Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 7 (2010) 473–476

Procedia Engineering

www.elsevier.com/locate/procedia

2010 Symposium on Security Detection and Information Processing

Improved Binary Discernibility Matrix Attribute Reduction Algorithm in Customer Relationship Management

Ju Li [a*], Xing Wang [b], Xiaowen Fan [a]

[a] *School of Computer Science and Engineering, Chang Shu Institute of Technology, Changshu, Jiangsu, 215500, China*
[b] *School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, 210044, China*

Abstract

This paper proposed an improved binary discernibility matrix, and attribute reduction algorithm based on the improved binary discernibility matrix. Sort and simplification links were added in the algorithm, so as to achieve the purpose of reducing complexity of the algorithm. Examples denoted this method reduced the amount of calculation ,and it is a practical approach.

Keywords: rough set; Binary Dicernibility Matrix ,Attribution Reduction, CRM

1. Introduction

By the Polish mathematician Pawlak in 1982, the rough set theory is proposed to deal with imprecise and incomplete data ,and it is an effective way. [1~2] Attribution reduction algorithm based on binary dicernibility matrix [3~8] is one of important attribute reductions Based on rough set,Many attribute reduction and the expansion of the work is on this basis.But,this method only applies to the compatibility decision table,Inconsistent decision table will get the error result of reduction.To overcome this inapplicability of distinction matrix definition to inconsistent decision tables,References presented a new concept of discernibility matrix of decision table,And discussed its nature.But,Of the inconsistent decision tables, in accordance with the references the definition of matrix will have many redundant information,affect calculation of reduction,And reduction of these algorithms is not high quality.To improve the efficiency reduction, This paper presents a modified binary discernibility matrix attribute reduction algorithm, combined with the actual situation, the establishment of a client classification rules mining model, using the model accurately classified customer level, customer relationship management decisions..

2. Customer relationship management

Customer relationship management is good customer service, improve customer service quality, to focus on customer satisfaction and maintain customer loyalty. The core idea was that customers are the most important assets, enterprises in order to more effectively acquire, develop and retain the most important assets one by one customer,

* Corresponding author. Tel.:+86-0512-52166136.
E-mail address: liju284532@163.com.

1877-7058
doi:10.1016/j.proeng.2010.11.078

you need to know what the needs of major customers. And customer contact in the process, provide for individual differences in customer demand is consistent with which the service plans. In customer relations, business always want to establish strong relationships with our customers the best and most efficient to put this relationship into profits, that is to retain old customers and develop new customers and target customers in the highest profit margins, this is the CRM to focus on the problem. To achieve this goal, companies need to understand customer behavior as possible, but this understanding can not be obtained through direct contact with customers, because businesses can not talk with customers one by one, and the information needed by enterprises often can not provide a single customer. Companies can do is to collect customer information as possible, using various analytical methods, through the disorder, the inner surface of the information dug up the knowledge and rules, which currently very popular data mining technology research. After a lot of information in the dug out, companies can use this information to these laws or the design model, the behavior did not occur to make the results of the forecast, the company integrated management decision-making, provide the basis for marketing.

3. The basic concept of rough set

To introduce improved binary discernibility matrix attribute reduction algorithm, we introduce some concepts of the following：

Definition 1: For a given decision-making systemS=(U, C∪D, V, f),reduction of condition attribute set C is a non-empty subset of C ——P. It meets：

① $\forall a \in P$, Can not be omitted by D

② POSp(D) = POSc(D)

Claimed : P is a reduction of C,the set of all reduction of C denoted RED(C).

By the reduction of the definition, every decision-making system reduction may have several, but reduction is equivalent,that is say they have the same classification ability. The reduction of nuclear is the most important attribute set, which includes all of the reduction. Definition 2：Identification of matrix is Proposed by Skowron Professor.System of decision-making table S=(U, R, V, f),R=PYD is the set of attributes,subset P={ai | i=1, 2, …, m}and D={d} Are called condition attributes set and decision attribute set,U={X1, X2, …, Xn} Is the domain,ak(xj) is the sample xj,s value on the property ak.CD(i,j)denotes dicernibility matrix the i row j column element, defined as the identification matrix CDis：

$$C_D(i,j) = \begin{cases} \{a_k \mid a_k \in p \Lambda a_k(x_i) \neq a_k(x_j)\}, d(x_i) \neq d(x_j); \\ 0, d(x_i) = d(x_j); \end{cases}$$

Among i,j=1, 2, …, n.

Recognition based on improved binary matrix attribute reduction algorithm based on discernibility matrix and make up logical combination of attribute reduction algorithm can only consistent decision table attribute reduction of the shortcomings of the two types of consistent and inconsistent decision tables to arrive at a correct result of attribute reduction.

4. attribute reduction algorithm based on improved binary dicernibility matrix

analysis of algorithms：First of all, scan the line of binary discernibility matrix,Obtain nuclear properties. If a row is only one element 1, then with the elements of a column where the corresponding property of the nuclear properties that must be included in the attribute reduction. Followed by scanning the binary discernibility matrix of the column, find the maximum number of elements in a column. Here elements of each column to measure the properties of a number of importance, a number of the more, the corresponding decision table the column attribute for the contribution of the overall classification ability of the larger, more important the attribute. When the two a number of the same, to the column where the value of a property line of a number of add, the smallest of the property taken and the most important attribute, because the number of lines can also be a measure of the importance of property When the line of a smaller number that can distinguish between two instances of the line corresponds to the properties of the less so that the more important of these properties. Then, from the binary discernibility matrix M by deleting the selected elements in one column and the corresponding line of the column,

the remainder constitute a recognition of a new binary matrix. Repeat the above steps until M is empty. All the columns selected in turn constitutes an attribute reduction.

5. analyses

In this example, the first through the data extraction and discretization, to obtain the data in Table A, where the domain U = {1,2 ··· 8}, condition attribute set C = {a, b, c, d, e }, the decision attribute D = {type} = {customers high-value customers, the general value of the customer, low-value customers} = {0,1,2}, a = {Nature} = {foreign investment enterprises, state-owned, private} = {0 , 1,2}, b = {year turnover} = {high, medium, low} = {0,1,2}, c = {product quality requirements of high, medium, low} = {0,1 , 2}, d = {corporate-owned processing and manufacturing industry, product Agencies} = {0,1}, e = {large-scale enterprises, medium, small} = {0,1,2}

Table 1 a decision-making system

U	a	b	c	d	e	D
1	0	2	2	0	2	2
2	1	2	1	1	1	2
3	2	1	0	1	0	0
4	2	0	0	1	2	1
5	2	0	0	1	0	0
6	1	0	1	1	1	1
7	0	0	2	0	0	0
8	2	1	1	0	0	0
9	1	0	2	0	2	2
10	0	2	2	0	0	1

U/ind(C)= {{1}，{2}，{3}，{4}，{5},{6}，{7},{8},{9}，{10} } ,U/{a}={{1，7，10}，{2，6，9},{3,4,5,8}}, U/{b}={{1，2，10}，{4,5, 6,7,9},{3, 8}}, U/{c}={{1，7,9，10}，{2, 6,8},{3, 4,5}}, U/{d}={{1，7，8,9,10}，{2,3,4,5, 6}}, U/{e}={{1，4，9}，{2, 6 },{3, 5,7,8,10}}, U/D={{1，2，9}，{3，5，7，8},{4，6，10}}，U/{a，b}={{1，10}，{2}，{3,8}，{4,5}{7}，{6,9} }，U/{a，b}={{1，10}，{2}，{3,8}，{4,5}{7}，{6,9} }，U/{b，e}={{1}，{2}，{3,8}，{4,9}，{5,7}，{6}，10}，U/{C-a}={{1}，{2}，{3}，{4}，{5},{6}，{7},{8},{9}，{10} }，U/{C-b}={{1}，{2，6}，{3，5}，{4}，{5},{7，10},{8},{9}}，U/{C-c}={{1}，{2}，{3}，{4}，{5},{6}，{7},{8},{9}，{10} }，U/{C-d}={{1}，{2}，{3}，{4}，{5},{6}，{7},{8},{9}，{10} }，U/{C-e}={{1，10}，{2}，{3}，{4，5},{6}，{7},{8},{9}}

POS(C)(D)= {{1}，{2}，{3}，{4}，{5},{6}，{7},{8},{9}，{10} }，POS{C-a }(D)=POS(C)(D)，POS{C-b}(D)= {{1}，{4}，{5},{8},{9}} ≠POS(C)(D)，POS{C-c}(D)= POS(C)(D)，POS{C-d}(D)= POS(C)(D)，POS{C-e}(D)= {2}，{3}，{6}，{7},{8},{9}}≠POS(C)(D)，so，b and e are not omitted. So the core of C, D is {b, e}.

6. Conclusion

Rough set theory with its unique advantages to win more and more researchers are concerned about, and access to a wide range in various fields of application. In this paper, based on improved binary discernibility matrix attribute reduction algorithm to increase the sorting and simplification of the links, so the algorithm can reduce the complexity of subsequent steps. The method for extracting decision rules in the customer database, customer classification, customer relationship management to provide decision support through the establishment of the model can accurately find the high-value customers, and specific description of their behavior in order to extract the behavior of characteristics. These behavioral characteristics, there are differences could be developed. Other marketing strategies to improve their customer satisfaction. It is proved that the method is feasible.

References

[1] Paw1alc ZW. Rough sets[J]. International Journal of Information and Computer Science, 1982，11(5): 341-356.
[2] Pawlak Z. Rough Sets Theoretical Aspects of Reasoning about Data[M]. Kluwer Academic Publishers, 1991.

[3] Skowron A, Rauszer C. The discernibility matrices and functions in information systems[C] Intelligent Decision Support Handbook of Applications and Advances of the Rough Sets Theory. Dordrecht : Kluwer Academic Publishers , 1992; 331-362.
[4] Hu Xiaohua, Cercone N. Learning in relational data bases: a rough set approach[J]. *Computational Intel ligence* 1995; **11(2)**: 323-337.
[5] Wang Guoyin. Decision table calculation of nuclear properties [J]. *Computers* 2003; **26(5)**: 611-615.
[6] Yan Deloitte. Incompatible information systems paradigm and the difference between the specifications matrix [J]. *Computer Engineering and Applications* 2004; **36:** 45-46.
[7] Yun-Dong, Feng Yanbin, et al.. Discernible matrix and logic operations based on the attribute reduction algorithm [J]. Pattern Recognition and Artificial Intelligence 2004; **17(1)**: 119-123.
[8] Yi. J elonek attribute reduction algorithm is an improvement [J]. *Electronics* 2000; **28(12)**: 81-82.