Research of Default Rules Mining Model Based on Reduced Lattice

Xinyuan Lu [1], Huili Zhang [2], and Jinlong Zhang [3]

[1] Department of Information Management, Huazhong Normal University, Wuhan, 430079, China, luxy@mail.ccnu.edu.cn

[2] College of Humanities, Xi'an University of Archtecture and Technology, Xi'an, 710055, China, xy-hx@163.com

[3] College of Management, Huazhong University of Science & Technology, Wuhan 430074, China, jlzhang@mail.hust.edu.cn

Abstract. In order to solve the decision question with incomplete information and uncertainty of risk factors during the risk decision, the concept of reduced lattice is introduced into project risk management in this paper, then the default rule mining model based on the combination of rough set and reduced lattice is constructed, and created a series of subsystems from the known decision system at different reduced levels, then form a reduced lattice, and then deduce its own rule set at each reduced lattice. At last, an example is introduced to demonstrate the method and model above detailed.

1 Introduction

During the reasoning of rough set, It always supposes that a attribute of each object has only one value matched it, so the equivalent relationship of the value region can be make certain uniquely, which means it is very convenient to dispose the data. But in many occasion, the information are incomplete or lacked, and the descriptions of the objects are not reasonable, so the decision of incomplete information in the complicated system is inevitable. Assume that a decision support system is expressed by $S = (U, C \cup D, V, f)$ using the theory of rough set. Given a class $E_i \in U / Ind(C)$, and if all the objects in E_i are reflex to the same decision class $X_i \in U / Ind(D)$, then all the rules are certain, which means its reliability is 100%. But in facts, the objects can not reflex to the same decision class completely because of the missing information, and it means the rules are not 100% reliable, which is result in the coming into being of default rules.

Some scholars have make correlative research in this fields, such as: Skowron.A put forwarded the method of Boundary Region Thinning (BRT) to mine the default rules[1], Mollestad presented the method of selecting the default rules from incomplete information using theory of lattice and searching tree[2].Wang Yaying concluded the method of reduced lattice, and classify the knowledge into several kinds, then mined the default rules in different kind[3,4,5].Under this background mentioned above, this paper construct the default model mining model based on rough set and reduced lattice. The main process of this method are: given the reliability value μ of the rules, and construct the lattice according to different reduced levels using the rough set theory, then find out all the rule which matched the value of μ for each node in different layer. Finally, using those rules to make decision and reasoning, and find the result according to top-priority of each rule.

2 Construction of reduced lattice

The concept of lattice was presented by Wille.R in 1982, which is a two-dimensional hiberarchy, and it is also a useful mathematics tool to analyze the data. The lattice includes two meanings: inherent and extension characters of objects. In essential, the lattice describes the relationship between the object and it attributes, and it also indicates the relationship between generalities and particularity. The lattice is only influenced by the data itself, but not influenced by the order of the data or the attributes [6,7].

Suppose N expresses the number of decision attributes in a *DSS*, the initial lattice can be constructed through the method of projection. The top layer has only one node, all the node in the same layer has the same attributes (equal to the number of layer), with the decrease of layer, the number of node is decreased, so there is no node in the layer *0*. Obviously, the number of node in the layer k is C_N^k, which are deleted a attribute from the layer k+1, and the relationship between the neighbor nodes are successive. But in the initial lattice, each node is the subsystem of the initial system, the rules of each node is repeated because of the attributes are redundancy, which means that the computing process is very perplexing, so it is necessary to simplify the initial lattice.

Define1: Use the concept of reduced in Rough set, suppose there is a lattice of the node, if the attributes set of the node is positive to the set of decision attribute, then we can define the node as reduced node, otherwise, the node is un-reduced. If all the nodes in the lattice are reduced, then the lattice is defined as *Reduced Lattice.*

Define2: In a reduced lattice, if a node in layer k is un-reduced, then its reduced form can be found out (the reduced maybe more than one form). Generally, assume the number of attributes is $m(m\leqslant k)$ in a reduced form, if node in the layer from m+1 to k whose attribute set includes this reduced, then the node can be deleted, which means the node is superfluous, and this does not influence the obtaining of default rules.

The main course of construction of reduced lattice include: Firstly, the attribute sets can be ascertained from the initial data, then choose the value range of the attribute sets, and disposal those data preliminarily such as discrete, so the *DSS* of

$S=(U,C\cup D,V,f)$ can be gained. Secondly, construct the initial lattice according to the number of attributes, then the reduced lattice can also be calculated from the initial lattice. The main steps are as follows:

Step1：Initialize the nodes in each layer, $node[i]=C_N^i\ (i=0,1,2\ldots N)$, and all the nodes are signed with "*un-reduced"*, *i*=*N*, which means the layers of the information system;

Step2：Set *j* = *node*[*i*], which means the number of the nodes in each layer;

Step3：For the un-reduced node *j* in layer *i*, find out every reduce form of the node;

Step4：If the reduce form is itself, then change the sign symbol is changed with "*reduced"*, and set *node*[*i*]＝*node*[*i*]-1, then switch to Step6;

Step5：Change all the sign symbol of each *reduced* node, and delete the superfluous *un-reduced* node according to character of reduce lattice, then calculate the new numbers of the *un-reduced* node which denoted with *node*[*k*], where *k* means the layer having been modified;

Step6：Set *j* = *node*[*i*], if $j\neq 0$, switch to Step3;

Step7：Set *i* = *i*-1， if *i* =1then stop the algorithm, otherwise, switch to Step2.

3、The searching and optimizing of default rules

In the initial lattice, use $N_i^{(q)}$ denote the node of "*i*" in layer *q*, and the decision support system is $S=(U,C_i^{(q)}\bigcup D)$, all the rules in this layer corresponding are listed:

$$U\ /\ Ind(D)=\{Y_1,Y_2,\ldots Y_m\}$$

$$U\ /\ Ind(C_i^{(q)})=\{X_1,X_2,\ldots X_t\}$$

If （$X_k\cap Y_j$）$\neq 0$ (k=1,2,⋯,t; j=1,2,⋯,m), then a default rule can be created: $Des(X_k)\rightarrow Des(Y_j)$,its reliability is expressed with $R(X_i,Y_j)$:

$$R(X_i,Y_j)=\frac{Card(X_i\bigcap Y_j)}{Card(X_i)} \tag{1}$$

Its support is expressed with $S(X_i,Y_j)$:

$$S(X_i,Y_j)=\frac{Card(X_i\bigcap Y_j)}{Card(U)} \tag{2}$$

Suppose R_{tr} denotes the given criterion reliability, If $R(X_i,Y_j)=R_{tr}$,then the rule corresponding can be put into the rules set. If some $R(X_i,Y_j)<R_{tr}$, then all the rules' reliability of the successive node are also less than μ_{tr} , which means if some $R(X_i,Y_j)<R_{tr}$,then all the process of calculating can be overleaped, and

we can only conserve the rules which accord with $R(X_i, Y_j) > R_{tr}$. So the reduce lattice and rules can be simplify continuatively.

To each node in the reduce lattice, the rule can be found out correspondingly. If the rules set is empty, then the node can be deleted, and if there more than one node accord with the R_{tr}, then construct the new rules set which are suited the criterion. If the rules set is no-empty, then choose the top-priority rule as the output rule.

The top-priority of appraisement criterion in every default rules are as follows:

1) The rules in top-layer have the top-priority, which means the connotative information can be used totally.

2) If the rules in the same layer, then assuring rules are top-priority, which means that the $R = 1$ has the top-priority;

3) If some rules are in the same layer, and each rule have the character of $R = 1$, then the more "support" has the more top- priority;

4) If some rules are in the same layer, and all the $R < 1$, then the number of appraisement function is bigger means the rules has the more top-priority. The appraisement function is defined as follows:

Assume all the rules have some conclusion of （$d_1, d_2, \ldots, d_r$）, and the rule of d_i is （$r_1, r_2, \ldots, r_r$）,then the appraisement function of d_i is below:

$$v(d_i) = \frac{\sum_{j=1}^{m} S_j}{\sum_{j=1}^{m} (S_j / R_j)} \qquad (3)$$

Where, S_j、R_j express the *Support* and the *Reliability* of the rule r_j respectively.

4、Example

Suppose there is a IT project ,ant its decision system is $S = (U, C \cup D, V, f)$, and the sample number of each object is expressed with *K*, the believe threshold of each rule is denoted by $R_{tr} = 0.75$,which are showed in Tab.1.

The main algorithms and steps are as follows:

Step1: Construction the initial lattice, and set $N=4$, which means there are 4 condition attributes. The number of node in each layer is: $node[i] = C_4^i = \{1, 4, 6, 4, 1\}$ $(i = 0, 1, 2, 3, 4)$, which are showed in Fig.1.

Step2: In the initial lattice, each node is signed with no-reduction symbol. Set $i = 4$, and simplify the nodes in i=4 layer. There is j=*node*[4]=1 node which is no-reduced in this layer, this node is {*abcd*}. The two reduced sets {*b*,*c*,*d*} and {*a*,*b*}can be calculated, so the node "*bcd*" and "*ab*" in the initial lattice can be signed with reduced symbol. As a result , *node*[3]=3，*node*[2]=5. So we can deleted the node "*abcd*" which included the reduced of {*b*,*c*,*d*} in i=4 layer, the *node*[4]=0. By the same way, the node of "*abc*、*abd*" can also be deleted which included the reduced {*a*,*b*} in i=3 layer, so the *node*[3]=1.Because of j=*node*[4]=0， i=i-1=3, so it switch to i=3 layer.

Step3：set i = 3, all the nodes can be reduced in this layer. Because there is only j=*node*[3]=1 no-reduced node "*acd*", and its reduced is {*a*,*d*}, so the node of "*ad*" can be signed with the symbol of reduced. Then *node*[2]=4, and the node "*acd*" which included {*a*,*d*} can be deleted, so the j=*node*[3]=0， i=i-1=2, and it switch to i=2 layer.

Step4：set i = 2, the node in this layer can be reduced: there are j=*node*[2]=4 no-reduced nodes, because each node's reduced is itself, so each node can be signed with the symbol of reduced. Then *node*[2]=0, i=i-1=1, so all the nodes are reduced and finish all the process. As showed in Fig.2:

Tab.1 Decision table of information

U	Sample	Condition attributes				Decision attributes
	K	a	b	c	d	D
x_1	15	1	1	2	2	1
x_2	10	2	1	2	3	1
x_3	10	2	2	1	2	1
x_4	60	1	2	1	3	2
x_5	10	1	3	2	3	2
x_6	30	2	3	1	2	3
x_7	25	3	2	1	2	3
x_8	40	3	2	3	1	3

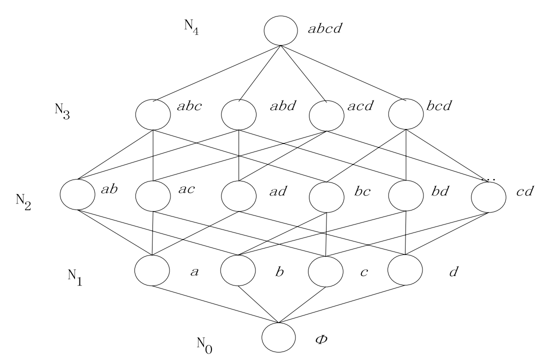

Fig.1 Initial lattice

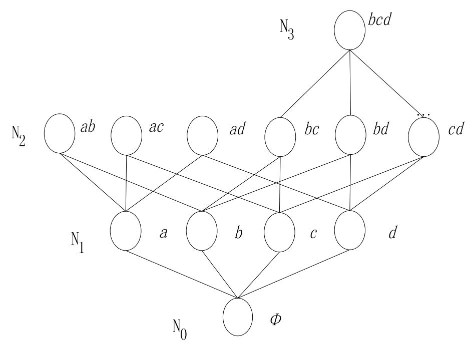

Fig.2 Reduced lattice

To each node of the reduced lattice in Fig.2, find out all the rules for each node, and put those rules into a set, the reliability and support can also be calculated according to formula（1）and（2）. For example, the rule and its reliability and support of the node "*bcd*" are as following:

$b_1c_2d_2 \rightarrow D_1$，$R$=15/15=1.000，$S$=15/200=0.075；
$b_1c_2d_3 \rightarrow D_1$，$R$=10/10=1.000，$S$=10/200=0.050；
$b_2c_1d_2 \rightarrow D_1$，$R$=10/(10+25)=0.286，$S$=10/200=0.050；
$b_2c_1d_3 \rightarrow D_2$，R=60/60=1.000，$S$=60/200=0.300；
$b_3c_2d_3 \rightarrow D_2$，$R$=10/10=1.000，$S$=10/200=0.050；
$b_3c_1d_2 \rightarrow D_3$，$R$=30/30=1.000，$S$=30/200=0.150；
$b_2c_1d_2 \rightarrow D_3$，R=25/(10+25)=0.714，$S$=25/200=0.125；
$b_2c_3d_1 \rightarrow D_3$，$R$=40/40=1.000，$S$=40/200=0.200；

In the same way, all the rules, its reliability and its support of each node can be calculated with the method mentioned above, which are showed in Tab.2:

Tab.2 The rules set of reduced nodes

Node	*Rule*	*Reliability (R)*	*Support (S)*	*Node*	*Rule*	*Reliability (R)*	*Support (S)*
a	$a_1 \rightarrow D_1$	0.176	0.075	*ad*	$a_1d_2 \rightarrow D_1$	1.000	0.075
	$a_1 \rightarrow D_2$	0.824	0.350		$a_1d_3 \rightarrow D_2$	1.000	0.350
	$a_2 \rightarrow D_1$	0.400	0.100		$a_2d_3 \rightarrow D_1$	1.000	0.050
	$a_2 \rightarrow D_3$	0.600	0.150		$a_2d_2 \rightarrow D_1$	0.250	0.050
	$a_3 \rightarrow D_3$	1.000	0.325		$a_2d_2 \rightarrow D_3$	0.750	0.150
b	$b_1 \rightarrow D_1$	1.000	0.125		$a_3d_2 \rightarrow D_3$	1.000	0.125
	$b_2 \rightarrow D_1$	0.074	0.300		$a_3d_1 \rightarrow D_1$	1.000	0.200
	$b_2 \rightarrow D_2$	0.444	0.050	*bc*	$b_1c_2 \rightarrow D_1$	1.000	0.125
	$b_2 \rightarrow D_3$	0.482	0.325		$b_2c_1 \rightarrow D_1$	0.105	0.050
	$b_3 \rightarrow D_2$	0.250	0.050		$b_2c_1 \rightarrow D_2$	0.632	0.300
	$b_3 \rightarrow D_3$	0.750	0.150		$b_2c_1 \rightarrow D_3$	0.263	0.125
c	$c_1 \rightarrow D_1$	0.080	0.050		$b_2c_3 \rightarrow D_3$	1.000	0.200
	$c_1 \rightarrow D_2$	0.480	0.300		$b_3c_1 \rightarrow D_3$	1.000	0.150
	$c_1 \rightarrow D_3$	0.440	0.275		$b_3c_2 \rightarrow D_2$	1.000	0.050

	$c_2 \to D_1$	0.714	0.125		$b_1d_2 \to D_1$	1.000	0.075
	$c_2 \to D_2$	0.286	0.050		$b_1d_3 \to D_1$	1.000	0.050
	$c_3 \to D_1$	1.000	0.200		$b_2d_1 \to D_3$	1.000	0.200
	$d_2 \to D_1$	0.312	0.125	*bd*	$b_2d_2 \to D_1$	0.286	0.050
	$d_2 \to D_3$	0.688	0.275		$b_2d_2 \to D_3$	0.714	0.125
d	$d_3 \to D_2$	0.750	0.350		$b_2d_3 \to D_2$	1.000	0.300
	$d_3 \to D_1$	0.250	0.050		$b_3d_2 \to D_3$	1.000	0.150
	$d_1 \to D_3$	1.000	0.200		$b_3d_3 \to D_2$	1.000	0.050
	$a_1b_1 \to D_1$	1.000	0.075		$c_2d_2 \to D_1$	1.000	0.075
	$a_1b_2 \to D_2$	1.000	0.300		$c_1d_3 \to D_3$	1.000	0.300
	$a_1b_3 \to D_2$	1.000	0.050		$c_2d_3 \to D_2$	0.500	0.050
ab	$a_2b_1 \to D_1$	1.000	0.050	*cd*	$c_2d_3 \to D_1$	0.500	0.050
	$a_2b_2 \to D_1$	1.000	0.050		$c_1d_2 \to D_1$	0.154	0.050
	$a_2b_3 \to D_3$	1.000	0.150		$c_1d_2 \to D_3$	0.846	0.275
	$a_3b_2 \to D_3$	1.000	0.325		$c_3d_1 \to D_3$	1.000	0.200
	$a_1c_2 \to D_1$	0.600	0.075		$b_1c_2d_2 \to D_1$	1.000	0.075
	$a_1c_2 \to D_2$	0.400	0.050		$b_2c_1d_3 \to D_2$	1.000	0.300
	$a_1c_1 \to D_2$	1.000	0.300		$b_3c_2d_3 \to D_2$	1.000	0.050
	$a_2c_1 \to D_1$	0. 250	0.050		$b_1c_2d_3 \to D_1$	1.000	0.050
ac	$a_2c_1 \to D_3$	0. 750	0.150	*bcd*	$b_2c_1d_2 \to D_1$	0.286	0.050
	$a_2c_2 \to D_1$	1.000	0.050		$b_3c_1d_2 \to D_3$	1.000	0.150
	$a_3c_1 \to D_2$	1.000	0.125		$b_2c_1d_2 \to D_3$	0.714	0.125
	$a_3c_3 \to D_3$	1.000	0.200		$b_2c_3d_1 \to D_3$	1.000	0.200

All the rules can be extracted from Tab.2, which accord with $R_{tr} \geq 0.75$, and the rules set can be showed in Tab.3:

Tab.3 The rules set of reduced nodes（$R_{tr} \geq 0.75$）

Node	*Rule*	*Reliability (R)*	*Support (S)*	*Node*	*Rule*	*Reliability (R)*	*Support (S)*
a	$a_1 \to D_2$	0.824	0.350		$a_1d_2 \to D_1$	1.000	0.075
	$a_3 \to D_3$	1.000	0.325		$a_1d_3 \to D_2$	1.000	0.350
b	$b_1 \to D_1$	1.000	0.125	*ad*	$a_2d_3 \to D_1$	1.000	0.050
	$b_3 \to D_3$	0.750	0.150		$a_2d_2 \to D_3$	0.750	0.150
c	$c_3 \to D_1$	1.000	0.200		$a_3d_2 \to D_3$	1.000	0.125
d	$d_3 \to D_2$	0.750	0.350		$a_3d_1 \to D_1$	1.000	0.200
ab	$a_1b_1 \to D_1$	1.000	0.075	*bd*	$b_1d_2 \to D_1$	1.000	0.075
	$a_1b_2 \to D_2$	1.000	0.300		$b_1d_3 \to D_1$	1.000	0.050
	$a_1b_3 \to D_2$	1.000	0.050		$b_2d_1 \to D_3$	1.000	0.200
	$a_2b_1 \to D_1$	1.000	0.050		$b_2d_3 \to D_2$	1.000	0.300
	$a_2b_2 \to D_1$	1.000	0.050		$b_3d_2 \to D_3$	1.000	0.150

	$a_2b_3\to D_3$	1.000	0.150		$b_3d_3\to D_2$	1.000	0.050
	$a_3b_2\to D_3$	1.000	0.325	cd	$c_2d_2\to D_1$	1.000	0.075
ac	$a_1c_1\to D_2$	1.000	0.300		$c_1d_3\to D_3$	1.000	0.300
	$a_2c_1\to D_3$	0. 750	0.150		$c_1d_2\to D_3$	0.846	0.275
	$a_2c_2\to D_1$	1.000	0.050		$c_3d_1\to D_3$	1.000	0.200
	$a_3c_1\to D_2$	1.000	0.125	bcd	$b_1c_2d_2\to D_1$	1.000	0.075
	$a_3c_3\to D_3$	1.000	0.200		$b_2c_1d_3\to D_2$	1.000	0.300
bc	$b_1c_2\to D_1$	1.000	0.125		$b_3c_2d_3\to D_2$	1.000	0.050
	$b_2c_3\to D_3$	1.000	0.200		$b_1c_2d_3\to D_1$	1.000	0.050
	$b_3c_1\to D_3$	1.000	0.150		$b_3c_1d_2\to D_3$	1.000	0.150
	$b_3c_2\to D_2$	1.000	0.050		$b_2c_3d_1\to D_3$	1.000	0.200

It can be concluded from the Tab.3 above, the default rules mining model of educed lattice can provide the satisfying decision rule according to the different layer. Due to its mechanism resemble humans' thinking, those method is very useful and flexible, it can conveniently arrive at a conclusion as good as possible at the matching rule sets according to the given criterion.

5、Conclusion

The advantages of the method mentioned above in this paper are as follows: 1) The reliability and support can be given when the data and information are incomplete. 2) The satisfying results can be approved when the data is inconsistent, which means that the system has the character and ability of allowable error. 3) The decision maker can choose the different layer and node freely.4) The default rules are simple and regular, especially unrepeated, which is propitious to make decision according to the new data.

6、Acknowledgement

This research was supported by the National Natural Science Foundation of China（70571025）, and also supported by the DanGui Project of Huazhong Noraml University（06DG024）.

References

1. Skowron A, Polkowski L, Komorowski J. Learning Tolerance relations by Boolean descriptors, *Automatic Feature Extraction from Data Tables.* In:Tsumoto[C]. Kobayashi,Yokomor,Tanka,and Nakamura.RSFD, 1996:11-17
2. Mollestad T.A Rough set approach to default rules data mining[D]. Norway: the Norwegian Institute of Technology,1996
3. Wang Yaying, Shao Huihe. A Rough Set Model to Mine Default Rule *Journal of Shanghai Jiaotong University*. 2000 ,5:691-694
Hu Keyun; Lu Yuchang; Shi Chunyi. Advances in concept lattice and its application[J].

Journal of Tsinghua University. 2000,9:77-81
4. Marzena Kryszkiewicz. Rough set approach to incomplete information systems[J]. *Information science,* 1998, 112: 39-49
Marzena Kryszkiewicz. Rules in incomplete information systems[J]. *Information science,* 1999, 113: 271-292
5. Zdzislwa Pawlak. Rough sets and intelligent data analysis [J], *Information sciences,* 2002,147:1-12
6. Wu Bing. Web mining model based on rough set theory. *Journal of southeast university*[J]. 2002.3:54-58
7. Wille R. Restructuring Lattice theory: An Approach Based on Hierarchies of Concepts[C].In: Rival I ed. Ordered Set.Dorerecht-Boston:Reidel.1982:445-470
8. Lu Xinyuan.*A sutydy on the risk decision rules mining of IT project* [M], Hubei People Press,2006,
9. MO Xiangqun. Risk management of software project schedule [J]. *Computer and Information Technology*,2002,4:67-70
10. Jian Lirong, Da Qing li,Chen Weida. A method of rule induction based on rough set theory in inconsistent information system [J]. *Chinese Journal of Management Science*. 2003,8: 91-95
11. Pang Sulin, Li Rongzhou, The decision mechanism of credit risk for banks with imperfect information [J]. *Journal of South China University of Technology,* 1999.8
12. Pawlak Z. Rough sets[J].*International Journal of Information and Computer Science*,1982,11:341-356.
13. Wang Guoyin. *The Rough set theory and the knowledge acquirement [M*]. Xi'an:Xi'an Jiaotong university press,2001,.5.
14. Holt, Gary D. Classifying construction contractors: a case using cluster analysis (Periodical style). *Building Research & Information,* 1997, 25(11): 374-378
15. Chinyio, E. A., Olomolaiye, P. O., Kometa, S. T. and Harris, F. C., A needs-based methodology for classifying construction clients and selecting contractors (Periodical style). Construction Management & Economics, 1998, 16(1): 91-95
16. Koen Milis, Roger Mercken, The use of the balanced scorecard for the evaluation of Information and Communication Technology projects (Periodical style). International Journal of Project Management, 2004(22): 87-97
17. M. Banerjee, S. Mitra, S.K. Pal, Rough fuzzy MLP: knowledge encoding and classification, IEEE Trans. *Neural Networks 9* (Nov) (1998) 1203 – 1216.
18. XU Zeshui. A method for multi-objective decision making with incomplete information [J]. *Operations research and management science* 2001.6
19. Liao Xiuwu; Tang Huanwen. A new group decision-making approach with incomplete information [J]. *Journal of Dalian University of Technology* 2002.1