Granular association rules with four subtypes

Fan Min\textsuperscript{a}, Qinghua Hu\textsuperscript{b} and William Zhu\textsuperscript{a}
\textsuperscript{a}Lab of Granular Computing, Zhangzhou Normal University, Zhangzhou 363000, China.
\textsuperscript{b}Tianjin University, Tianjin 300072, China
Email: minfanphd@163.com, huqinghua@hit.edu.cn, williamfengzhu@gmail.com

Abstract—Relational data mining approaches look for patterns that involve multiple tables; therefore they become popular in recent years. In this paper, we introduce granular association rules to reveal connections between concepts in two universes. An example of such an association might be “men like alcohol.” We present four meaningful explanations corresponding to four subtypes of granular association rules. We also define five measures to evaluate the quality of rules. Based on these measures, the relationships among different subtypes are revealed. This work opens a new research trend concerning granular computing and associate rule mining.

Keywords—Granular computing, relational data mining, granular association rule, complete match, partial match.

I. INTRODUCTION

Two types of rules are more popular in associate rule mining [1], [2], [3]. The first type, called boolean association rule, reveals the connection between two disjoint subsets of the same universe. A well known application is mining basket data that stores items purchased on a per-transaction basis [1]. An example of such rule is “30% of transactions that contain beer also contain diapers; 2% of all transactions contain both of these items.” Here 30% is called the confidence of the rule, and 2% is the support of the rule. The second type, called quantitative association rule, reveals the relationships among attribute values of an object. A well known application is mining information of people. An example of such rule is “10% of married people between age 50 and 60 have at least 2 cars; 3% of all people queried satisfy this rule” [2]. These two types are based on a single data table.

As an important direction of relational data mining (RDM) [4], [5], relational association rule mining [5], [6], [7] looks for patterns that involve multiple tables. Jensen et al. [8] redefined support and confidence measures under the new context. The frequency of an itemset over multiple relations is expressed in the number of occurrences in the result of a join of two tables. Goethals et al. [7] argued that with this definition, “it is hard to determine the true cause of the frequency.” Then they proposed to count the occurrences of unique objects in the first and the second table, respectively.

In this paper, we introduce a new type, called granular association rule, to reveal the connection between concepts within different universes. Compared with the work in [6] which considers one-to-many relationship, the new type considers many-to-many relationship. While compared with the work in [7] which may involve more than two universes, the new type has more semantic meanings. The term “granular” comes from granular computing [9], [10], [11]. It indicates that concepts can take any granular specified by an attribute subset. Two examples of granular association rules are “Chinese men like France alcohol” and “young women like white products.” Here the left part of the first rule is concerned with county and gender; while the left part of the second one is concerned with age and gender. In contrast, multilevel association rules [12] follow a predefined concept hierarchy with a tree structure. For example, both part of a multilevel association rule should start from the category, followed by country, color, price, and so on in sequence.

We present four subtypes of granular association rules with different explanations to a rule. Let us consider the following granular association rule “men like alcohol.” The first subtype is complete match with the explanation “all men like all kinds of alcohol.” The second subtype is left-hand side partial match with an explanation “40% men like all kinds of alcohol.” The third subtype is right-hand side partial match with an explanation “all men like at least 30% kinds of alcohol.” The fourth subtype is partial match with an explanation “40% men like at least 30% kinds of alcohol.” In fact, left-hand side match and right-hand side match are more general than complete match; at the same time they are special cases of partial match.

To evaluate the quality of rules, we define five measures, namely source coverage, target coverage, support, confidence, and target confidence. The definitions of source coverage and target coverage are shared by all subtypes. In contrast, other measures may have different definitions on different subtypes. In most cases, different measures of a rule can be computed independently. However, for partial match granular association rules, there is a tradeoff between confidence and target confidence. Hence we should specify one of them before computing the other one.

II. PRELIMINARIES

In this section, we revisit two popular association rule mining problems through examples. Boolean associate rules were proposed by Agrawal et al. [1] and named by Srikant et al. [2] to distinguish it from other types. Quantitative
association rules were proposed by Srikant et al. [2] to address information systems containing both quantitative and symbolic attributes. Han and Kamber [3] presented meaningful and detailed discussions on these issues.

A. Boolean association rules

The data model for boolean association rules is a binary relation on two universes. It is defined as follows:

Definition 1: \( R \subseteq U \times V \) is a binary relation from \( U \) to \( V \), where \( U = \{x_1, x_2, \ldots, x_n\} \) and \( V = \{y_1, y_2, \ldots, y_m\} \) are two sets of objects.

\[
R(x) = \{y \in V| (x, y) \in R\}, \tag{1}
\]

\[
R^{-1}(y) = \{x \in U| (x, y) \in R\}. \tag{2}
\]

In supermarket transaction analysis, \( U \) is the set of transactions, and \( V \) is the set of items. \((x_i, y_j) \in R\) indicates that item \( y_j \) is included in transaction \( x_i \). This type of data are often referred to as basket data [1]. \( R \) is often represented as an \( n \times m \) boolean matrix, and stored as a boolean table in a database.

Table I illustrates an example basket data, where transactions are named after the customers, 1 indicates that an item is included in respective transaction, and 0 for otherwise. The \( support \) of any \( V' \subseteq V \) is the percentage of transactions in \( U \) that contains all elements in \( V' \). That is,

\[
support(V') = \frac{|\{x \in U| R(x) \supseteq V'\}|}{|U|}, \tag{3}
\]

where \(| \cdot | \) denotes the cardinality of a set. A boolean association rule is an implication of the form

\[
(BR) : V_1 \Rightarrow V_2, \tag{4}
\]

where \( V_1, V_2 \subseteq V \) and \( V_1 \cap V_2 = \emptyset \). It reveals the connection between two disjoint itemsets. Transaction information is employed to evaluate the quality of the association. The \( support \) of the boolean association rule is

\[
support(BR) = support(V_1 \cup V_2), \tag{5}
\]

It reflects the usefulness of the rule. The \( confidence \) of the boolean association rule is

\[
confidence(BR) = \frac{support(V_1 \cup V_2)}{support(V_1)}. \tag{6}
\]

It reflects the certainty of the rule. The following problem is the most often addressed in association rule mining.

Table I

<table>
<thead>
<tr>
<th>( U \setminus V )</th>
<th>Bread</th>
<th>Diaper</th>
<th>Pork</th>
<th>Beef</th>
<th>Wine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ron</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Michelle</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Shun</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Yamago</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Wang</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table II

<table>
<thead>
<tr>
<th>( U )</th>
<th>Age</th>
<th>Gender</th>
<th>Married</th>
<th>Country</th>
<th>Income</th>
<th>NumCars</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ron</td>
<td>20-29 Male No USA 60k-69k</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Michelle</td>
<td>20-29 Female Yes USA 80k-89k</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shun</td>
<td>20-29 Male No China 40k-49k</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yamago</td>
<td>30-39 Female Yes Japan 80k-89k</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang</td>
<td>30-39 Male Yes China 90k-99k</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Problem 1: The boolean association rule problem.

Input: A binary relation \( R \) representing supermarket transactions or similar data; a minimal support threshold \( ms \); a minimal confidence threshold \( mc \).

Output: All boolean association rules satisfying

\[
support(BR) \geq ms \text{ and } confidence(BR) \geq mc.
\]

Given \( R \) as illustrated Table I. Let \( ms = 0.4 \) and \( mc = 0.6 \). We obtain many boolean association rules, e.g.,

\[
\{\text{Diaper}\} \Rightarrow \{\text{Beer}\} \quad \text{[support = 60%, confidence = 100%];} \tag{7}
\]

\[
\{\text{Beer}\} \Rightarrow \{\text{Diaper}\} \quad \text{[support = 60%, confidence = 75%]; and} \tag{8}
\]

\[
\{\text{Bread, Beef}\} \Rightarrow \{\text{Wine}\} \quad \text{[support = 40%, confidence = 67%].} \tag{9}
\]

Problem 1 was essentially proposed by Agrawal et al. in 1993 [1]. Then the Apriori algorithm was designed in 1994 [13]. Han et al. designed the FP-growth algorithm [14] to avoid candidate generation. With these algorithms, Problem 1 can be tackled efficiently on large datasets.

B. Quantitative association rules

The data model for quantitative association rules is an information system. It is defined as follows:

Definition 2: \( S = (U, A) \) is an information system, where \( U = \{x_1, x_2, \ldots, x_n\} \) is the set of all objects, \( A = \{a_1, a_2, \ldots, a_m\} \) is the set of all attributes, and \( a_j(x_i) \) is the value of \( x_i \) on attribute \( a_j \) for \( i \in \{1..n\} \) and \( j \in \{1..m\} \).

An information system is often stored as an information table in a database or a text file in a file system. The data type of \( a_j(x_i) \) may be boolean, symbolic, numeric, interval, or more complex ones. Information systems with symbolic and numeric data are also called quantitative information systems [2], with an example illustrated in Table II. Note that the domains of Age, Income and NumCars are already partitioned into a number of intervals. Respective techniques, called discretization and symbolic value partition, have been extensively studied in the literature [15], [16]. In the following context, we will assume that the information system contains only symbolic values.

In an information system, any \( A' \subseteq A \) induces an equivalent relation [17]

\[
E_A = \{(x, y)| \forall a \in A', a(x) = a(y)\}, \tag{10}
\]
and partitions $U$ into a number of disjoint subsets called blocks. The block containing $x \in U$ is

$$E_A(x) = \{y | \forall a \in A', a(y) = a(x)\}. \quad (11)$$

From another viewpoint, a pair $C = (A', x)$ where $x \in U$ is called a concept. The extension of the concept is

$$ET(C) = ET(A', x) = E_A(x); \quad (12)$$

while the intension of the concept is the conjunction of respective attribute-value pairs, i.e.,

$$IT(C) = IT(A', x) = \bigwedge_{a \in A'} \langle a : a(x) \rangle. \quad (13)$$

The support of the concept is the size of its extension divided by the size of the universe, namely,

$$support(C) = support(\bigwedge_{a \in A'} \langle a : a(x) \rangle) = support(E_A(x)) = \frac{|ET(A', x)|}{|U|}. \quad (14)$$

A quantitative association rule is an implication of the form

$$(QR) : \bigwedge_{a \in A_1} \langle a : a(x) \rangle \Rightarrow \bigwedge_{a \in A_2} \langle a : a(x) \rangle, \quad (15)$$

where $\emptyset \subset A_1, A_2 \subset A$ and $A_1 \cap A_2 = \emptyset$. It reveals the relationships among attribute values of an object. In fact, a quantitative association rule has the same form as a decision rule. The difference lies in that a decision rule has one or more predefined decision attributes.

The boolean information system contains only boolean values. Moreover, the boolean association rule in (4) can be interpreted as

$$\bigwedge_{a \in V_1} \langle a : 1 \rangle \Rightarrow \bigwedge_{a \in V_2} \langle a : 1 \rangle. \quad (16)$$

Hence in the form the boolean association rule is a special case of the quantitative association rule. However, boolean values in the boolean information system discussed in Section II-A is viewed not as an attribute value of the object, but the relationship between objects in $U$ and $V$. Therefore, from the semantic point of view they are different.

The support of the quantitative association rule $QR$ given by Equation (15) is

$$support(QR) = support(\bigwedge_{a \in A_1 \cup A_2} \langle a : a(x) \rangle). \quad (17)$$

The confidence of $QR$ is

$$confidence(QR) = \frac{support(\bigwedge_{a \in A_1 \cup A_2} \langle a : a(x) \rangle)}{support(\bigwedge_{a \in A_1} \langle a : a(x) \rangle)} \quad (18)$$

The following problem is most often addressed under this context.

**Problem 2**: The quantitative association rule problem.

<table>
<thead>
<tr>
<th>$V$</th>
<th>Country</th>
<th>Category</th>
<th>Color</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bread</td>
<td>Australia</td>
<td>Food</td>
<td>Black</td>
<td>1.9</td>
</tr>
<tr>
<td>Diaper</td>
<td>China</td>
<td>Daily</td>
<td>White</td>
<td>1.9</td>
</tr>
<tr>
<td>Pork</td>
<td>China</td>
<td>Meat</td>
<td>Red</td>
<td>1.9</td>
</tr>
<tr>
<td>Beef</td>
<td>Australia</td>
<td>Meat</td>
<td>Red</td>
<td>10.19</td>
</tr>
<tr>
<td>Beer</td>
<td>France</td>
<td>Alcohol</td>
<td>Black</td>
<td>10.19</td>
</tr>
<tr>
<td>Wine</td>
<td>France</td>
<td>Alcohol</td>
<td>White</td>
<td>10.19</td>
</tr>
</tbody>
</table>

**Input**: An information system $S = (U, A)$; a minimal support threshold $ms$; a minimal confidence threshold $mc$.

**Output**: All quantitative association rules satisfying $support(QR) \geq ms$ and $confidence(QR) \geq mc$.

Given $S$ as illustrated Table II. Let $ms = 0.4$ and $mc = 0.6$. We obtain many associate rules, e.g. [2].

(19) 

$$(\text{Age} : 30..39) \wedge (\text{Married: Yes}) \Rightarrow (\text{NumCars: 2}) \quad [support = 40\%, confidence = 100\%]; \quad \text{and}$$

(20) 

$$(\text{NumCars: 0..1}) \Rightarrow (\text{Married: No}) \quad [support = 40\%, confidence = 67\%].$$

An information system can be converted into one with only boolean data through scaling [16], [18]. In this way Problem 2 is mapped into Problem 1 [2].

**III. GRANULAR ASSOCIATION RULES**

In this section, we first discuss the data model for granular association rules. The we present four types of rules corresponding to four different explanation of granular association rules. A number of measures are also proposed to evaluate the quality of these rules.

**A. The data model**

The data model for granular association rules is the many-to-many relation in databases. In relational databases, a many-to-many relation involves two universes and a relation. It can be defined as follows.

**Definition 3**: A many-to-many entity-relationship system (MMER) is a 5-tuple $ES = (U, A, V, B, R)$, where $(U, A)$ and $(V, B)$ are two information systems, and $R \subseteq U \times V$ is a binary relation from $U$ to $V$.

$R$ has been defined through Definition 1 and an example has been given by Table I. $(U, A)$ has been defined through Definition 2 and an example has been given by Table II. $(V, B)$ shares the same definition as $(U, A)$ and an example is given by Table III. Therefore MMER is completely defined by Definitions 1, 2 and 3. An example of MMER is given by Tables I, II and III.

**B. Granular association rule with four subtypes**

A granular association rule is an implication of the form

$$(GR) : \bigwedge_{a \in A'} \langle a : a(x) \rangle \Rightarrow \bigwedge_{b \in B'} \langle b : b(y) \rangle. \quad (21)$$
where \( A' \subseteq A \) and \( B' \subseteq B \). An example path from the root to a leaf is: \( \text{All} \rightarrow \text{Computer} \rightarrow \text{Laptop} \rightarrow \text{IBM} \). 

According to Equation (14), the set of objects meeting the left-hand side of the granular association rule is

\[
LH(GR) = E_{A'}(x);
\]

while the set of objects meeting the right-hand side of the granular association rule is

\[
RH(GR) = E_{B'}(y).
\]

We define two measures to evaluate the generality of the granular association rule. The source coverage of \( GR \) is

\[
scoverage(GR) = \frac{|LH(GR)|}{|U|};
\]

while the target coverage of \( GR \) is

\[
tcoverage(GR) = \frac{|RH(GR)|}{|V|}.
\]

In most cases, rules with higher source coverage and target coverage tend to be more interesting. We present a granular association rule for discussion.

\[
\langle \text{Gender: Male} \rangle \Rightarrow \langle \text{Category: Alcohol} \rangle \\
[\text{scoverage} = 60\%, \text{tcoverage} = 33\%].
\]

A direct explanation of Rule (26) is “men like alcohol.” However, this explanation is ambiguous and the following questions may arise: Do all men like alcohol? Do men like all kinds of alcohol? To avoid such ambiguous, more accurate specification of the rule is needed. We propose four different explanations of this rule, as illustrated in Fig. 1, and will discuss them one by one. Note that example rules discussed in the following context may not comply to the MMER illustrated in Tables I, II and III.

1) Complete match: The first explanation of Rule (26) is “all men like all alcohol,” or equivalently, “100% men like 100% alcohol.” This can be formally expressed by the following definition.

**Definition 4:** A granular associate rule \( GR \) is called a complete match granular association rule iff

\[
LH(GR) \times RH(GR) \subseteq R.
\]

It is also called a complete match rule for brevity. We need to know the percentage of objects in \( U \) matching the rule. It is called the support of the rule and defined by

\[
support_c(GR) = scoverage(GR) = \frac{|LH(GR)|}{|U|},
\]

where the suffix \( c \) stands for complete. Although the support is equal to the source coverage, we still define this measure since in other subtypes they are different. One may obtain the following rule

\[
\langle \text{Gender: Male} \rangle \Rightarrow \langle \text{Category: Alcohol} \rangle \\
[\text{scoverage} = 60\%, \text{tcoverage} = 33\%],
\]

which is read as “all men like all kinds of alcohol; 60% of all people are men; 33% of all products are alcohol.” Note that Rules (26) and (29) have the same form. However the explanation of Rule (29) causes no ambiguity.

2) Left-hand side partial match: The second explanation of Rule (26) is “some men like all alcohol,” or equivalently, “at least one man like 100% alcohol.” Because “some” appears on the left-hand side, the rule is called “left-hand side partial match.” Consequently, we define a subtype of granular associate rule as follows.

**Definition 5:** A granular associate rule \( GR \) is called a left-hand side partial match rule iff there exists \( x \in LH(GR) \) such that

\[
R(x) \supseteq RH(GR).
\]

In applications, however, if only very few men like all kinds of alcohol, this rule is not quite useful. We need to know how many men like alcohol, and the percentage of men that like alcohol. The support of the rule is

\[
support_{lp}(GR) = \frac{|\{x \in LH(GR) \mid R(x) \supseteq RH(GR)\}|}{|U|}.
\]

In other words, only men that like all kinds of alcohol are counted. Moreover, the confidence of the rule is

\[
confidence_{lp}(GR) = \frac{|\{x \in LH(GR) \mid R(x) \supseteq RH(GR)\}|}{|LH(GR)|}.
\]

One may obtain the following rule

\[
\langle \text{Gender: Male} \rangle \Rightarrow \langle \text{Category: Alcohol} \rangle \\
[\text{scoverage} = 60\%, \text{tcoverage} = 33\%, \text{confidence}_{lp} = 40\%],
\]

which is read as “40% men like all kinds of alcohol; 60% of all people are men; 33% of all products are alcohol.” We deliberately avoid the support measure in this explanation; the reason will be discussed in the next subsection.

3) Right-hand side partial match: The third explanation of Rule (26) is “all men like some kinds of alcohol,” or equivalently, “100 % men like at least one kind of alcohol.” Because “some” appears on the right-hand side, the rule is called “right-hand side partial match.” Consequently, we define a subtype of granular associate rule as follows.

**Definition 6:** A granular associate rule \( GR \) is called a right-hand side partial match rule iff \( \forall x \in LH(GR) \)

\[
R(x) \cap RH(GR) \neq \emptyset.
\]

Similar to the case of complete match, the support of the rule is equal to the source coverage. It is given by

\[
support_{rp}(GR) = scoverage(GR) = \frac{|LH(GR)|}{|U|}.
\]

In the case of complete match and left-hand side partial match, bigger target coverage values indicate stronger rules.
Unfortunately, in the case of right-hand side partial match, bigger target coverage values indicate weaker rules. Consider one extreme case as follows: “all people like at least one kind of all product.” The rule always holds, and both the source coverage and the target coverage of the rule are 100%, but the rule is totally useless.

Hence we need to know men like how many kinds of alcohol. Here we introduce a new measure called target confidence for this purpose. The target confidence of the right-hand side partial match rule is

$$t_{\text{confidence}}_{rp}(GR) = \min_{x \in LH(GR)} \frac{|R(x) \cap RH(GR)|}{|RH(GR)|}.$$  \hfill (36)

With existing measures, we may obtain the following rule

$$\langle \text{Gender: Male} \rangle \Rightarrow \langle \text{Category: Alcohol} \rangle$$

$$[s_{\text{coverage}} = 60\%, t_{\text{coverage}} = 33\%,$$  \hfill (37)

$$t_{\text{confidence}}_{rp} = 30\%].$$

which is read as “all men like at least 30% of alcohol; 60% of all people are men; 33% of all products are alcohol”

4) Partial match: The fourth explanation of Rule (26) is “Some men like some kinds of alcohol,” or equivalently, “at least one man like at least one kind of alcohol.” Because “some” appears on both sides, the rule will be simply called “partial match.” Consequently, we define a subtype of granular associate rule as follows.

Definition 7: A granular associate rule $GR$ is called a partial match granular association rule iff there exists $x \in LH(GR)$ and $y \in RH(GR)$ such that

$$(x, y) \in R.$$  \hfill (38)

It is also called a partial match rule for brevity.

There is a tradeoff between the confidence and the target confidence of a rule. Consequently, neither value can be obtained directly from the rule. To compute any one of them, we need to specify the threshold of the other. Let $tc$ be the target confidence threshold. The support of the partial match rule is

$$\text{support}_{tp}(GR) = \frac{|\{x \in LH(GR) \mid |R(x) \cap RH(GR)| \geq tc\}|}{|LH(GR)|}. $$  \hfill (39)

The confidence of the partial match rule is

$$\text{confidence}_{tp}(GR) = \frac{|\{x \in LH(GR) \mid |R(x) \cap RH(GR)| \geq tc\}|}{|RH(GR)|}. $$  \hfill (40)

Let $mc$ be the confidence threshold, and

$$\frac{|\{x \in LH(GR) \mid |R(x) \cap RH(GR)| \geq K + 1\}|}{|LH(GR)|}$$

$$\leq \frac{|\{x \in LH(GR) \mid |R(x) \cap RH(GR)| \geq K\}|}{|LH(GR)|}.$$  \hfill (41)

The target confidence of the partial match rule is

$$t_{\text{confidence}}_{tp}(GR) = \frac{K}{|RH(GR)|}. $$  \hfill (42)

In fact, the computation of K is non-trivial. First, for any $x \in LH(GR)$, we need to compute $tc(x) = |R(x) \cap RH(GR)|$ and obtain an array of integers. Second, we sort the array in a descending order. Third, let $k = \lfloor mc \times |LH(GR)| \rfloor$, $K$ is the $k$-th element in the array.

With existing measures, we may obtain the following rule

$$\langle \text{Gender: Male} \rangle \Rightarrow \langle \text{Category: Alcohol} \rangle$$

$$[s_{\text{coverage}} = 60\%, t_{\text{coverage}} = 33\%,$$  \hfill (43)

$$t_{\text{confidence}}_{tp} = 40\%, t_{\text{confidence}}_{tp} = 30\%].$$

which is read as “40% men like at least 30% of alcohol; 60% of all people are men; 33% of all products are alcohol”.

C. Discussion of measures

We have presented five measures to evaluate the quality of granular association rules. The source coverage is given by $\frac{|LH(GR)|}{|U|}$ and the target coverage is given by $\frac{|RH(GR)|}{|U|}$. There are relationships among different subtypes.

1) A left-hand side partial match rule $GR$ is a complete match rule iff $\text{confidence}_{lp}(GR) = 100\%$.

2) A right-hand side partial match rule $GR$ is a complete match rule iff $\text{confidence}_{rp}(GR) = 100\%$.

3) A partial match rule $GR$ is a left-hand (right-hand) side partial match rule iff $\text{confidence}_{lp}(GR) = 100\%$ $(\text{confidence}_{rp}(GR) = 100\%)$.

4) A partial match rule $GR$ is a complete match rule iff $\text{confidence}_{rp}(GR) = \text{confidence}_{lp}(GR) = 100\%$.

---

**Figure 1.** Four explanations of “men like alcohol.”
For all four subtypes, there is a direct connection among the support, source coverage and confidence of a rule.

\[ \text{support}_s(\text{GR}) = \text{source coverage}(\text{GR}) \times \text{confidence}_s(\text{GR}), \]

where the suffix “s” could be replaced by \( lp, rp \) and \( p \).

IV. CONCLUSIONS AND FURTHER WORKS

In this paper, we have proposed granular association rule with four subtypes. Different measures have been defined to evaluate the quality of these rules. In further works, we will study other types of data concerning interval-values [19], neighborhood [20], and costs [21], [22], [23]. Since rules form a covering of either universe, we will also employ covering-based rough sets [24], [25] for rule mining.

ACKNOWLEDGEMENTS

This work is supported in part by the National Natural Science Foundation of China under Grant No. 61170128, the Natural Science Foundation of Fujian Province, China, under Grant Nos. 2011J01374 and 2012J01294, and the Education Natural Science Foundation of Fujian Province, China, under Grant No. JA11176.

REFERENCES


