Interpretations of Association Rules by Granular Computing

Yuefeng Li [a] and Ning Zhong [b]

aSchool of Software Engineering & Data Communications
Queensland University of Technology, Brisbane, OLD 4001, Australia
E-mail: y2.li@qut.edu.au
bDepartment of Systems and Information Engineering
Maebashi Institute of Technology, Maebashi 371-0816, Japan
E-mail: zhong@maebashi-it.ac.jp

Abstract

This paper presents interpretations for association rules. It first introduces Pawlak's method, and the corresponding algorithm of finding decision rules (a kind of association rules). It then uses extended random sets to present a new algorithm of finding interesting rules. It proves that the new algorithm is faster than Pawlak's algorithm. The extended random sets are easily to include more than one criterion for determining interesting rules. They also provide two measures for dealing with uncertainties in association rules.

1. Introduction

Mining association rules has received much attention currently [7]. The frequency of occurrence is a well-accepted criterion for mining association rules. Apart from the frequency, the rules should reflect real world patterns [1] [4]. It is desirable to use some mathematical models to interpret association rules in order to obtain real world patterns.

The patterns hidden in data can be characterized by rough set theory [6], in which the premises of association rules (or called decision rules in [6]) are interpreted as condition granules, and the post-conditions are interpreted as decision granules. The measure of uncertainties for decision rules is based on well-established statistical models. This only reveals the objective aspect of decision rules. However, knowledge in some applications is based on "subjective" judgments.

In this paper, we use granular computing to interpret association rules. We first introduce Pawlak's method [6] and formally describe the corresponding algorithm for determining strengths and certainty factors of decision rules. We then present a new interpretation of association rules using extended random sets [3]. An effective algorithm of finding interesting rules is proposed using the new interpretation. We also show that an extended random set can be interpreted as a probability function (which can provide an "objective" interpretation) or a belief function (which can provide a "subjective" interpretation).

2. Databases to Decision Tables

Let U be a non-empty finite set of objects (a set of records), and A be a set of attributes (or fields). We call a pair $S = (U, A)$ an *information table* if there is a function for every attribute $a \in A$ such that $a: U \to V_a$, where V_a is the set of all values of a. We call V_a the domain of a.

Let B be a subset of A. B determines a binary relation $I(B)$ on U such that $(x, y) \in I(B)$ if and only if $a(x) = a(y)$ for every $a \in B$, where $a(x)$ denotes the value of attribute a for element $x \in U$. it is easy to prove that $I(B)$ is an equivalence relation, and the family of all equivalence classes of $I(B)$, that is a partition determined by B, is denoted by $U/I(B)$ or simply by U/B. The classes in U/B are referred to *B-granules* or *B-elementary sets*. The class which contains x is called *B-granule* induced by x, and is denoted by $B(x)$.

A user may use some attributes of a database. We can divide the user used attributes into two groups: condition attributes and decision attributes, respectively. We call the tripe (U, C, D) a *decision table* of (U, A) if $C \cap D = \varnothing$ and $C \cup D \subseteq A$, where (U, C, D) is a set of classes and each class is the representative of a group of records.

For example, we assume that there is an information table (relation) that includes 1000 records of vehicle accidents, where the set of attributes is $A = \{driver, vehicle\ type, weather, road, time, accident\}$. If the user only uses 4 attributes and let $C = \{weather, road\}$, and $D = \{time, accident\}$. $C \cup D$ determines a binary relation $I(C \cup D)$ on U, and U is classified into 7 equivalence classes, as shown in Table 1 (i.e. a decision table), where N is the number of records in the corresponding class.

Using Table 1, we also can get the set of condition granules, $U/C = \{\{1,7\}, \{2,5\}, \{3,6\}, \{4\}\}$, and decision granules, $U/D = \{\{1\}, \{2,3,7\}, \{4\}, \{5,6\}\}$, respectively.

In the following we let $U/C = \{c_1, c_2, c_3, c_4\}$ and $U/D = \{d_1, d_2, d_3, d_4\}$.

Table 1. A decision table

Class	Weather	Road	Time	Accident	N
1	Misty	Icy	Day	Yes	80
2	Foggy	Icy	Night	Yes	140
3	Misty	Not icy	Night	Yes	40
4	Sunny	Icy	Day	No	500
5	Foggy	Icy	Night	No	20
6	Misty	Not icy	Night	No	200
7	Misty	Icy	Night	Yes	20

3. Pawlak's Method

Every class in a decision table can be mapped into an association rule (or called decision rule in [6]), e.g., class 2 in Table 1 can be read as "*if the weather is foggy **and** road is icy **then** the accident occurred at night*" in 140 cases.

For our convenience, we assume $A = \{a_1,..., a_k, a_{k+1}, ..., a_m\}$, $C = \{a_1, ..., a_k\}$, and $D = \{a_{k+1}, ..., a_m\}$, where $k>0$, and $m>k$. Every class f determines a sequence $a_1(f)$, ..., $a_k(f)$, $a_{k+1}(f)$, ..., $a_m(f)$. The sequence can determine a decision rule:

$$a_1(f), ..., a_k(f) \rightarrow a_{k+1}(f), ..., a_m(f)$$

or in short $f(C) \rightarrow f(D)$.

The strength of the decision rule $f(C) \rightarrow f(D)$ is defined as $|C(f) \cap D(f)| / |U|$; and the certainty factor of the decision rule is defined as $|C(f) \cap D(f)| / |C(f)|$.

According to the above definitions, we can use the following algorithm to calculate strengths and certainty factors for all decision rules, where we assume N_i denotes the number of records in class i, and UN denotes the total number of records in U.

Algorithm 1.
1. let $UN = 0$;
2. for ($i = 1$ to n) // n is the number of classes
   $UN = UN + N_i$;
3. for ($i = 1$ to n)
   $\{strength(i) = N_i/UN$; $CN = N_i$;
   for ($j = 1$ to n)
      if (($j \neq i$) and ($f_j(C) == f_i(C)$))
      $CN = CN + N_j$;
   $certainty\_factor(i) = N_i/CN$; $\}$.

If we assume the basic operation is the comparison between two classes (i.e., $f_j(C) == f_i(C)$), then the time complexity is $(n-1) \times n = O(n^2)$, where n is the number of

classes in the decision table. It also needs a similar algorithm to determine interesting rules for Pawlak's method.

4. Extended Random Sets

Let U/C be the set of condition granules and U/D be the set of decision granules. To describe the relationship between condition granules and decision granules, we can rewrite the decision rules in Table 1 as follows:

$c_1 \rightarrow \{ (d_1, 80/100), (d_2, 20/100) \}$
$c_2 \rightarrow \{ (d_2, 140/160), (d_4, 20/160) \}$
$c_3 \rightarrow \{ (d_2, 40/240) (d_4, 200/240) \}$
$c_4 \rightarrow \{ (d_3, 500/500) \}$.

These determine a mapping Γ from U/C to $2^{(U/D) \times [0,1]}$, such that

$$\sum_{(fst,snd) \in \Gamma(c_i)} snd = 1 \text{ for all } c_i \in U/C$$

where $\Gamma(c_i)$ is a set of decision-granule numeral pairs.

Now we consider the support degree for each condition granule. The obvious way is to use the frequency in the decision table, that is,

$$w(c_i) = \sum_{x \in c_i} N_x \qquad (1)$$

for every condition granule c_i, where N_x is the number of records in class x. By normalizing, we can get a probability function P on U/C such that

$$P(c_i) = \frac{w(c_i)}{\sum_{c_j \in U/C} w(c_j)}$$

for all $c_i \in U/C$.

Based on the above analysis, we can use a pair (Γ, P) to represent what we can obtain from an information table. We call the pair (Γ, P) an *extended random set*.

According to the definitions in the previous section, we can obtain the following decision rules:

$$c_i \rightarrow fst_{i,1}, \ c_i \rightarrow fst_{i,2}, \ ..., \text{and } c_i \rightarrow fst_{i,|\Gamma(c_i)|}$$

for a given condition granule c_i, where

$$\Gamma(c_i) = \{(fst_{i,1}, snd_{i,1}), ..., (fst_{i,|\Gamma(c_i)|}, snd_{i,|\Gamma(c_i)|})\} \qquad (2)$$

We call $P(c_i) \times snd_{i,1}, ..., P(c_i) \times snd_{i,|\Gamma(c_i)|}$ the strengths of these decision rules, respectively; and $snd_{i,1}, ..., snd_{i,|\Gamma(c_i)|}$ the certainty factors, respectively.

From the above definitions, we have

$$snd_{i,j} = \frac{|c_i \cap fst_{i,j}|}{|c_i|}.$$

The above definitions about strengths and certainty factors are the same as Pawlak's definitions.

5. Determining Interesting Rules

Given an extended random set (Γ, P), it can provide a new representation for decision rules. Figure 1 shows a such

example for representing the extended random set that is obtained from Table 2, where U/C is the set of condition granules, and $\Gamma(c_i)$ is the set of conclusions of premise c_i ($i = 1, \ldots, |U/C|$). Algorithm 2 shows the process of creating extended random sets, and the process of calculating of strengths and certainty factors of the decision rules.
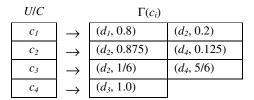
U/C		$\Gamma(c_i)$	
c_1	$\rightarrow$	$(d_1, 0.8)$	$(d_2, 0.2)$
c_2	$\rightarrow$	$(d_2, 0.875)$	$(d_4, 0.125)$
c_3	$\rightarrow$	$(d_2, 1/6)$	$(d_4, 5/6)$
c_4	$\rightarrow$	$(d_3, 1.0)$	

Figure 1. An extended random set

Algorithm 2.
1. let $UN = 0$, $U/C = \varnothing$;
2. for ($i = 1$ to n)
   $UN = UN + N_i$;
3. for ($i = 1$ to n do) // create the data structure
   if ($f_i(C) \in U/C$)
     $insert((f_i(D), N_i))$ to $\Gamma(f_i(C))$;
   else
     $add(f_i(C))$ into U/C, and set $\Gamma(f_i(C)) = \varnothing$;
4. for ($i = 1$ to $|U/C|$)
   $P(c_i) = (1/UN) \times (\sum_{(fst, snd) \in \Gamma(c_i)} snd)$;
5. for ($i = 1$ to $|U/C|$) // normalization
   { $temp = 0$;
     for (j = 1 to $|\Gamma(c_i)|$)
       $temp = temp + snd_{i,j}$;
     for (j = 1 to $|\Gamma(c_i)|$)
       $snd_{i,j} = snd_{i,j}/temp$; }
6. for ($i = 1$ to $|U/C|$) // calculate rule strengths
   for (j = 1 to $|\Gamma(c_i)|$)
     { $strength(c_i \rightarrow fst_{i,j}) = P(c_i) \times snd_{i,j}$;
       $certainty\_factor(c_i \rightarrow fst_{i,j}) = snd_{i,j}$; }.

Because steps 4, 5, and 6 all traverse pairs in $\Gamma(c_i)$ ($i = 1, \ldots, |U/C|$), and the number of pairs in all $\Gamma(c_i)$ ($i = 1, \ldots, |U/C|$) is just n (the number of classes in the decision table), the time complexity of this algorithm is determined by step 3. In step 3, checking $f_i(C) \in U/C$ takes O($|U/C|$), so the time complexity of the algorithm is O($n \times |U/C|$), where the basic operation is still the comparison between classes. Since $|U/C| \leq n$, Algorithm 2 is better than Algorithm 1 for the time complexity.

A decision rule $c_i \rightarrow fst_{i,j}$ is an *interesting rule* if $pr(fst_{i,j} | c_i) - pr(fst_{i,j})$ is greater than a suitable constant.

From the definition of mapping Γ, we have $pr(fst_{i,j} | c_i) = snd_{i,j}$. To decide the probability on the set of decision granules, we present the following function:

$pr : (U/D) \rightarrow [0,1]$ such that
$$pr(d) = \sum_{c_i \in (U/C), (d, snd) \in \Gamma(c_i)} P(c_i) \times snd \qquad (3)$$

We can prove that pr is a probability function on (U/D). The algorithm of determining pr is only to traverse the data structure as showed in Figure 1.

Table 2. Probability function on decision granules

Decision Granule	Description	pr
d_1	*Accident occurred at night*	0.08
d_2	*Accident occurred in daytime*	0.20
d_3	*Accident not occurred in daytime*	0.50
d_4	*Accident not occurred at night*	0.22

Table 2 shows the probability function on the set of decision granules. From Figure 1 and Table 2 we can obtain the probability of $pr(fst_{i,j} | c_i)$ for every decision rule $c_i \rightarrow fst_{i,j}$, where $fst_{i,j} \in \{d_1, d_2, d_3, d_4\}$. We can get 4 interesting rules (seeTable 3) if we assume that a decision rule $c_i \rightarrow fst_{i,j}$ is an interesting rule iff $pr(fst_{i,j} | c_i) - pr(fst_{i,j}) > 0$.

Table 3. Interesting rules

Rule Description	$pr(fst_{i,j} | c_i)$	$pr(fst_{i,j})$	Interesting rule
$c_1 \rightarrow d_1$	0.8	0.08	*Yes*
$c_2 \rightarrow d_2$	0.875	0.20	*Yes*
$c_3 \rightarrow d_2$	0.167	0.20	*No*
$c_4 \rightarrow d_3$	1	0.5	*Yes*
$c_2 \rightarrow d_4$	0.125	0.22	*No*
$c_3 \rightarrow d_4$	0.833	0.22	*Yes*
$c_1 \rightarrow d_2$	0.20	0.20	*No*

6. Discussions

In this section, we discuss other advantages of our approach except the time complexity. We first discuss the weight functions for condition granules. We also introduce another uncertain measure on decision granules.

IEEE
COMPUTER
SOCIETY

6.1 Weight Functions for Condition Granules

The extended random sets are easily to include other criteria apart from the well-accepted criterion "frequencies" when determining association rules.

Although the frequency is a well-accepted criterion for data mining, it is not the only criterion for support degree because some condition granules with high frequencies may be meaningless. For example, when we use keywords to represent the meaning of documents, we usually consider both keywords frequency and inverse document frequency (e.g., the popular technique $tf*idf$ in information retrieval) because some words (like "information") may have high frequencies in a document but they may appear in most documents in a collection (e.g., "information" may appears in most documents in the information table collection).

In order to use the above idea, we assume there is a collection which contains many databases. Given a decision table (U, C, D) of a database, we can define a new weight function w on U/C to instead of the weight function in Eq. (1), which satisfies

$$w(c_i) = (\sum_{x \in c_i} N_x) \times log(M/n_i)$$

for every $c_i \in (U/C)$, where M is the total number of databases, and n_i is the number of databases which contain the given condition granule c_i. It is easy to do so because this definition does not affect the rest calculation of decision rules.

6.2 Uncertain Measures on Decision Granules

The obvious way to measure the uncertainty of a set of decision granules is using a probability function. For a given set of decision granules $X = \{d_1, \dots, d_s\}$, we may use $\sum_{x \in X} pr(x)$ to represent the probability of (d_1 or … or d_s). However, this measure is very sensitive to the frequencies of records.

To consider a relative stable measure, we consider a random set (ξ, P) (see [2] [5]) which is derived from the extended random set (Γ, P):

$\xi : U / C \to 2^{U/D}$ such that

$\xi(c_i) = \{fst | (fst, snd) \in \Gamma(c_i)\}$ for every $c_i \in (U/C)$.

The random set (ξ, P) determines a Dempster-Shafer mass function (see [2]) m_P on U/D such that

$$m_P(X) = P(\{c_i | c_i \in (U/C), \xi(c_i) = X\}) \qquad (4)$$

for every $X \subseteq U/D$.

This mass function can decide a belief function and plausibility function (see [2]) as well. They are defined as follows:

$bel_m : 2^{U/D} \to [0,1]$, $pl_m : 2^{U/D} \to [0,1]$ and

$$bel_m(X) = \sum_{Y \subseteq X} m_P(Y), \quad pl_m(X) = \sum_{Y \cap X \neq \varnothing} m_P(Y) \quad (5)$$

for every $X \subseteq U/D$.

We can prove that

$$bel_m(X) \leq \sum_{x \in X} pr(x) \leq pl_m(X) \text{ for every } X \subseteq U/D.$$

Domain experts can use the interval $[bel_m, pl_m]$ to check if their "subjective" judgments for some descriptions are correct. Table 4 shows the uncertainty measures for some descriptions.

Table 4. Uncertain measures

Description	Subset	Pr	m_P	$[bel_m, pl_m]$
"d_1 or d_2"	$\{d_1, d_2\}$	0.28	0.1	[0.1, 0.5]
"d_3 or d_4"	$\{d_3, d_4\}$	0.72	0.0	[0.5, 0.9]
"d_2 or d_3"	$\{d_2, d_3\}$	0.70	0.0	[0.5, 1.0]
"d_1 or d_4"	$\{d_1, d_4\}$	0.30	0.0	[0.0, 0.5]

7. Conclusions

This paper uses granular computing to interpret association rules. The main contribution of this paper is that the concept of extended random sets is used to describe the relationships between condition granules and decision granules. It presents a new efficient algorithm to find interesting rules in databases. Apart from the "frequencies", the extended random sets are easily to include other criteria when determining association rules. The extended random sets also provide more than one measure for dealing with uncertainties in the association rules significantly.

References

[1] E. Cohen, et al., Finding interesting associations without support pruning, *IEEE Transactions on Knowledge and Data Engineering*, 2001, 13(1): 64-78.

[2] R. Kruse, E. Schwecke and J. Heinsoln, *Uncertainty and vagueness in knowledge based systems (Numerical Methods)*, Springer-Verlag, New York, 1991.

[3] Y. Li, Extended random sets for knowledge discovery in information system, in Proc. *the 9th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, China, 2003, 524-532.

[4] T. Y. Lin, The lattice structure of database and mining multiple level rules, *Bulletin of International Rough Set Society*, 2002, 6(1/2): 11-16.

[5] D. Liu and Y. Li, The interpretation of generalized evidence theory, *Chinese Journal of Computers*, 1997, 20(2): 158-164.

[6] Z. Pawlak, In pursuit of patterns in data reasoning from data, the rough set way, *3rd International Conference on Rough Sets and Current Trends in Computing*, USA, 2002, 1-9.

[7] R. Rastogi and K. Shim, Mining optimized association rules with categorical and numeric attributes, *IEEE Transactions on Knowledge and Data Engineering*, 2002, 14(1): 29-50.

IEEE COMPUTER SOCIETY