ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0018

Study on the Improved Reduction Algorithm of Telecom CRM Knowledge Base Property

Weilai Hao

School of Lingnan Normal University, Commercial College, Zhanjiang 524000, China

Abstract

Telecom enterprises have a high rate of disconnection. Through the data analysis in GRM system, the behavior dynamics of users can be understood.According to the defects of the difference matrix attribute reduction algorithm in the existing analysis method, in order to solve the problem of solving the reduction attribute of the difference matrix too complex,the algorithm is improved,by grouping the conditional properties,extract representative records to generate the difference matrix, simplifying the order of the difference matrix and the complexity of the reduction attribute.Therefore, the temporal complexity and spatial complexity of the algorithm are optimized, which saves the computational time and spatial complexity.Examples show that the algorithm can effectively reduce the properties, can obtain ideal results, and the improved algorithm is simple and efficient.

Keywords

CRM; Difference Matrix; Reduction; Algorithm.

1. Introduction

Data mining techniques and methods as a scientific method of data analysis has been applied to various industry sectors;this method can be found in vast amounts of customer data, useful data, and later adopt a different strategy to achieve certain social and economic benefits.Most telecom companies already have a lot of detailed data existing in CRM systems,but the data implied a lot of valuable information has not been fully tapped use. This paper studies using differential matrix from a lot of, not complete, there is noise in the data extraction of implicit, people do not know in advance of useful information and knowledge, so as to achieve the purpose of attribute reduction.

2. Attribute Reduction Algorithm

So-called knowledge reduction in knowledge ability is classified under the condition of invariable, delete not related or not important knowledge. In large amounts of data, if want to obtain valuable information must undertake attribute reduction, delete the interference attributes.

2.1 The Basic Concepts of Attribute Reduction

A knowledge expression system can be represented as a quaterple S= (U, A, V, f) [1], among:

U: System of non-empty finite set of objects, called discourse domain.

A: Properties of non-empty finite set, usually expressed as $C \cup D$, C is condition attributes set, D is decision attribute set.

$V = \cup Va$ range for the set of object attributes, Va for the attribute $a \in A$ in the range.

f: $U \times A \rightarrow V$ is an information function, it gives each object a message for each property value, have randomly $a \in A$, $x \in U$, $f(x, a) \in Va$.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0018

Definition 1.Let P, R be the two families of equivalence relations defined on the theoretic domain U, and R⊆P, if:

$$ind(P) = ind(R) \tag{1}$$

R is an independent, said R is a reduction of P.[2].

Definition 2. P is defined in the U a equivalence relation, the collection of all the necessary relationships in P, known as the nuclear of P, denoted as core (P).

Definition 3. Suppose |U|=n, discernibility matrix Ms is an n × n matrix in decision table S, in which any element mij represents the i line and j row element set. Among [3]:

$$\begin{cases} \{\alpha \in C: f(x_i,\alpha) \neq f(x_j,\alpha)\} \\ f(x_i,\alpha) \neq f(x_j,\alpha) and\ min\{d(x_i), d(x_j)\} = 1 \\ \phi \quad other \end{cases} \tag{2}$$

Clearly, Ms is a symmetric matrix diagonal is empty, so our analysis, just consider the matrix of the upper half or lower half. In the attribute reduction is the most concerned about how to obtain the best set of attributes, but in order to seek this property set how to got the nuclear as the began, Ms in a discernibility matrix element mij is the condition attributes C, constitute a single attribute, then the element mij is the nuclear core, core = core ∪ mij.

Definition 4. Set X⊆U, R is an equivalence relation, said RX={x|x ∈ U} and RX={x|x ∈ U,and[x]R⊆X},is set X of R lower approximation; and $\bar{R}X = \{x|x \in U, and[x]_R \cap X \neq \phi\}$for the set X of R upper approximation. X, R-boundary region for the $\mathrm{BN_R(X)} = \bar{R}X - \mathrm{RX}$,the lower approximation and upper approximation of set difference. X is called R-domain, $\mathrm{NEGR(X)} = \mathrm{U} - \overline{\mathrm{R}}\mathrm{X}$ known as R of x negative domain [4].

2.2 Differential Matrix Generation

The existing discernibility matrix method in attribute screening exist serious redundancy , thus causing the time complexity of the algorithm. For example in the Discernibility matrix, if mij={a, b} element item, at the same time, there is mnp={a, b, c} element item, then (a, b) element key to determine decisions, (a, b, c) element item becomes redundant, and in the screening of the property to delete the result will be spending time algorithm, resulting in duplication of attributes, so in order to exclude statistical properties of redundancy, should be excluded in the discernibility matrix element mnp.

Conventional discernibility matrix method can produce an n × n matrix. So the space complexity of the algorithm is high. And when the data and information classification was not considered decision attributes, just according to the condition attributes. For example: records with the same condition attribute Xj and Xi, in the discernibility matrix will have the same elements, so the only choose one of the record to generate differential matrix, using this method the new domain for $U = \{x_1', x_2', \ldots\ldots x_n'\}$, one for the same condition attribute set of records. It would be difficult to get the decision attribute differences caused by the condition attributes; therefore the attribute set accuracy is not high. In this case, in the condition attributes are classified according to the decision attribute, before classification.

2.3 Attribute Reduction Algorithm

In the generated difference matrix, the single attribute element item constitutes the kernel of the attribute, how to seek the simplified attribute if there is no single attribute element item,in the fourth references use the importance of the attributes (SFG(a, C0, D)) used to filter, SFG(a, C0, D) larger

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0018

the attribute more important. But this method is more complex, and spending more time. But in the discernibility matrix elements, attributes represent the item caused by different factors, decision attribute value, namely the important attribute of attribute, so if the ergodic [5] again discernibility matrix, statistical matrix of each attribute, it is relatively easy to extract attributes successively from Fre (r).

Specific algorithm is as follows:

Input: one decision table S = (U, C ∪ D, V, f).

Output: Decision table and an optimal core nuclear attribute set of R.

Step1 For the U / R of the decision table, the U was classified first $U = \{x_1', x_2', \ldots\ldots x_n'\}$.

Step2 Select records from U generate discernibility matrix M.

Step3 On discernibility matrix redundant processing.

Step4 Search the discernibility matrix elements of a single attribute r, and add to the nuclear and attributes, core=core∪r， R=R∪r,remove the element of the difference matrix M containing this property r. If no such elements, the attribute r with the highest frequency is selected to merge into the property set R.

Step5 Treatment of the difference matrix, delete the elements of attribute r in discernibility matrix M, and delete the attributes of the elements contain attribute R.

Step6 if matrix M=∅, go to Step7, else choose single attribute elements, and add to attributes set R, if do not have this kind of elements, then select the maximum attribute of frequencies set into attributes R and removed from the discernibility matrix. Go to Step5.

Step7 Out put core and R.

R as the best attribute sets for us going to build BP neural network in the next.

3. Example analysis

3.1 Example

Table 1. Decision Table

	a_1	a_2	a_3	a_4	D
x_1	1	1	1	1	0
x_2	2	2	2	1	1
x_3	1	1	1	1	0
x_4	2	2	2	1	1
x_5	3	1	2	1	0
x_6	1	2	3	2	2
x_7	2	3	1	2	3
x_8	3	1	2	1	1
x_9	1	2	3	2	2
x_{10}	3	1	2	1	1
x_{11}	2	3	1	2	3
x_{12}	4	3	4	2	1
x_{13}	1	2	3	2	3
x_{14}	4	3	4	2	2

ISSN: 2414-1895 DOI: 10.6919/ICJE.202204_8(4).0018

Through the classification and get the sets U={{x1,x3},{x5,x8,x10},{x2,x4},{x7,x11},{x6,x9,x13},{x12,x14}},then select the condition attributes of different records from U1 and U2, and got a new decision table.

Table 2. Simplify Decision Tables

		a	b	c	d
x_1		2	2	0	1
x_2		1	2	0	0
x_3		1	2	0	1
x_4		0	0	0	0
x_5		1	0	1	0
x_6		2	0	1	1

The number of properties in the difference matrix obtained from Table 2 is successively Fre(a1)=9,Fre(a2)=11,Fre(a3)=11,Fre(a4)=8 the reduced difference matrix is:

Table 3. Simplified Discernibility Matrixes

	x_1	x_3	x_6
x_2	a	ϕ	abc
x_4	ab	ab	ac
x_5	abc	bc	a

Then the attribute reduction[6] is performed for the difference matrix.

Table 4. Attribute reduction

	ab	ac	bc	results
Step 1			bc	a
Step 2				ab

The reduction of attribute reduction sets (a, b). Attribute (a) as the core attribute selected from table 4, then selected attribute b from table 4, because Fre (b) larger than Fre (c).

3.2 Algorithm Analysis

For decision table is classified into the time complexity is $O(C_n^{[n/2]})$, n is decision table records. Simplified discernibility matrix space complexity is O(m1×m2), m1, m2 for simplifying the records of decision table a and b. Time complexity is O(m1×m2) for Fre(r) , For simplified discernibility matrix time complexity is O(m1×m2), Space complexity is O(m1×m2), So the algorithm greatly saves time and space.

4. Algorithmic Complexity Analysis

By using this method and the original methods attribute set, structural neural network to predict the customer data applications. Select 500 data as test data, the selection of 50 data attribute reduction,

and using the data of 50 to neural network is trained, then the 500 data on actual prediction, and the forecast results and practical results, comparing the accuracy as below:

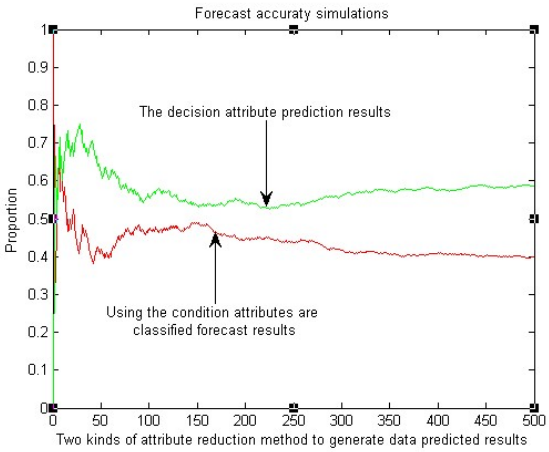

Figure 1. Accurate distribution

In the above figure, green curve is the accuracy of this paper method, red curve is the accuracy of original method. From the figure 1 we can seen that, this method accuracy high than original method. Using the decision attribute decision table is divided into two kinds of data, one kind is customized business data, and another kind is not customized business data. Using two kinds of data generated discernibility matrix can be effectively extracted key attributes. Thus causes two methods of different accuracy.

5. Conclusion

A From the above example shows, this method has some improvement, based on the condition attributes categorize grouping, extracting representative to generate differential matrix, simplify the discernibility matrix rank number and the complexity of the attribute reduction. Therefore in the time and space complexity of the algorithm is doing some optimization. Save the time and space complexity of the algorithm. Example shows that the algorithm can effectively get attributes sets, achieve ideal result. And the improved algorithm is simple and effective.

Acknowledgments

Fund project:Zhanjiang City Science and Technology Plan Project,project number:2021B01203.

References

[1] X.Fan,H.M.Chen:Stepwise Optimized Feature Selection Algorithm Based on Discernibility Matrix and mRMR (Computer Science,2020) p.87-95. (In Chinese).

[2] Majid Abdolrazzagh-Nezhad,Homa Radgohar,Seyede Najme Salimian: Enhanced cultural algorithm to solve multi-objective attribute reduction based on rough set theory (Mathematics and Computers in Simulation, 2020),p.170-172.

[3] Y.Yao,X,Y.Zhang:Class-specific attribute reducts in rough set theory.(Information Sciences,2017) p.418-419.(In Chinese).

[4] L.Feng,D.Q.Miao:Witold Pedrycz. Granular multi-label feature selection based on mutual information. (Pattern Recognition,2017) p.67-70.

[5] Y.H.Qian,X.Y. Liang,Q. Wang: Local rough set: A solution to rough data analysis in big data (International Journal of Approximate Reasoning,2018) p.97-100.

[6] C.J.YANG,H.Ge,Z.S.Wang:Overview of attribute reduction based on rough set(Computer Application Research, 2012) p.16-20.(In Chinese).