.0	57,000
4	Ph. D.	1.2	47,000	15	Bachelor	5.0	36,000
5	Master	7.5	53,000	16	Master	6.2	50,000
6	Bachelor	1.5	26,000	17	Bachelor	0.5	23,000
7	Bachelor	2.3	29,000	18	Master	7.2	55,000
8	Ph. D.	2.0	50,000	19	Master	6.5	51,000
9	Ph. D.	3.8	54,000	20	Ph. D.	7.8	65,000
10	Bachelor	3.5	35,000	21	Master	8.1	64,000
11	Master	3.5	40,000	22	Ph. D.	8.5	70,000

Firstly, two continuous attributes, denoted by E and S respectively, are fuzzified. The Kohonen's feature-map algorithm is used to determine the center m_i and. Secondly, The fuzzy similarity relations of three attributes are constructed. Then given the level values, the fuzzy similarity matrix is transformed to the normal similarity matrix. Thirdly, the fuzzy partition $U/IND(\tilde{R}_\lambda^{\{a\}})$ given the fuzzy similarity relation $\tilde{R}_\lambda^{\{a\}}$ is calculated, where the level values are as follows: $\lambda_D=1.0$, $\lambda_E=0.7$ and $\lambda_S=0.7$. Finally, calculate the attribute reduction and there is no redundant attribute in this case. Then we use FRILA to generate 6 fuzzy rules.

In Reference [4], the ID3-like algorithms, named by FCLS, tend to involve more attributes than FRILA. In other words, the rules induced by FCLS algorithms have redundant attributes and are not more concise than FRILA. The more concise rules and fewer rules lead to a more efficient classification; on the other hand, more rules lead to a higher classification accuracy. These two factors have to be traded off to satisfy application-dependent specifications. The comparison of FCLS and FRILA is shown as Table 2, where the level values $\lambda_D=1.0$ and $\lambda_S=0.7$.

Table 2. Comparison of FRIFA and FCLS

Algorithms	FCLS	FRILA			
		$\lambda_E=0.6$	$\lambda_E=0.7$	$\lambda_E=0.9$	$\lambda_E=0.95$
Number of rules (num)	17	4	6	7	12
Coverage rate (%)	100	84	100	100	100
Accuracy rate (%)	100	64	77	91	100

Compared to the decision tree algorithms, the proposed algorithm has the following advantages:

- 1) The method can deal with both discrete and continuous attributes.
- 2) The induced rules are more concise. Due to the root node attribute exists all the rules, the rules induced by decision tree have redundant attributes.
- 3) The method generates fewer fuzzy rules. It is shown that when both accuracy rate and coverage rate are 100%, FRILA generates 12 rules, whereas FCLS generates 17 rules.

REFERENCES

1. Z. Pawlak, AI and intelligent industrial applications: the rough set perspective. *Cybernetics and Systems: An International Journal*, 31(4), pp. 227-252, 2000
2. Z. Pawlak, Rough sets and fuzzy sets. *Fuzzy Sets and Systems*, 17(1), pp.88-102, 1985
3. D. Dubois and H. Prade, Rough fuzzy sets and fuzzy rough sets[J]. *Int. J. General Systems*, 17, pp.191-208,1990
4. S. M. Chen and M. S. Yeh, Generating fuzzy rules from relational database systems for estimating null values. *Cybernetics and Systems: An International Journal*, 28(8), pp. 695-723, 1997