Analysis of Risk Decision of E-Commerce Project Based on Data Mining Of Rough Sets

LI WANQING, MA LIHUA, MENG WENQING, DU FENGQIU
College of Economics and Management
Hebei University of Engineering
Guangming South Street 199,Handan,056038
CHINA

Abstract: - In the paper, the objectivity and feasibility of Rough Sets applying is analyzed in the procedure of decision support system. In order to solve the problem of risk decision of E-Commerce project, a new method of data mining based on Rough Sets is proposed by analyzing project including uncertain factors. Firstly, the set of factors is established including condition attribute and decision attribute. Secondly, experts qualitatively describe risk factors and establish a decision database, called decision table. Thirdly, the attribute reduction algorithm based on Rough Sets is used to eliminate the redundant risk factor and its value of decision table. Fourthly, the minimum decision rules are abstracted based on data mining technology. Finally, the procedure of risk decision based on data mining of Rough Sets is analyzed in a case study. The method is more convenient and practical compared with the traditional one.

Key-Words: - Risk decision, data mining, minimum decision rule, Rough Sets, attribute reduction, E-Commerce, project

1 Introduction

E-Commerce is a high-risk project because of many uncertain causes, including complex technology, specialized equipment, special environment personnel disposition and so on [1]. How to control and decrease the risk is a difficulty problem [2]. The procedure of decision-making is making right decision through right information and right way. So information plays an important role in the course of decision-making procedure. Information of proper quantities is necessary to make decision, that is to say, information of quantitative and qualitative influences directly the result of decision [3]. As far as the deciders are concerned, they hold the massive information in an E-Commerce project. Then it is necessary to use the algorithm to discover valuable knowledge and make right decision. Data mining is a kind of method to disposal massive data and to discover some implied rules that are useful to make a decision [4]. Since 1990s, many kinds of data mining are proposed by scientists, such as clustering, artificial neural network, decision tree, machine learning, Rough Sets and so on [5,6]. Rough Sets theory applies to data mining supplying the mathematics tool for dealing with uncertain knowledge [7]. And it also supplies the method for acquiring and refining knowledge. Based on the basic conception of Rough Sets theory, the paper discusses the method in terms of Rough Sets theory for analyzing information, extracting useful characteristics and producing the least-making regulation [7]. Data mining-based project for the exploitation of potential E-Business clients is discussed, and the advantages and procedure of the technology and the detailed implementing plan are introduced [8]. Liu Qing and Zeng Huanglin studied the characteristic and application of Rough Sets theory [9, 10]. Yang Shanlin discussed the procedure of data analyzing based on data mining of Rough Sets, and proposed the application of this method to decision support system [11].

In the risk decision process of an E-Commerce project, it is necessary to use the algorithm to discover valuable knowledge and make a right decision. The rest of the paper is organized as follows. In section 2, data mining based on Rough Sets are introduced, including some concepts of Rough Sets theory, reduction algorithm based on Rough Sets, and the computational process. In section 3, first, the objectivity and feasibility of Rough Sets applying is analyzed in the procedure of a decision support system. And a procedure of data mining is analyzed in an E-Commerce project. In this course, first, the set of factors is established, including condition factor and decision factor. Secondly, experts qualitatively describe risk factors and establish a decision knowledge database, called decision table. Thirdly, the algorithm based on Rough Sets is used to eliminate the redundant risk

factor and its value from the decision table. Fourthly, the minimum decision rules are created based on data mining technology. Finaly,the meaning of minimum decision rules are analyzed in a case study.

2 Data Mining Based on Rough Sets

Data mining is a kind of method to process massive data and to find out some implied rules that are useful to make decisions [4]. Rough Sets theory proposed by Z.Plawlak in 1980s is one of such techniques. It is a novel mathematic method to study uncertain data, deficiency of data, incomplete data, or even inconsistent data [12, 13].

2.1 Some concepts of Rough Sets theory

2.1.1 Indiscernibility relation

In Rough Sets, the relation is close between knowledge and classification, and knowledge is defined as an ability to classify. Suppose $K=(U,R)$ is a knowledge base, where U is a nonempty finite set called domain, R is the equivalence relations of U , U/R is all the equivalence classes of R . $[X]_R$ is an equivalence class of R including element $x \in U$. If $P \subseteq R$ and $P \neq \Phi$, then all intersection of equivalence relations are an equivalence relation in P , called indiscernibility relation about P , as in $ind(P)$, $[x]_{ind\{R\}} = \bigcap_{R \in P} [x]_R$, $P \subseteq R$.

2.1.2 Upper approximation, lower approximation and boundary of Rough Sets

In Rough Sets, accuracy concepts are signified by two accuracy sets including upper approximation and lower approximation. In a knowledge base $K=(U,R)$,for each subset $x \in U$ and an equivalence relation $R \in ind(K)$, suppose two subsets are as follows.

$$R_{-}(X) = \{x | [x]_R \subset X, x \in U\}$$

$$R^{-}(X) = \{x | [x]_R \cap X \neq \phi, x \in U\}$$

Then $R_{-}(X)$ and $R^{-}(X)$ are the upper and lower approximation sets of X about R . Suppose boundary domain of X about R is $bn_R(X) = R^{-}(X) - R_{-}(X)$. And suppose $posR(X) = R_{-}(X)$ is the positive region of X about R , $negR(X) = U - R_{-}(X)$ is the negative region of X about R .

2.1.3 Information system and decision table

In Rough Sets, the information system takes the form of relation table. Knowledge system with condition attribute and decision attribute is a decision table. A decision table is a kind of critical knowledge system. Suppose $S=(U,A,V,f)$ is a knowledge system, where $S=(x_1,x_2,\cdots,x_n)$ is a finite set of object, $A=(a_1,a_2,\cdots,a_n)$ is a finite set of attribute, here in V is field composed of attribute A , $f: U \times A \to V$ is an information function, each element of U with a unique value that is a about V , $A = C \bigcup D$, C is the condition attribute set, D is the decision attribute set.

2.2 Reduction algorithm based on Rough Sets

Simplified table is the result of simplifying condition attribute, and the classification function remains to be. And simplified decision table contains less complicated condition attributes. We know a simplified condition is necessary in making decisions. The algorithm has 2 steps, i.e. attribute reduction and attribute value reduction as follows.

2.2.1 Attribute reduction

For an information system $S=(U,A,V,f)$, $A = C \bigcup D$, $B \subseteq C$, if $\gamma_C(D) = \gamma_B(D)$ and B is individual in relation to D , then B is the simplification of attribute D in relation to C , as in $RED_D(C)$. The calculation is shown as follows.

Input: C , D , and U

Output: attribute reduction C in relation to D

Step 1 $s \leftarrow 0$, $RED(s) \leftarrow \phi$;

Step 2 $i \leftarrow 1$;

Step 3 $j \leftarrow 1, m \leftarrow 0$;

Step 4 For subset $C(i,j)$ of C, covering j subset of element i

(1) $t \leftarrow 0$

(2) If $(RED(t) \neq \phi) \wedge (RED(t) \subseteq C(i,j))$, $m \leftarrow m+1$, if $m = C^i_{|C|}$, turn to Step 7, else turn to Step 5

(3) If $t \geq s$ turn to (5)

(4) $t \leftarrow t+1$, turn to (2)

(5) If $\gamma_C(D) = \gamma_{C(i,j)}(D)$ turn to (6), else turn to Step 5

(6) $s \leftarrow s+1, RED(s) \leftarrow C(i,j)$

Step 5 If $j \geq C^i_{|C|}$ turn to Step 6, else $j \leftarrow j+1$,

turn to Step4

Step 6 If $i \geq |C|$ ends, else $i \leftarrow i+1$, turn to Step 3

Step 7 Output $RED(s)$

2.2.2 Attribute value reduction

For in information system $S = (U, RED_D(C) \bigcup D, V, f)$, the calculation is shown as follows.

Input: $S = (U, RED_D(C) \bigcup D, V, f)$,

$RED(s) = \{C_1, C_2, \cdots, C_n\}$

Output: core value table S' of S

Step1 $S' = (U, C \bigcup D, V' \leftarrow Null, f')$

Step2 For each condition attribute C_k (repeat as follows)

For each $x_i \in U$ and $C'_k(x_i) = Null$ (repeat as follows)

If

$$\exists x_i \left((x_j \neq x_i) \wedge \forall C_l \left(C_l \neq C_k \wedge C_l(x_j) = C_l(x_i) \right) \right) \wedge \left(D(x_j) \neq D(x_i) \right)$$

Then $C'_k(x_j) = C_k(x_j), C'_k(x_i) = C_k(x_j)$

Step3 Output S'

3 Case Study

3.1 The analysis of risk decision procedure based on data mining of Rough Sets

Risk control is a procedure from identification to settlement during the course of decision. And the procedure of data management contains 3 stages, collection, process and application. From Fig.1, a procedure of decision support system includes 5 sections, i.e. data collection, rough database, decision knowledge base, decision rules base, and decision interface. Rough Sets are essential in two steps, both of which are used from original database to knowledge base and from knowledge base to decision rules base. The analysis of decision procedure based on data mining of Rough Sets is as follows Fig.1.

3.2 Condition attribute sets and decision attribute sets

The risk factor sets are called condition attribute sets, which reflect E-Commerce project risks, including technical feasibility *a*, amount of investment *b*, capital-raising ability of project *c*, and market expectation *d*. Decision attribute sets include enterprise scale *e* and risk process methods *f*. We can get information to make decision from a decision table, called risk decision table. The table is composed of rows and arrays to represent attributes and objects. We study an E-Commerce project to abstract Table 1 as follows.

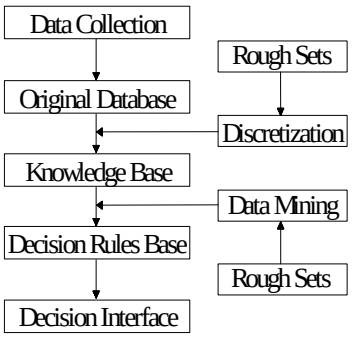

Fig.1. The analysis of risk decision procedure based on data mining of Rough Sets

3.3 Dispersing attribute and establishing knowledge base

Dispersing condition and decision attributes are used to establish the knowledge base. Firstly, using condition attribute sets from above, we disperse the results of expert evaluation as follows. Technical feasibility is divided into 3 grades {*1,2,3*} to represent {*low,average,high*}. Similarly, the amount of investment is also divided into 4 grades {*1,2,3,4*} to represent {*lower,low,high,higher*}. Capital-raising ability of project is divided into 2 grades {*1,2*} to represent {*bad,good*}. Market expectation is divided into 3 grades {*1,2,3*} to represent {*bad,average,good*}. Secondly, we use decision attribute sets above, and the results of expert evaluation are represented as follows. Enterprise scale is divided into 2 grades {*1,2*} to represent {*small,big*}. Risk process methods are divided into 3 kinds {*1,2,3*}, including risk bearing, risk sharing and risk avoiding. Risk decision table of E-Commerce project is as follow Table 1.

3.4 Attribute reduction

From Table 1, the redundant attributes are eliminated and core attributes are preserved. The minimum decision rules are composed of core attributes without redundant attributes.The new table is called reduced attribute table as follow Table 2.

Table 1. Risk decision table of E-Commerce project

U \ A	a	b	c	d	e	f
1	2	3	1	2	2	3
2	2	2	1	2	2	3
3	2	2	1	1	2	3
4	1	1	1	1	1	3
5	1	2	1	2	1	3
6	2	2	1	3	2	2
7	2	3	1	3	2	2
8	3	3	1	3	2	2
9	3	3	2	3	2	1
10	3	4	2	3	2	1
11	3	4	2	2	2	1
12	3	3	2	2	2	1
13	3	2	2	2	2	1

Table 2. Reduced attribute table

U \ A	a	c	d	e	f
1	2	1	2	2	3
2	2	1	1	2	3
3	1	1	1	1	3
4	1	1	2	1	3
5	2	1	3	2	2
6	3	1	3	2	2
7	3	2	3	2	1
8	3	2	2	2	1

3.5 Attribute value reduction

From Table 2, we get attribute value reduced table as follows.

Table 3. Reducted attribute value table

U \ A	a	c	d	e	f
1	2	-	2	2	3
2	2	-	1	2	3
3	1	-	-	1	3
4	1	-	-	1	3
5	-	-	3	2	2
6	-	1	-	2	2
7	-	2	-	2	1
8	-	-	-	2	1

3.6 Interpretation analysis

From Table 3, we know decision rules as follows.

$a_2 d_2 \to (2,3)$

or $a_2 d_1 \to (2,3), a_1 \to (1,3), d_3 \leftarrow (2,2)$

or $c_1 \to (2,2), c_2 \to (2,1)$.

From above, we know 4 decision rules as follows.

(1) $a_2 d_2 \vee a_2 d_1 \to (2,3)$.

(2) $a_1 \to (1,3)$.

(3) $d_3 \vee c_1 \to (2,2)$.

(4) $c_2 \to (2,1)$.

From 4 decision rules above, we get strategy of risk decision as follows.

(1) When the technical feasibility is “average”, and market expectation is “average” or “bad”, the strategy of big enterprise is risk avoiding. That is to say, that will give up the project or modify it.

(2) When the technical feasibility is “low”, the strategy of small enterprise is risk avoiding. That is to say, that will give up project or modify it.

(3) When market expectation is “good” or capital-raising ability of project is “bad”, the strategy of big enterprise is risk sharing. That is to say, that will share risks and profits with cooperators.

(4) When capital-raising ability of project is “good”, the strategy of big enterprise is risk bearing. That is to say, that will solely bear the risk in full.

4 Conclusions

The risk decision of E-Commerce project is sort of multiple attribute decision-making processes. It is certain that we have to deal with massive data. So data mining technology is necessary, for it can abstract implicit regularity from massive data. In this paper, the reduction algorithm based on Rough Sets is proposed as a practical data mining technology. A new decision method is proposed in order to solve the risk decision problem of an E-Commerce project. The factor set is established including condition attribute and decision attribute. Then experts qualitatively describe risk factors and create a decision table, and the attribute reduction algorithm based on Rough Sets is used to eliminate the redundant risk factor and its value of decision table. Finally, the minimum decision rules are created based on data mining results. And the rules with explain meanings are abstracted in the case study. We can make proper decision from the rules to improve precision and explanatory ability in practice.

References:

[1] Electronic Commerce Association of China. Project Management of International

E-Commerce [M]. Posts &Telecom Press, 2004. (in Chinese)
[2] Bennet P.Lientz,Kathryn P.Rea,Shen T. E-Commerce Project Performance Management[M].Beijing: Electronic Industry Press, 2003. (in Chinese)
[3] Zhou Sanduo, Chen Chuanming.Management[M].Higher Education Press,2000. (in Chinese)
[4] Lv Anmin, Lin Zongjian, Li Chenming. Approaching of Data Mining and Knowledge Discovery [J]. Survey Science, 2000,12(4): 36-39. (in Chinese)
[5] Vasant Dhar, Roger Stein. Intelligent Dicision Support Methods. The Science of Knowledge Work. Printice Hall, 1997:171-180.
[6] L Polkowski, A skowron. Rough Sets:A Perspective. Rough Sets in Knowledge Difcovery(1,2), Physica-Verlag, Heidelberg,1998.
[7] Wu Bing, Li Weixiang, Zhao Lindu. Least Decision-Making Method Based on Rough Set Theory [J]. Information Technique, 2002, (4): 4-5. (in Chinese)
[8] Zheng Xiuli, Wang Lening,Chen Zhongzhu. Data Mining-based Project for Electronic Business Client Potential Exploitation [J]. Computer Engineering and Application, 2002, (5): 194-195. (in Chinese)
[9] Liu Qing. Rough Sets and Rough Inference [M]. Beijing: Science Press, 2001. (in Chinese)
[10] Zeng Huanglin. Rough Sets and Application [M].Chongqing University Press,1996. (in Chinese)
[11] Yang Shanlin. Intelligent Decision Method and Intelligent Decision Support System [M]. Science Press, 2005. (in Chinese)
[12] Z. Pawlak. Rough Sets-Theoretical Aspects of Reasoning about Data [M]. Klystron Academic Publisher, 1994.
[13] Z. Pawlak. Rough Sets.International Journal of Computer and Information Sceinces, 1982, (11): 341-356.