International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.7, No.3 (2014), pp.193-202
http://dx.doi.org/10.14257/ijsip.2014.7.3.16

Reduction Algorithm for Decision Table Combined with Grey Relational Analysis

Jin Dai[1], Xin Liu[1] and Feng Hu[2]

[1]*School of Software Engineering, Chongqing University of Posts and Telecommunications, Chongqing, 400065, P. R. China*
[2]*College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, 400065, P. R. China*
daijin@cqupt.edu.cn

Abstract

The fuzziness and randomness of decision table affect hugely on the performance of knowledge acquisition in rough set. In order to reduce their influence, a novel reduction algorithm based on grey relational analysis is proposed. In the algorithm, every value of decision table is converted to the same domain. Moreover, on the basis of grey relational analysis, the grey relational matrix for the converted decision table is constructed to describe the equivalence relations between samples of decision table. Finally, the samples with the same similar level are adopted as the coarser granularity. The experiments fully show that the reduction decision table achieved almost the same recognition rate with less than one tenth of the former conditions. It fully shows the effectiveness of the algorithm.

Keywords: *rough set; decision table; grey relational analysis; equivalence relation*

1. Introduction

Decision Table is an essential tool for knowledge acquisition in rough set [1], which is widely used in different fields such as machine learning, data mining and so on [2, 3].

The fuzziness and randomness of decision Table affect hugely on the performance of knowledge acquisition in rough set. In order to reduce their influence, many scholars have done a great deal of research, but most of these are mainly focused on the reduction of decision Table attributes [4-7]. In the paper [8], the redundant samples is extracted by radix sorting in the typical rough set attribute values, which is based on the equivalence relations of strict attribute values matching. While in real world, the attribute values of decision table mostly are continuous values. It's difficult for equivalence relations of typical rough set to divide the samples. The paper [9] simplified the decision table by the Boolean attribute discernibility matrix, but it can only deal with the situation when the value of attributes is integer type. The paper [10] designed a heuristic function to calculate the discernibility object pair on the basis of the discernibility matrix, which is generated in the condition, attributes set of reduction decision Table. When the dataset is large-scale, the reduction method of binary discernibility matrix attribute has some shortages. The paper vertically divided the dataset for distributed calculating to solve the problem [11]. It obtained a reduction decision Table from another way, but the complexity is higher.

Grey theory [12, 13] is an important acquisition model of uncertainty knowledge. It has been successfully applied in forecasting and decision-making by lots of important projects. And, its prediction accuracy is fairly high [12-15]. In order to reduce the influence of the fuzziness and randomness of decision Table, a novel reduction algorithm based on grey

ISSN: 2005-4254 IJSIP

relational analysis is proposed. In the algorithm, every value of decision Table is converted to the same domain. Moreover, on the basis of grey relational analysis, the grey relational matrix for the converted decision Table is constructed to describe the equivalence relations between samples of decision Table. Finally, the samples with the same similar level are adopted as the coarser granularity. The experiments fully show that the reduction decision table achieved almost the same recognition rate with less than one tenth of the former conditions.

In this paper, the 2nd section introduces the basic concepts of rough set and grey relational analysis. The 3rd section proposes a reduction algorithm for decision Table based on grey relational degree. The experiment results and analysis are given in the 4th section. At last, a conclusion of the algorithm is presented.

2. Fundamental Concepts

2.1. Reduction Decision Table in Rough Set

Definition 1 (decision Table [16])

Given $S =< U, A, V, f >$, S is an information system, $U = \{x_1, x_2, ..., x_n\}$ is a domain, A is a set of attributes, V is a set of attribute values, F is mapping to $U \times A \to V$. If $A = C \cup D, C \cap D = \varphi$, C is called condition attribute set, D is called decision attribute set, and the whole information system S is called decision table.

Definition 2 (positive region [16])

Given $S =< U, C \cup D, V, f >$, set $U / D = \{D_1, D_2, ..., D_k\}$ to represent the division of decision attribute set D to universe) U, $U / P = \{P_1, P_2, ..., P_m\}$ represents the division of decision attribute set $P (P \subseteq C)$ to domain U, $POS_P(D) = \bigcup_{D_i \in U/D} P\_(D_i)$ is called positive region of P on D.

Definition 3 (the attribute reduction based on positive region [16])

In the decision Table $S =< U, C \cup D, V, f >$, if $\forall B \subseteq C$, $POS_B(D) = POS_C(D)$, and B is independent from D, then we call B is the attribute reduction of C that relate to D.

Theorem 1: Given the decision Table $S =< U, C \cup D, V, f >$,

$$POS_C(D) = \bigcup_{X \in U/C \wedge \forall x, y \in X \Rightarrow f(x,D) = f(y,D)} X . \quad (1)$$

Proof: It can be achieved by the definition.

The theorem describes that the positive regions of C on D are consisting of basic blocks, which have the only values of decision attributes. Therefore, any basic blocks (any equivalence class of U / C is the basic block) that isn't an only attribute in decision attributes will not exist in the positive region. The following definition can be achieved by the theorem.

Definition 4 (Reduction Decision Table)

In the decision Table $S =< U, C \cup D, V, f >$, note $U / C = \{[u'_1]_C, [u'_2]_C, ..., [u'_m]_C\}$, $U' = \{u'_1, u'_2, ..., u'_m\}$, suppose $POS_C(D) = [u'_{i_1}]_C \cup [u'_{i_2}]_C \cup ... \cup [u'_{i_t}]_C$ based on Theorem 1, there are $\forall u'_{i_s} \in U'$

and $|[u'_{i_s}]_C / D| = 1 (s = 1, 2, ..., t)$; note $U'_{POS} = \{u'_{i_1}, u'_{i_2}, ..., u'_{i_t}\}$, then $U'_{NEG} = U' - U'_{POS}$. $S' = < U', C \cup D, V, f >$ is the reduction decision table.

2.2. Grey Relational Analysis

Grey relational analysis is a method can be used for quantitatively describing and comparing the dynamic development process of a system [12]. By using grey relational degree, the similarity between the reference sequence and the comparable sequence's geometric shape will be achieved. The higher grey relational degree is, the closer the development and rate between sequences are. And their relationship will be closer too. Grey relational analysis method has made up the disadvantages of the statistical method for system analyzing. It is applicable to samples with different sizes as well as samples without regularity. Meanwhile, it has lowered computational complexity and is much more convenient, and the quantitative conclusions will be always consistent with qualitative analysis results.

Grey relational degree refers to the closeness level between two grey systems or two factors of one grey system with the variation of time and objects. In the development process of grey system, if the variation of two factors is consistent, their grey relational degree will be higher, and vice versa.

Definition 5 (grey Absolute Relational Degree [12])

Proposing X_i and X_j have the same length , and the length is 1 time interval sequence, $X_i^0 = (x_i^0(1), x_i^0(2), ..., x_i^0(n))$ and $X_j^0 = (x_j^0(1), x_j^0(2), ..., x_j^0(n))$ are the starting-point annihilating images[12] of X_i and X_j, then:

$$\varepsilon_{ij} = \frac{1 + |s_i| + |s_j|}{1 + |s_i| + |s_j| + |s_i - s_j|} \tag{2}$$

where $|s_i - s_j| = |\sum_{k=2}^{n-1} (x_i^0(k) - x_j^0(k)) + \frac{1}{2}(x_i^0(n) - x_j^0(n))|$, $|s| = |\sum_{k=2}^{n-1} x^0(k) + \frac{1}{2} x^0(n)|$.

ε_{ij} Is where denoted as the grey absolute relational degree between X_i and X_j .

On the basis of grey relational degree, a grey relational matrix, which is used for grey correlation clustering analysis, can be obtained.

$$A = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1m} \\ & \gamma_{22} & & \gamma_{2m} \\ & & \ddots & \vdots \\ & & & \gamma_{mm} \end{bmatrix} \tag{3}$$

In the above matrix, γ_{ij} is the grey relational degree between X_i and X_j . In addition, there is the relative relational degree in the grey relational analysis, whose construction is similar to the absolute relational degree. The only difference between them is the initial value image should be done before calculating starting-point value images.

3. Reduction Algorithm for Decision Table based on Grey Relational Analysis

The core problem of decision table reduction is which equivalence relation is adopted to characterize the relations of samples. Currently, there are two methods, one is equivalence relation based on strict attribute value matching, and the other one is fuzzy similarity relation based on membership function. While the former is unable to handle the problem when taking continuous attribute value, and the latter's determination band of membership function is subjective. So if only relying on the priority knowledge given by the expert, or considering less about the random distribution of taking attribute values, it will lead to a low recognition rate with reduction.

By introducing the grey relational degree of samples, the similarity relations can be measured between samples. On this basis, the samples of the same decision can be clustered by dynamic clustering method. Each center is the most typical representation of polymerization samples. It has the better information representation capability. The core idea of the algorithm is to extract the typical sample from the samples with the same similarity level, and constructs a new decision table.

In the selection of clustering method, *k*-means algorithm [17], which is an efficient dynamic clustering algorithm, is adopted. However, *k*-means algorithm has a problem that it needs manually specify the number of clusters. Furthermore, it's easy to fall into optimal solution defects. Based on the analysis, this paper proposes a decision table clustering algorithm with a novel *k*-means method by the step-by-step advancing strategy.

Algorithm 1: Decision Table Sample Dynamic Clustering Algorithm

Input: Decision table S_c (decision attribute value is D_k)
Output: Polymerization matrix D'
Algorithm steps:

(1) Calculating grey similarity matrix $D_c = \{T_i, T_j, sim, Clusterid\}$ of S_c // *sim* is the grey relational degree between the samples T_i, T_j $(i < j)$, *Clusterid* is the clustering number, and its initial value is 0.

(2) Extracting all the unduplicated *sim* to construct a category vector $C' = \{sim, Clusterid\}$ with ascending order.

(3) Calculating the threshold $e, e = \overline{s}_n$ // e is the end control condition of cluster completed

(4) Initial Category $K = 1$, $v = 0$, //v is a loop control variable

(5) Do

1) Constructing a center category table TC : C' is divided averagely to $K+1$ composition adding TC to the right point of fist K composition as the initial category of C' under the situation K ; At the same time, the *Clusterid* of each element in C' is set to 0.

2) Set a temporary control variable $e_1 = 0$

3) When $e_1 \neq v$, performing the follow loop: // when the clustering is stable, the standard deviation of each class will converge to a stable value.

a) $e_1 \neq v$

b) Calculating the distance between each value of C' and each category of TC, and integrating it into the minimum distance category.

c) Correcting the center distance of each category in TC based on the weighted average.

d) Calculating the standard deviation S_i of each category in T, set $v = \min(S_i)$.

4) $K = K + 1$

$While(v > e)$ //when $v \le e$, the polymerization degree of each is relatively good, the clustering is end.

(6) According to the *sim* in C', updating the *Clusterid* of D_c.

(7) D_c is processed according to the order T_i ascending and *Clusterid* descending as follows:

1) Set $t_k \in T_i$, $c_{i\max}$ as the largest category number when $T_i = t_k$. Then the most similar sample set is $T = \{t_k\} \cup \{t_m \mid t_m \in T_j \wedge T_i = t_k \wedge Clusterid = c_{i\max}\}$ of t_k in D_c. // according to the clustering algorithm, if the *Clusterid* is larger, the similarity degree is higher.

2) $D' = D' \cup \{(T, c_{i\max})\}$ // only keeping the most similar sample set.

(8) Return D'.

With the clustering processing of algorithm 1, and when the decision table has been processed by the grey similarity relation, the sample equivalence clusters will be obtained. The same *Clusterid* is a similarity sample. On this basis, this paper proposes a reduction algorithm for decision table based on grey relational analysis.

Algorithm2: The Reduction Algorithm for Decision Table based on Grey Relational Analysis.

Input: Decision Table S

Output: Reduction decision Table S'

Algorithm steps:

(1) $S' = \phi$; $D' = \phi$ // D' is the polymerization matrix.

(2) Convert S to the same domain with probability and statistics method

(3) Performing the following operations for each D_k of D (D is the decision attribute set of S):

(1) Using algorithm 1, a clustering matrix $D'_k = \{T, Clusterid\}$ of the sample in decision D_k can be obtained

2) $D' = D' \cup D'_k$

(4) According to D', S is divided into several equivalence clusters, extract each center sample from every cluster to S';

(5) Return S'.

4. Experiment and Analysis

In order to evaluate the performance of algorithm 2 (abbreviated as GR-DTRA), some datasets are selected from UCI datasets [17] for experiments.

The experiments are mainly focusing on the knowledge acquisition capability of the different decision Tables in rough set. As a comparison, the decision Table reduction algorithm based on fuzzy similarity relation (abbreviated as FSR-DTRA) [9, 18], which is the most widely used at present, is adopted.

Firstly, two algorithms are used to reduce the decision-making Table. The results are shown in Table 1;

Table 1. The Average Reduction Performance Comparison between the Algorithms

Dataset	GR-DTRA		FSR-DTRA ($\varepsilon = 0.1$)	
	Samples	Reduction Rate	Samples	Reduction Rate
Iris	19	87.33%	13	91.33%
Wine	22	87.64%	18	89.89%
Glass	28	86.92%	24	88.79%
Average		87.30%		90.00%

Moreover, the above reduction decision tables are divided into three parts, two of them as the training set, the remaining part as test set. Using the Nguyen greedy algorithm [19] to discretize the training sets, and then process the attribute reduction. The attribute reduction is a very important research field in rough set, which can greatly improve the clarity of the potential knowledge in decision-making table by removing redundant attributes. In order to observe the influence of the different decision-making Tables to attribute reduction, the attribute reduction discernibility based on discernibility matrix algorithm (abbreviated as DMAR), the attribute reduction based on information entropy algorithm (abbreviated as IEAR) and the attribute reduction based on genetic algorithm (abbreviated as GAAR) are adopted.

After attribute reduction treatment, the value reduction is required, which is the rule extraction procedure in Rough Set. Similarly, in order to observe the influence of the different decision-making Tables to the value reduction, the value reduction based on general algorithm (abbreviated as GVR), the value reduction based on decision matrix algorithm (abbreviated as DMVR) and the value reduction based on heuristic algorithm (abbreviated as HVR) are adopted to obtain the rule sets.

Finally, apply the rule sets to identify the test sets.

The experimental hardware conditions are shown as follows: CPU: Intel Core2 2.0GHz, RAM: 2GB, Operating system: Windows XP, development tools for the VC++6.0. The results are shown in Table 2, Table 3 and Table 4.

Table 2. The Test Result based on GR-DTRA

Algorithms				Data Sets		
				Iris	Wine	Glass
GR-DTRA	DMAR	GVR	Attributes	3	10	8
			Rules	10	13	15
			Correction Rate (%)	97.2	92.5	82.1
		DMVR	Attributes	3	11	9
			Rules	9	12	14
			Correction Rate (%)	96.5	90.8	81.5
		HVR	Attributes	3	11	9
			Rules	10	13	18
			Correction Rate (%)	97	92.1	83.2
	IEAR	GVR	Attributes	4	12	7
			Rules	10	14	17
			Correction Rate (%)	96.2	90.2	81.7
		DMVR	Attributes	3	12	9
			Rules	10	13	16
			Correction Rate (%)	97.1	92.5	82.5
		HVR	Attributes	3	11	8
			Rules	11	12	17
			Correction Rate (%)	96.8	92.3	82.4
	GAAR	GVR	Attributes	3	12	9
			Rules	9	11	16

			Correction Rate (%)	97.1	91	80.8
		DMVR	Attributes	3	11	10
			Rules	10	13	17
			Correction Rate (%)	97	90.8	81.5
		HVR	Attributes	3	11	9
			Rules	11	13	18
			Correction Rate (%)	96.8	90.8	82.9

Table 3. The Test Result based on FSR-DTRA

Algorithms				Data Sets		
				Iris	Wine	Glass
FSR-DTRA ($\varepsilon = 0.1$)	DMAR	GVR	Attributes	3	9	10
			Rules	9	11	12
			Correction Rate (%)	82.1	81.5	72
		DMVR	Attributes	3	10	10
			Rules	9	12	14
			Correction Rate (%)	82.5	81.8	71.5
		HVR	Attributes	3	10	9
			Rules	10	10	12
			Correction Rate (%)	80.8	80.1	81.2
	IEAR	GVR	Attributes	2	9	8
			Rules	10	11	12
			Correction Rate (%)	80.2	79.9	80.7
		DMVR	Attributes	3	9	9
			Rules	9	10	11
			Correction Rate (%)	83.1	80.5	79.5
		HVR	Attributes	3	10	8
			Rules	10	10	9
			Correction Rate (%)	84.1	83.1	80.2
	GAAR	GVR	Attributes	3	11	10
			Rules	9	9	10
			Correction Rate (%)	81.1	79.5	78.5
		DMVR	Attributes	3	11	10
			Rules	9	10	12
			Correction Rate (%)	81.2	79.8	80.5
		HVR	Attributes	3	10	9
			Rules	9	9	11
			Correction Rate (%)	80.2	78.8	77.9

Table 4. The Average Performance between the Algorithms

Dataset	GR-DTRA		FSR-DTRA	
	Reduction Rate	Correction Rate	Reduction Rate	Correction Rate
Iris	87.33%	96.85%	91.33%	81.70%
Wine	87.64%	91.44%	89.89%	80.56%
Glass	86.92%	82.06%	88.79%	78.21%
Average	***87.30%***	***90.12%***	***90.00%***	***80.16%***

From Table 4, the average reduction rate of GR-DTRA is slightly lower than FSR-DTRA and the correct rate is apparently higher than FSR-DTRA (Figure 1).

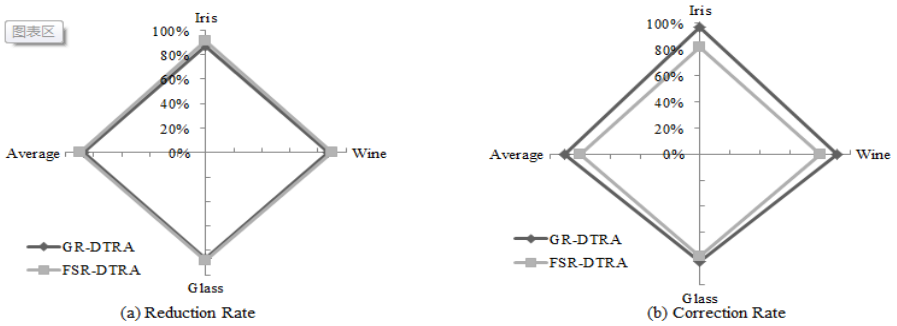

Figure 1. Performance Comparison between the Algorithms

In addition, FSR-DTRA requires experience knowledge to determine the appropriate threshold, which increases the difficulty of large-scale adoption. Combined the time complexity and the correction rate to recognize the samples, it is fully proved that GR-DTRA is an efficient reduction algorithm. The core reason is that the reduction algorithm based on the grey relational analysis is better in keeping the characteristics of the original decision table, and the extracted coarse-grained samples have the better knowledge representation capability.

6. Conclusions

Although rough set theory has been increasingly mature, the practical application is not widely used. An important reason is that the decision table reduction algorithm based on rough set theory is not efficient when dataset is the large-scale. With the impact of the objective world and awareness, it's hard to get a decision Table, which is completely accurate or containing no redundant information. It's necessary to realize the reduction based on samples before using reduction Table to realize knowledge acquisition. The reduction algorithm for decision Table based on grey relational analysis has better considered the distribution characteristics. Therefore, the achieved coarse granularity samples have better knowledge representation capability. And, it effectively improves the further application of rough set.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (NSFC) under Grant no. 61309014; the Natural Science Foundation Project of CQ CSTC under Grant no. cstc2013jcyjA40009, no. cstc2013jcyjA40063; the Natural Science Foundation Project of CQUPT under Grant no. A2012-96.

References

[1] Z. PAWLAK, "Rough Set", International Journal of Computer and Information Sciences, vol. 11, **(1982)**, pp. 341-356.
[2] Z. PAWLAK, J. GRZYMALA-BUSSE, R. SLOWINSKI and W. ZIARKO, "Rough sets", Communications of the ACM, vol. 38, no. 11, **(1995)**, pp. 89-95.
[3] Z. PAWLAK, J. MYCIELSKI, G. ROZENBERG and A. SALOMAA, "Vagueness-a rough set view", In, eds. Structures in Logic and Computer Science：A selection of Essays in Honor of A Ehrenfeucht, Springer-Verlag, **(1997)**, pp. 106-117. Berlin.
[4] Y. QIAN, J. LIANG, Y. YAO, *et al.*, "MGRS: A multi-granulation rough set", Information Sciences, **(2010)**, vol. 180, no. 6, pp. 949-970.

[5] D. Zhang, J. Qiu and X. Li, "Attribute Reduction Based on Equivalence Classes with Multiple Decision Values in Rough Set Proceedings of the International Conference on Information Engineering and Applications (IEA)", Springer London, vol. 2013, **(2012)**, pp. 505-512.
[6] S. K. Mandal, F. T. S. Chan and M. K. Tiwari, "Leak detection of pipeline: An integrated approach of rough set theory and artificial bee colony trained SVM", Expert Systems with Applications, vol. 39, no. 3, **(2012)**, pp. 3071-3080.
[7] W. Xu, Y. Li and X. Liao, "Approaches to attribute reductions based on rough set and matrix computation in inconsistent ordered information systems", Knowledge-Based Systems, vol. 27, **(2012)**, pp. 78-91.
[8] Z. Y. Xu, Z. P. Liu, B. R. Yang and W. Song, "A quick attribute reduction algorithm with complexity of max(O(|C||U|),O(|C||U/C|))", Chinese Journal of Computers, vol. 29, no. 3, **(2006)**, pp. 391-399.
[9] H. Ge, L. S. Li and C. J. Yang, "Algorithm for Computing Core Based on Attribute Boolean Discernibility Matrix", Journal of Chinese Computer Systems, vol. 33, no. 2, **(2012)**, pp. 275-279.
[10] Z. D. Han, Z. L. Wang and J. Gao, "Efficient Attribute Reduction Algorithm Based on the Idea of Discernibility Object Pair Set", Journal of Chinese Computer Systems, vol. 32, no. 2, **(2011)**, pp. 299-304.
[11] C. J. Yang, H. Ge and L. S. Li, "Attribute reduction of vertically partitioned binary discernibility matrix", Control and Decision, vol. 28, no. 4, **(2013)**, pp. 563-568.
[12] J. L. Deng, "The Foundation of Grey System", Wuhan: Huazhong University of Science and Technology Press, **(2002)**, pp. 1-46.
[13] S. F. Liu, M. L. Hu and Y. J. Yang, "Progress of Grey System Models", Transactions of Nanjing University of Aeronautics and Astronautics, vol. 29, no. 2, **(2012)**, pp. 103-111.
[14] J. Song, Y. G. Dang and Z. M. Hua, "Study on group decision-making method based on grey cluster model", Control and Decision, vol. 25, no. 10, **(2010)**.
[15] P. Li and S. F. Liu, "Interval-valued Intuitionistic Fuzzy Numbers Decision-making Method Based on Grey Relational analysis and D-S Theory of Evidence", Acta Automatica Sinica, vol. 37, no. 8, **(2011)**, pp. 993-998.
[16] G. Y. Wang, "Rough Set Theory and Knowledge Acquisition", Xian: Jiaoda Press, **(2001)**.
[17] S. M. Yuan and X. Q. Cheng, "Clustering Method for Mining Quantitative Association Rules", Chinese Journal of Computers, (in Chinese), vol. 23, no. 8, **(2002)**, pp. 866-871.
[18] J. Dai and F. Hu, "Research and application of text classification based on incomplete information system", Journal of Chongqing University of Posts and Telecommunications (Natural Science), vol. 18, no. 3, **(2006)**, pp. 397-401.
[19] S. H. Nguyen and A. Skowron, "Quantization of real value attributes-rough set and Boolean reasoning approach", Proc. of the 2nd Joint Conf. on Information Science, **(1995)**, pp. 34-37.

