Relative Association Rules Based on Rough Set Theory

Shu-Hsien Liao[1], Yin-Ju Chen[2], and Shiu-Hwei Ho[3]

[1] Department of Management Sciences, Tamkang University,
No.151 Yingzhuan Rd., Danshui Dist., New Taipei City 25137, Taiwan R.O.C
[2] Graduate Institute of Management Sciences, Tamkang University,
No.151 Yingzhuan Rd., Danshui Dist., New Taipei City 25137, Taiwan R.O.C
[3] Department of Business Administration,
Technology and Science Institute of Northern Taiwan,
No. 2, Xueyuan Rd., Peitou, 112 Taipei, Taiwan, R.O.C
michael@mail.tku.edu.tw, s5515124@ms18.hinet.net,
succ04.dba@msa.hinet.net

Abstract. The traditional association rule that should be fixed in order to avoid the following: only trivial rules are retained and interesting rules are not discarded. In fact, the situations that use the relative comparison to express are more complete than those that use the absolute comparison. Through relative comparison, we proposes a new approach for mining association rule, which has the ability to handle uncertainty in the classing process, so that we can reduce information loss and enhance the result of data mining. In this paper, the new approach can be applied for finding association rules, which have the ability to handle uncertainty in the classing process, is suitable for interval data types, and help the decision to try to find the relative association rules within the ranking data.

Keywords: Rough set, Data mining, Relative association rule, Ordinal data.

1 Introduction

Many algorithms have been proposed for mining Boolean association rules. However, very little work has been done in mining quantitative association rules. Although we can transform quantitative attributes into Boolean attributes, this approach is not effective, is difficult to scale up for high-dimensional cases, and may also result in many imprecise association rules [2]. In addition, the rules express the relation between pairs of items and are defined in two measures: support and confidence. Most of the techniques used for finding association rule scan the whole data set, evaluate all possible rules, and retain only those rules that have support and confidence greater than thresholds. It's mean that the situations that use the absolute comparison [3]. The remainder of this paper is organized as follows. Section 2 reviews relevant literature in correlation with research and the problem statement. Section 3 incorporation of rough set for classification processing. Closing remarks and future work are presented in Section 4.

B.-L. Lu, L. Zhang, and J. Kwok (Eds.): ICONIP 2011, Part II, LNCS 7063, pp. 185–192, 2011.

2 Literature Review and Problem Statement

In the traditional design, Likert Scale uses a checklist for answering and asks the subject to choose only one best answer for each item. The quantification of the data is equal intervals of integer. For example, age is the most common type for the quantification data that have to transform into an interval of integer. Table 1 and Table 2 present the same data. The difference is due to the decision maker's background. One can see that the same data of the results has changed after the decision maker transformation of the interval of integer. An alternative is the qualitative description of process states, for example by means of the discretization of continuous variable spaces in intervals [6].

Table 1. A decision maker

No	Age	Interval of integer
t_1	20	20–25
t_2	23	26–30
t_3	17	Under 20
t_4	30	26–30
t_5	22	20–25

Table 2. B decision maker

No	Age	Interval of integer
t_1	20	Under 25
t_2	23	Under 25
t_3	17	Under 25
t_4	30	Above 25
t_5	22	Under 25

Furthermore, in this research, we incorporate association rules with rough sets and promote a new point of view in applications. In fact, there is no rule for the choice of the "right" connective, so this choice is always arbitrary to some extent.

3 Incorporation of Rough Set for Classification Processing

The traditional association rule, which pays no attention to finding rules from ordinal data. Furthermore, in this research, we incorporate association rules with rough sets and promote a new point of view in interval data type applications. The data processing of interval scale data is described as below.

First: Data processing—Definition 1—Information system: Transform the questionnaire answers into information system $IS = (U, Q)$, where $U = \{x_1, x_2, \cdots x_n\}$ is a finite set of objects. Q is usually divided into two parts, $G = \{g_1, g_2, \cdots g_i\}$ is a finite set of general attributes/criteria, and $D = \{d_1, d_2, \cdots d_k\}$ is a set of decision attributes. $f_g = U \times G \to V_g$ is called the information function, V_g is the domain of the attribute/criterion g, and f_g is a total function such that $f(x, g) \in V_g$ for each $g \in Q$; $x \in U$. $f_d = U \times D \to V_d$ is called the sorting decision-making information function, V_d is the domain of the decision attributes/criterion d, and f_d is a total function such that $f(x, d) \in V_d$ for each $d \in Q$; $x \in U$.

Example: According to Tables 3 and 4, x_1 is a male who is thirty years old and has an income of 35,000. He ranks beer brands from one to eight as follows: Heineken, Miller, Taiwan light beer, Taiwan beer, Taiwan draft beer, Tsingtao, Kirin, and Budweiser.

Then:

$$f_{d_1}=\{4,3,1\} \qquad f_{d_2}=\{4,3,2,1\} \qquad f_{d_3}=\{6,3\} \qquad f_{d_4}=\{7,2\}$$

Table 3. Information system

U \ Q	General attributes G		Decision-making D
	Item1: Age g_1	Item2: Income g_2	Item3: Beer brand recall
x_1	30 g_{1_1}	35,000 g_{2_1}	As shown in Table 4.
x_2	40 g_{1_2}	60,000 g_{2_2}	As shown in Table 4.
x_3	45 g_{1_3}	80,000 g_{2_4}	As shown in Table 4.
x_4	30 g_{1_1}	35,000 g_{2_1}	As shown in Table 4.
x_5	40 g_{1_2}	70,000 g_{2_3}	As shown in Table 4.

Table 4. Beer brand recall ranking table

U	D the sorting decision-making set of beer brand recall							
	Taiwan beer d_1	Heineken d_2	light beer d_3	Miller d_4	draft beer d_5	Tsingtao d_6	Kirin d_7	Budweiser d_8
x_1	4	1	3	2	5	6	7	8
x_2	1	2	3	7	5	6	4	8
x_3	1	4	3	2	5	6	7	8
x_4	3	1	6	2	5	4	8	7
x_5	1	3	6	2	5	4	8	7

Definition 2: The Information system is a quantity attribute, such as g_1 and g_2, in Table 3; therefore, between the two attributes will have a covariance, denoted by $\sigma_G = Cov(g_i, g_j)$. $\rho_G = \dfrac{\sigma_G}{\sqrt{Var(g_i)}\sqrt{Var(g_j)}}$ denote the population correlation coefficient and $-1 \le \rho_G \le 1$.

Then:

$$\rho_G^+ = \{g_{ij} | 0 < \rho_G \le 1\} \qquad \rho_G^- = \{g_{ij} | -1 \le \rho_G < 0\} \qquad \rho_G^0 = \{g_{ij} | \rho_G = 0\}$$

Definition 3—Similarity relation: According to the specific universe of discourse classification, a similarity relation of the decision attributes $d \in D$ is denoted as U/D

$$S(D) = U|D = \{[x_i]_D | x_i \in U, V_{d_k} > V_{d_l}\}$$

Example:

$$S(d_1) = U/d_1 = \{\{x_1\}, \{x_4\}, \{x_2 x_3, x_5\}\}$$

$$S(d_2) = U/d_2 = \{\{x_3\}, \{x_5\}, \{x_2\}, \{x_1, x_4\}\}$$

Definition 4—Potential relation between general attribute and decision attributes: The decision attributes in the information system are an ordered set, therefore, the attribute values will have an ordinal relation defined as follows:

$$\sigma_{GD} = Cov(g_i, d_k) \qquad \rho_{GD} = \frac{\sigma_{GD}}{\sqrt{Var(g_i)}\sqrt{Var(d_k)}}$$

Then:

$$F(G, D) = \begin{cases} \rho_{GD}^{+} & : 0 < \rho_{GD} \le 1 \\ \rho_{GD}^{-} & : -1 \le \rho_{GD} < 0 \\ \rho_{GD}^{0} & : \rho_{GD} = 0 \end{cases}$$

Second: Generated rough associational rule—Definition 1: The first step in this study, we have found the potential relation between general attribute and decision attributes, hence in the step, the object is to generated rough associational rule. To consider other attributes and the core attribute of ordinal-scale data as the highest decision-making attributes is hereby to establish the decision table and the ease to generate rules, as shown in Table 5. $DT = (U, Q)$, where $U = \{x_1, x_2, \cdots x_n\}$ is a finite set of objects, Q is usually divides into two parts, $G = \{g_1, g_2, \cdots g_m\}$ is a finite set of general attributes/criteria, $D = \{d_1, d_2, \cdots d_l\}$ is a set of decision attributes. $f_g = U \times G \to V_g$ is called the information function, V_g is the domain of the attribute/criterion g, and f_g is a total function such that $f(x, g) \in V_g$ for each $g \in Q$; $x \in U$. $f_d = U \times D \to V_d$ is called the sorting decision-making information function, V_d is the domain of the decision attributes/criterion d, and f_d is a total function such that $f(x, d) \in V_d$ for each $d \in Q$; $x \in U$.

Then:

$$f_{g_1} = \{\text{Price, Brand}\} \qquad f_{g_2} = \{\text{Seen on shelves, Advertising}\}$$

$$f_{g_3} = \{\text{purchase by promotions, will not purchase by promotions}\}$$

$$f_{g_4} = \{\text{Convenience Stores, Hypermarkets}\}$$

Definition 2: According to the specific universe of discourse classification, a similarity relation of the general attributes is denoted by $U/_G$. All of the similarity relation is denoted by $K = (U, R_1, R_2 \cdots R_{m-1})$.

$$U|G = \{[x_i]_G | x_i \in U\}$$

Example:

$$R_1 = \frac{U}{g_1} = \{\{x_1, x_2, x_5\}, \{x_3, x_4\}\} \qquad R_6 = \frac{U}{g_2 g_4} = \{\{x_1, x_3, x_4\}, \{x_2, x_5\}\}$$

$$\vdots \qquad\qquad \vdots$$

$$R_5 = \frac{U}{g_1 g_3} = \{\{x_1, x_2, x_5\}, \{x_3, x_4\}\} \qquad R_{m-1} = \frac{U}{G} = \{\{x_1\}, \{x_2, x_5\}, \{x_3, x_4\}\}$$

Table 5. Decision-making

Q	General attributes				Decision attributes	
U	Product Features g_1	Product Information Source g_2	Consumer Behavior g_3	Channels g_4	Rank	Brand
x_1	Price	Seen on shelves	purchase by promotions	Convenience Stores	4	d_1
x_2	Price	Advertising	purchase by promotions	Hypermarkets	1	d_1
x_3	Brand	Seen on shelves	will not purchase by promotions	Convenience Stores	1	d_1
x_4	Brand	Seen on shelves	will not purchase by promotions	Convenience Stores	3	d_1
x_5	Price	Advertising	purchase by promotions	Hypermarkets	1	d_1

Definition 3: According to the similarity relation, and then finding the reduct and core. If the attribute g which were ignored from G, the set G will not be affected; thereby, g is an unnecessary attribute, we can reduct it. $R \subseteq G$ and $\forall_g \in R$. A similarity relation of the general attributes from the decision table is denoted by $ind(G)$. If $ind(G) = ind(G - g_1)$, then g_1 is the reduct attribute, and if $ind(G) \neq ind(G - g_1)$, then g_1 is the core attribute.

Example:

$$U|ind(G) = \{\{x_1\}, \{x_2, x_5\}, \{x_3, x_4\}\}$$

$$U|ind(G - g_1) = U|(\{g_2, g_3, g_4\}) = \{\{x_1\}, \{x_2, x_5\}, \{x_3, x_4\}\} = U|ind(G)$$

$$U|ind(G - g_1 g_3) = U|(\{g_2, g_4\}) = \{\{x_1, x_3, x_4\}, \{x_2, x_5\}\} \neq U|ind(G)$$

When g_1 is considered alone, g_1 is the reduct attribute, but when g_1 and g_3 are considered simultaneously, g_1 and g_3 are the core attributes.

Definition 4: The lower approximation, denoted as $\underline{G}(X)$, is defined as the union of all these elementary sets, which are contained in $[x_i]_G$. More formally,

$$\underline{G}(X) = \cup\left\{[x_i]_G \in \frac{U}{G} \middle\| [x_i]_G \subseteq X\right\}$$

The upper approximation, denoted as $\overline{G}(X)$, is the union of these elementary sets, which have a non-empty intersection with $[x_i]_G$. More formally:

$$\overline{G}(X) = \cup\left\{[x_i]_G \subseteq \frac{U}{G} \middle\| [x_i]_G \cap X \neq \phi\right\}$$

The difference $Bn_G(X) = \overline{G}(X) - \underline{G}(X)$ is called the boundary of $[x_i]_G$.

Example: $\{x_1, x_2, x_4\}$ are those customers that we are interested in, thereby $\underline{G}(X) = \{x_1\}$, $\overline{G}(X) = \{x_1, x_2, x_3, x_4, x_5\}$ and $Bn_G(X) = \{x_2, x_3, x_4, x_5\}$.

Definition 5: Rough set-based association rules.

$$\frac{\{x_1\}}{g_1 g_3} : g_{1_1} \cap g_{3_1} \Rightarrow d_{d_1}^1 = 4 \qquad \frac{\{x_1\}}{g_1 g_2 g_3 g_4} : g_{1_1} \cap g_{2_1} \cap g_{3_1} \cap g_{4_1} \Rightarrow d_{d_1}^1 = 4$$

Algorithm-Step1

Input:
Information System (IS);
Output:
{Potential relation};
Method:

1. Begin
2. $IS = (U, Q)$;
3. $x_1, x_2, \cdots, x_n \in U$; /* where $x_1, x_2, \cdots, x_n$ are the objects of set U */
4. $G, D \subset Q$; /* Q is divided into two parts G and D */
5. $g_1, g_2, \cdots, g_i \in G$; /* where $g_1, g_2, \cdots, g_i$ are the elements of set G */
6. $d_1, d_2, \cdots, d_k \in D$; /* where $d_1, d_2, \cdots, d_k$ are the elements of set D */
7. For each g_i and d_k do;
8. compute $f(x, g)$ and $f(x, d)$; /* compute the information function in IS as described in definition1*/
9. compute σ_G; /* compute the quantity attribute covariance in IS as described in definition2*/

10. compute ρ_G; /* compute the quantity attribute correlation coefficient in IS as described in definition2*/
11. compute $S(D)$ and $S(D)$; /* compute the similarity relation in IS as described in definition3*/
12. compute $F(G,D)$; /* compute the **p**otential relation as described in definition4*/
13. Endfor;
14. Output {Potential relation};
15. End;

Algorithm-Step2
Input:
Decision Table (DT);
Output:
{Classification Rules};
Method:
1. Begin
2. $DT = (U,Q)$;
3. $x_1, x_2, \cdots x_n \in U$; /* where $x_1, x_2, \cdots x_n$ are the objects of set U */
4. $Q = (G,D)$;
5. $g_1, g_2, \cdots, g_m \in G$; /* where $g_1, g_2, \cdots, g_m$ are the elements of set G */
6. $d_1, d_2, \cdots, d_l \in D$; /* where $d_1, d_2, \cdots, d_l$ are the "trust value" generated in Step1*/
7. For each d_l do;
8. compute $f(x,g)$; /* compute the information function in DT as described in definition1*/
9. compute R_m; /* compute the similarity relation in DT as described in definition2*/
10. compute $ind(G)$; /* compute the relative reduct of DT as described in definition3*/
11. compute $ind(G - g_m)$; /* compute the relative reduct of the elements for element m as described in definition3*/
12. compute $\underline{G}(X)$; /* compute the lower-approximation of DT as described in definition4*/
13. compute $\overline{G}(X)$; /* compute the upper-approximation of DT as described in definition4*/
14. compute $Bn_G(X)$; /* compute the bound of DT as described in definition4*/
15. Endfor;
16. Output {Association Rules};
17. End;

4 Conclusion and Future Works

The quantitative data are popular in practical databases; a natural extension is finding association rules from quantitative data. To solve this problem, previous research partitioned the value of a quantitative attribute into a set of intervals so that the traditional algorithms for nominal data could be applied [1]. In addition, most of the techniques used for finding association rule scan the whole data set, evaluate all possible rules, and retain only the rules that have support and confidence greater than thresholds [3]. The new association rule algorithm, which tries to combine with rough set theory to provide more easily explained rules for the user. In the research, we use a two-step algorithm to find the relative association rules. It will be easier for the user to find the association. Because, in the first step, we find out the relationship between the two quantities attribute data, and then we find whether the ordinal scale data has a potential relationship with those quantities attribute data. It can avoid human error caused by lack of experience in the process that quantities attribute data transform to categorical data. At the same time, we known the potential relationship between the quantities attribute data and ordinal-scale data. In the second step, we use the rough set theory benefit, which has the ability to handle uncertainty in the classing process, and find out the relative association rules. The user in mining association rules does not have to set a threshold and generate all association rules that have support and confidence greater than the user-specified thresholds. In this way, the association rules will be a relative association rules. The new association rule algorithm, which tries to combine with the rough set theory to provide more easily explained rules for the user. For the convenience of the users, to design an expert support system will help to improve the efficiency of the user.

Acknowledgements. This research was funded by the National Science Council, Taiwan, Republic of China, under contract NSC 100-2410-H-032 -018-MY3.

References

1. Chen, Y.L., Weng, C.H.: Mining association rules from imprecise ordinal data. Fuzzy Sets and Systems 159, 460–474 (2008)
2. Lian, W., Cheung, D.W., Yiu, S.M.: An efficient algorithm for finding dense regions for mining quantitative association rules. Computers and Mathematics with Applications 50(3-4), 471–490 (2005)
3. Liao, S.H., Chen, Y.J.: A rough association rule is applicable for knowledge discovery. In: IEEE International Conference on Intelligent Computing and Intelligent Systems (ICIS 2009), ShangHai, China (2009)
4. Liu, G., Zhu, Y.: Credit Assessment of Contractors: A Rough Set Method. Tsinghua Science & Technology 11, 357–363 (2006)
5. Pawlak, Z.: Rough sets, decision algorithms and Bayes' theorem. European Journal of Operational Research 136, 181–189 (2002)
6. Rebolledo, M.: Rough intervals—enhancing intervals for qualitative modeling of technical systems. Artificial Intelligence 170(8-9), 667–668 (2006)