TELKOMNIKA Indonesian Journal of Electrical Engineering
Vol.12, No.4, April 2014, pp. 2969 ~ 2976
DOI: http://dx.doi.org/10.11591/telkomnika.v12i4.4795

Fuzzy Rough Set Conditional Entropy Attribute Reduction Algorithm

Jinsheng Ren*, Haitao Jia
School of computer science and engineering, University of Electronic Science and Technology of China
Chengdu, Sichuan, China, 611731
*Corresponding author, e-mail: jsren@uestc.edu.cn

Abstract

Modern science is increasingly data-driven and collaborative in nature. Comparing to ordinary data processing, big data processing that is mixed with great missing date must be processed rapidly. The Rough Set was generated to deal with the large data. The QuickReduct is a popular attribute algorithm as the attribute reduction of big database. But less effort has been put on fuzziness and vagueness data. Considering this requirement this paper proposes an improved attribute reduction based on condition entropy of fuzzy rough sets (FRCE) which can deal with the continuous and fuzzy data. This algorithm rewrites the expression of condition entropy by using the information theory. Last this paper takes the UCI database to simulate the efficiency of this algorithm.

Keywords: *fuzzy, rough set, attribute reduction*

1. Introduction

IBM estimates that every day 2.5 quintillion bytes of data are created so much that 90% of the data in the world today has been created in the last two years [1]. Big data is not only becoming more available but also more understandable to computers. But there are more challenges in big data period that great volume unstructured data must be rapidly processed and analyzed. Unstructured data refers to information that either does not have a pre-defined data model or missing data.

Rough-set theory, proposed by Pawlak and Skowron [2], has become a well-established mechanism for uncertainty management in a wide variety of applications related to unstructured data [3-4]. In this framework, an attribute set is viewed as a granular space, which partitions the universe into some knowledge granules or elemental concepts. In most information decision supporting problem, the decision is taken based on known values of information attribute repressed by the vector $A=[a_1,a_2,\cdots,a_n]$. The goal of decision supporting is to determine if the state x belongs to decision d_j or not, $j=1,2\cdots m$.

The one of most importance application of rough set is the attributes reduction for big data, which would find an attribute subset from the original attributes that contains the whole knowledge as the original one. Rough sets approach of attributes reduction can be used as a purely structural method for reducing dimensionality using information contained within the dataset and preserving the meaning of the features.

Nowadays most traditional attribute reduction algorithms were presented to deal with symbolic or real-valued databases. Those algorithms could not manage the fuzziness and vagueness database. One way to solve this problem is to discretize the attribute value before attribute reduction processing [5]. But this method will lead new errors into the system.

For the above attribute reduction issues of fuzzy rough set, Richard Jensen and Qiang Shen proposed QuickReduct algorithm to achieve attribute reduction of fuzzy rough set [6]. This is a algorithm which exploits dependent function to achieve attribute collection reduction. The algorithm is currently the most popular algorithm on attribute reduction of fuzzy rough set.

QuickReduct algorithm is based on a purely algebraic point of view, so it is relatively less intuitive to understand. This paper introduces mutual information theory to reduction algorithm of fuzzy rough set and proposes an improved attribute reduction based on condition

Received September 7, 2013; Revised October 27, 2013; Accepted November 18, 2013

entropy of fuzzy rough sets (FRCE). It is more intuitive and understandable, while its computation has also been reduced, so it advantages attribute reduction of massive data.

2. Theoretical Background

Central to fuzzy rough set rule generation is the concept of fuzzy indiscernibility. Here let U denote a finite and non-empty set called the universe. And A is a non-empty finite set of attributes such that $a: U \to V_a$ for every $a \in A$ .With every subset of attribute $B \subseteq A$ there are an indiscernibility relation:

$$IND(B) = \{(x, y) \in U^2 \mid \forall a \in B, x(a) = y(a)\} \quad (1)$$

Using $[x]_B$ we denote the equivalence class of $IND(B)$ including x. Given an arbitrary set $X \subseteq U$, one can characterize X by a pair of lower and upper approximations. The lower approximation X_R is the greatest definable set contained in X, and the upper approximation X^R is the least definable set containing X. Formally, the lower and upper approximation can be defined as follows [7]:

$$X_R = \{x \mid [x]_R \subset X\} \quad (2)$$

$$X^R = \{x \mid [x]_R \cap X \neq \varnothing\} \quad (3)$$

For $\forall P, Q \subseteq A$ is the indiscernibility relation the position can be defined as:

$$POS_P(Q) = \bigcup_{X \in U/IND(Q)} X_P \quad (4)$$

Fuzzy Rough Set was firstly proposed by Dubois and Prade [8] and then studied in [9]. Suppose U is a nonempty universe (may not be finite) and R is a binary fuzzy relation on U, then the fuzzy approximation operators can be summarized as follows. For every fuzzy set $A \in F(U)$:

$$\overline{R}_T A(x) = \sup_{u \in U} T(R(x,u), A(u)) \quad (5)$$

$$\underline{R}_S A(x) = \inf_{u \in U} S(N(R(x,u)), A(u)) \quad (6)$$

Here a triangular norm, or shortly T-norm, is a function T: $[0,1] \times [0,1] \to [0,1]$ that satisfies the following conditions: monotonicity [if $x < \alpha, y < \beta$,then $T(x, y) \leq T(\alpha, \beta)$], commutativity [$T(x, y) = T(y, x)$] , associativity [$T(T(x, y), z) = T(x, T(y, z))$], and boundary condition [$T(x,1) = x$]. The most popular continuous T-norm include the standard min operator $T_M(x, y) = \min\{x, y\}$ and the bounded intersection $T_L(x, y) = \max\{0, x + y - 1\}$ [10].

A triangular conorm, or shortly T-conorm, is a increasing, commutative, and associative function S: $[0,1] \times [0,1] \to [0,1]$ that satisfies the boundary conditions: $\forall x \in [0,1]$, $S(x,0) = x$. The most well-known continuous T-conorm include the standard max operator $S_M(x, y) = \max\{x, y\}$ and bounded sum $S_L(x, y) = \min\{1, x + y\}$ [11] .

3. QuickReduct Algorithm

The idea of QuickReduct is: first, take an empty set R, then add the condition attributes, which cause the dependent function $\gamma'_R(D)$ under the fuzzy rough sense get its maximum, to the collection, when $\gamma'_R(D)$ gets the maximum value, the algorithm ends. Figure 1 shows the specific steps of the algorithm.

Input: fuzzy decision table DT=(U, $\underset{\sim}{A}= C \bigcup D$,V, f)

Output: a relative reduction

step1 $R \leftarrow \varnothing, \gamma'_{best} \leftarrow 0, \gamma'_{pre} \leftarrow 0$;

step2 $T \leftarrow R$, $\gamma'_{pre} \leftarrow \gamma'_{best}$;

step3 for $\forall c_i \in C\text{-}R$, if $\gamma_{c_i \cup R}(D) > \gamma_T(D)$, then $T = R \bigcup \{c_i\}$, $\gamma_{best} \leftarrow \gamma_T(D)$, until traversing all c_i ;

step4 if $\gamma_{pre} = \gamma_{best}$, then go to Step5, else let $R \leftarrow T$, and go to Step2;

step5 output R, end

Figure 1. Flow of QuickReduct

As we can see from Figure 1, if the number of condition attribute decision table is n, then in the worst case, QuickReduct algorithm's time complexity is O(n(n+1)/2), while the time complexity of all the reduction is O(2n). So the search algorithm based on QuickReduct can not only get an optimal solution, even in the case of do not calculate all the possible relative reduction, but also lower computation complexity and faster search speed, so it has been applied widespread.

4. An Improved Attribute Reduction Based on Condition Entropy of Fuzzy Rough Sets (FRCE)

For newly given fuzzy attributes $\underset{\sim}{P}$ and $\underset{\sim}{Q}$, the divided results of the domain U are $X = U/\underset{\sim}{P} = \{\underset{\sim}{X}_1, \underset{\sim}{X}_2, \cdots, \underset{\sim}{X}_n\}$, $Y = U/\underset{\sim}{Q} = \{\underset{\sim}{Y}_1, \underset{\sim}{Y}_2, \cdots, \underset{\sim}{Y}_m\}$. According to the introduced fuzzy set membership function in fuzzy rough set, for $x_k \in U$, membership belonging to the fuzzy equivalence class $\underset{\sim}{X}_i \in X$ can also be expressed as the probability of belonging to the equivalence class, then the probability $p(\underset{\sim}{X}_i)$ of $\underset{\sim}{X}_i$ can be determined through each object, namely:

$$p(\underset{\sim}{X}_i) = \frac{\sum_{x_k \in U} \mu_{\underset{\sim}{X}_i}(x_k)}{\sum_{\underset{\sim}{X}_l \in X} \sum_{x_k \in U} \mu_{\underset{\sim}{X}_l}(x_k)}, \quad i = 1, 2, \cdots, n \tag{7}$$

Similarly,

$$p(\underset{\sim}{Y}_j) = \frac{\sum_{x_k \in U} \mu_{\underset{\sim}{Y}_j}(x_k)}{\sum_{\underset{\sim}{Y}_l \in X} \sum_{x_k \in U} \mu_{\underset{\sim}{Y}_l}(x_k)}, \quad j = 1, 2, \cdots, n \tag{8}$$

Now we deduct the joint probability expression $p(\underset{\sim}{X}_i \underset{\sim}{Y}_j)$, firstly according to the definition of fuzzy rough set, we get the partition result of domain U using set $\{P, Q\}$, namely:

$$U/A=\{\underset{\sim}{Z}_l=\underset{\sim}{X}_r \cap \underset{\sim}{Y}_t |\ \underset{\sim}{X}_r \in U/\underset{\sim}{P},\ \underset{\sim}{Y}_t \in U/\underset{\sim}{Q}\} \tag{9}$$

There in, $1 \le r \le n,\ 1 \le t \le m$. Using membership, the joint probability can be rewritten as:

$$p(\underset{\sim}{X}_i \underset{\sim}{Y}_j)=\frac{\sum_{x_k \in U} \mu_{\underset{\sim}{X}_i \cap \underset{\sim}{Y}_j}(x_k)}{\sum_{\underset{\sim}{Z}_l \in U/\{P,Q\}} \sum_{x_k \in U} \mu_{\underset{\sim}{Z}_l}(x_k)} \tag{10}$$

Where:

$$\mu_{\underset{\sim}{Z}_l}(x_k)=\mu_{\underset{\sim}{X}_r \cap \underset{\sim}{Y}_t}(x_k)=\min(\mu_{\underset{\sim}{X}_r}(x_k),\ \mu_{\underset{\sim}{Y}_t}(x_k)) \tag{11}$$

Because each object only using two values, 0 and 1, to denote whether it belongs to Q (decision attribute) equivalent class, so according to the formula (10), the joint probability $p(\underset{\sim}{X}_i \underset{\sim}{Y}_j)$ can be rewritten as:

$$p(\underset{\sim}{X}_i \underset{\sim}{Y}_j)=\frac{\sum_{x_k \in U} \mu_{\underset{\sim}{X}_i \cap \underset{\sim}{Y}_j}(x_k)}{\sum_{\underset{\sim}{X}_l \in X} \sum_{x_k \in U} \mu_{\underset{\sim}{X}_l}(x_k)} \tag{12}$$

Then the condition entropy of Q relative to P can be turned into the following form:

$$\begin{aligned} H(Q|P) &= -\sum_{i=1}^{n} p(\underset{\sim}{X}_i) \sum_{j=1}^{m} p(\underset{\sim}{Y}_j|\underset{\sim}{X}_i) \log p(\underset{\sim}{Y}_j|\underset{\sim}{X}_i) \\ &= -\sum_{i=1}^{n} p(\underset{\sim}{X}_i) \sum_{j=1}^{m} \frac{p(\underset{\sim}{X}_i \underset{\sim}{Y}_j)}{p(\underset{\sim}{X}_i)} \log \frac{p(\underset{\sim}{X}_i \underset{\sim}{Y}_j)}{p(\underset{\sim}{X}_i)} \\ &= -\sum_{i=1}^{n} \frac{\sum_{x_k \in U} \mu_{\underset{\sim}{X}_i}(x_k)}{\sum_{\underset{\sim}{X}_l \in X} \sum_{x_k \in U} \mu_{\underset{\sim}{X}_l}(x_k)} \sum_{j=1}^{m} \frac{\sum_{k=1}^{|U|} \mu_{\underset{\sim}{X}_i \cap \underset{\sim}{Y}_j}(x_k)}{\sum_{k=1}^{|U|} \mu_{\underset{\sim}{X}_i}(x_k)} \log \frac{\sum_{k=1}^{|U|} \mu_{\underset{\sim}{X}_i \cap \underset{\sim}{Y}_j}(x_k)}{\sum_{k=1}^{|U|} \mu_{\underset{\sim}{X}_i}(x_k)} \end{aligned} \tag{13}$$

It should be noted that whether different fuzzy equivalence class is empty can be judged from the membership denoting every object belongs to this intersection. if the membership of each object is 0, indicating that this intersection must be empty, because there is no object belonging to it, for $\underset{\sim}{X}_i$、$\underset{\sim}{Y}_j$、$\underset{\sim}{X}_i \cap \underset{\sim}{Y}_j$ in the formula(13), these fuzzy equivalence classes is not empty.

From formula(13), it can be seen that the value of the condition entropy here are entirely determine by each object's membership, so using the above transformation form, information entropy, condition entropy and mutual information can be applied to fuzzy rough set.

The same as information entropy in rough set, if $H(Q|P)$ smaller, then the extent which indicates how much information of attribute Q is determined by attribute P is larger, namely P is more important to Q. So using condition entropy to describe the attribute importance and make reduction is feasible, and the computation is much less than mutual information by using condition entropy to describe the attribute importance.

Similar with QuickReduct, attribute reduction algorithm based on condition entropy can also use condition entropy to determine attribute importance and gradually add attributes to the current reduction collection. Because the condition entropy is monotonically non-increasing function, so the idea of attribute reduction algorithm based on conditional entropy is adding the condition attribute which makes $H(D|R)$ reduce the most to the reduction collection, Figure 2 shows a flow chart of the algorithm.

In Figure 2, we set threshold η, mainly considering some properties, though it may be redundant for reduction set , it contains a certain amount of information, which makes it impossible to be equal to the conditional entropy of the original attribute set. For the amount of information is too small, almost negligible, so there is a need to set a threshold to terminate the algorithm, generally let $\eta=10^{-3}$.

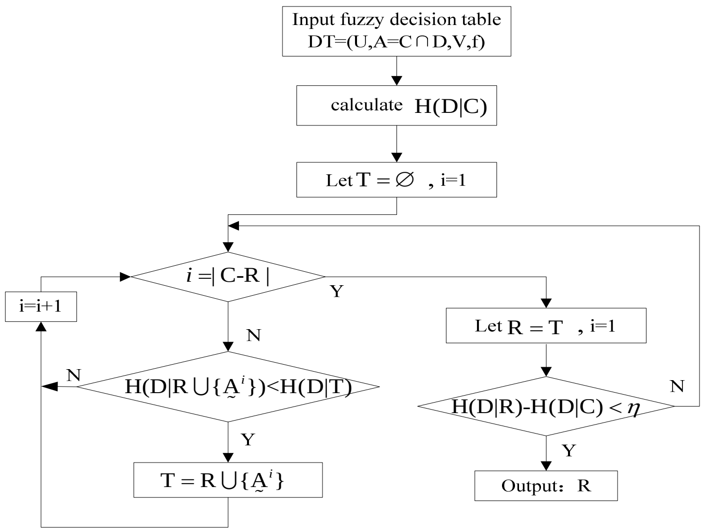

Figure 2. Flow Chart of FRCE Algorithm

5. Simulation

To verify the validity of FRCE algorithm, we choose the instance of frost resistance of concrete to perform the reduction experiment. Let $\eta=10^{-3}$, firstly we calculate the conditional entropy H(D|C), then we get $H(D|C)=0.4618$, Figure 3 shows the entire process of attribute reduction.

As can be seen from Figure 3, when $H(D|\{\underset{\sim}{A}^1,\underset{\sim}{A}^3,\underset{\sim}{A}^4,\underset{\sim}{A}^5\})-H(D|C)=0.0006<\eta$, the algorithm terminates, the reduction set we get is {$\underset{\sim}{A}^1$, $\underset{\sim}{A}^3$, $\underset{\sim}{A}^4$, $\underset{\sim}{A}^5$} and this result is the same with QuickReduct, which indicates the proposed reduction algorithm is effective and can obtain a reduction result of a decision table.

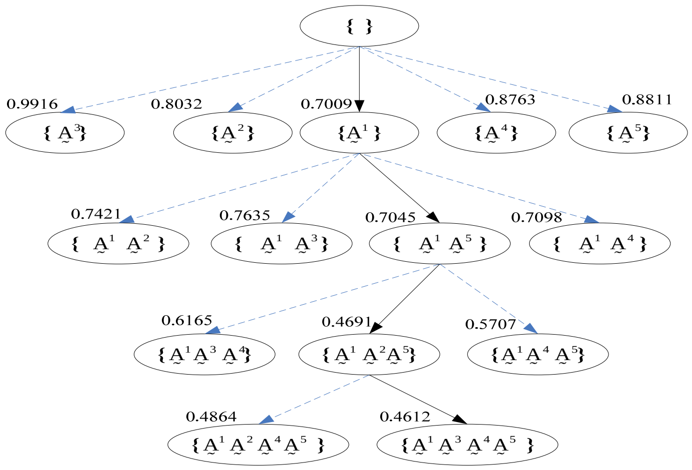

Figure 3. Attribute Reduction of the Instance of Frost Resistance of Concrete Based on FRCE

In order to verify the reduction efficiency of this algorithm, using QuickReduct and the proposed FRCE algorithm respectively perform attribute reduction on five kinds of data set, wine, iris, heart, glass, Ionosphere, from UCI database. Table 1 shows the properties and number information of these five kinds of data set.

Table 1. Information of Five Kinds of Data Set from UCI

	condition attribute	species number	sample number
Iris	4 continuous attributes	3	150
Wine	13 continuous attributes	3	178
Glass	9 continuous attributes	7	214
Heart	13 continuous attributes	4	270
Ionosphere	34 continuous attributes	2	351

First, we apply fuzzy method based on Adaption Fuzzy C-means Clustering to fuzz each data set, then we get reduction result which can be seen from Figure 4.

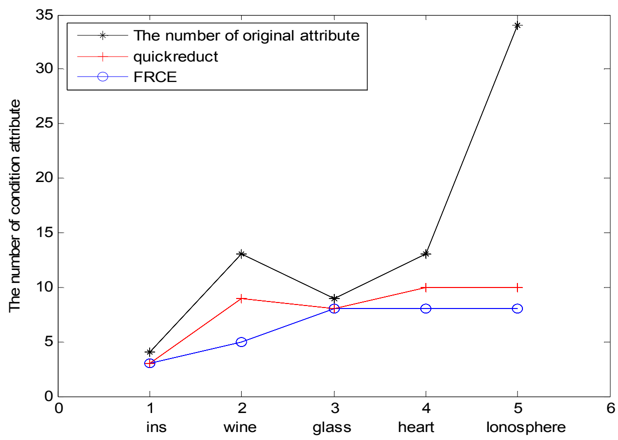

Figure 4. Reduction Result of FRCE and QuickReduct

As can be seen from Figure 4, compared to QuickReduct algorithm, the proposed FRCE algorithm can find a smaller or the same size of reduction set. For iris and glass data sets, the two algorithms get the same reduction result, while for wine, heart and Ionosphere data sets, the FRCE algorithm obtains a smaller reduction set than QuickReduct. Thus, for some high-dimensional data, the proposed reduction algorithm can find a more concise reduction set, compared to QuickReduct algorithm.

Figure 5 shows the time consumption of the two reduction algorithms, it can be seen from Figure 5, for the iris, Wine and glass, which possess relatively small sample number, the time consumption of the two algorithms is almost the same, while for heart and Ionosphere, which possess large sample number, time consumption of FRCE is significantly less than QuickReduct.

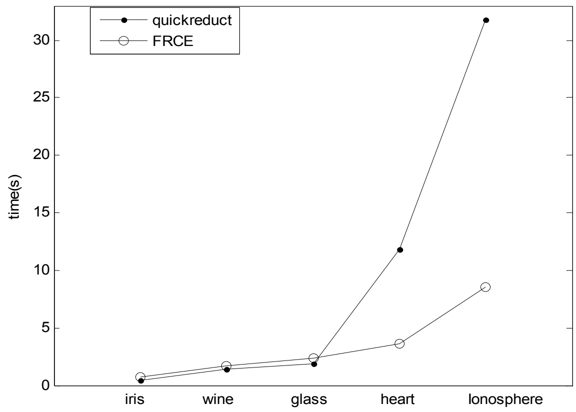

Figure 5. Time Consumption between FRCE and QuickReduct

The above analysis shows that the proposed FRCE algorithm can get an optimal reduction of decision table, especially the data sets or decision tables possess high dimensions and large number of objects, the proposed algorithm compared to QuickReduct can get a more concise reduction set and consume less time.

6. Conclusion

This paper analyzes the QuickReduct algorithm in rough set, which is less intuitive and understandable, and then we modify the expression of information entropy and conditional entropy in rough set, on this basis, we propose an improved attribute reduction based on condition entropy of fuzzy rough sets algorithm(FRCE). In order to prove the validity of the algorithm, we respectively perform attribute reduction on QuickReduct four kinds of UCI data sets using QuickReduct and FRCE algorithm. It can be seen from the experiment results, compared to QuickReduct, the algorithm can obtain a smaller or equal reduction set, it also consumes less time, so it is clear that the proposed FRCE algorithm is effective.

References

[1] Improving Decision Making in the World of Big Data http://www.forbes.com/sites/christopherfrank/ 2012/03/25/ improving-decision-making-in-the-world-of-big-dataM.
[2] Z Pawlak. *Rough Sets: Theoretical Aspects of Reasoning About Data System Theory*. Knowledge Engineering and Problem Solving, Kluwer Academic Publishers, Dordrecht, Netherlands. Kluwer. 1991; .9.
[3] Yi Zhang. Rough Set and DEA of Strategic Alliance Stable Decision-making Model. *TELKOMNIKA*. 2013; 11(12): 7295-7301.
[4] Yusu Xiong. A Clustering Algorithm Based on Rough Set and Genetic Algorithm. *TELKOMNIKA*. 2013; 11(10): 5782-5788.
[5] HS Nguyen, A Skowron. *Quantization of real value attributes.* Proceeding of second Joint Annual

Conference Information Science. Wrightsville Beach, North Carolina.1995: 34–37.
[6] Richard Jensen, Qiang Shen. *Fuzzy rough attribute reduction with application to web categorization*. Fuzzy sets and System. 2004; 141: 469-485.
[7] M Quafafou. α-RST: A generalization of rough set theory. *Information Science*. 2000; 124: 301–316.
[8] D Dubois, H Prade. *Putting rough sets and fuzzy sets together in Intelligent Decision Support*. Handbook of Applications and Advances of the Rough Sets Theory, R Slowinski, Ed. Norwell, MA: Kluwer. 1992: 203–232.
[9] Suyun Zhao, Eric CC Tsang, Degang Chen. The Model of Fuzzy Variable Precision Rough Sets. *IEEE Transactions on Fuzzy Systems.* 2009; 17(2): 451-467.
[10] Yan Li, Simon CK Shiu, Sankar K Pal. Combining Feature Reduction and Case Selection in Building CBR Classifiers. *IEEE Transactions on Knowledge and Data Engineering*. 2006: 18(3): 415-429.
[11] Q Shen, R Jensen. *Selecting Informative Features with Fuzzy-rough Sets and Its Application for Complex Systems Monitoring*. Pattern Recognition. 2004; 37: 1351–1363.