Mining Multidimensional Fuzzy Association Rules from a Normalized Database

Rolly Intan, Oviliani Yenty
Informatics Engineering Department
Petra Christian University, Surabaya, Indonesia
rintan@peter.petra.ac.id

Abstract

Mining association rules is one of the important tasks in the process of data mining application. In general, the input as used in the process of generating rules is taken from a certain data table by which all the corresponding values of every domain data have correlations one to each others as given in the data table. A problem arises when we need to generate the rules expressing the relationship between two or more domains that belong to several different tables in a normalized database. To overcome the problem, before generating rules it is necessary to join the participant tables into a general table by a process called Denormalization..

This paper shows a process of mining Multidimensional Fuzzy Association Rules from a normalized database. The process consists of two sub-process, namely sub-process of join tables (Denormalization) and sub-process of mining fuzzy rules. In general, some parts of mining the fuzzy association rules has been discussed in our previous papers [3,4,5,6].

1. Introduction

Association rule finds interesting association or correlation relationship among a large data set of items [1,9]. The discovery of interesting association rules can help in decision making process.

Association rule mining that implies a single predicate is referred as a single dimensional or intradimension association rule since it contains a single distinct predicate with multiple occurrences (the predicate occurs more than once within the rule). The terminology *of* single dimensional or intradimension association rule is used in multidimensional database by assuming each distinct predicate in the rule as a dimension [1].

Here, the method of *market basket analysis* can be extended and used for analyzing any context of database. For instance, database of medical track record patients is analyzed for finding association (correlation) among diseases taken from the data of complicated several diseases suffered by patients in a certain time. For example, it might be discovered a Boolean association rule "Bronchitis $\Rightarrow$ Lung Cancer" representing relation between "Bronchitis" and "Lung Cancer" which can also be written as a single dimensional association rule as follows:

Rule-1

$$Dis(X, \text{"Bronchitis"}) \Rightarrow Dis(X, \text{"Lung Cancer"}),$$

where *Dis* is a given predicate and X is a variable representing patient who have a kind of disease (i.e. "Bronchitis" and "Lung Cancer"). In general, "Lung Cancer" and "Bronchitis" are two different data that are taken from a certain data attribute, called *items*. In general, *Apriori* [1,9] is used an influential algorithm for mining frequent itemsets for mining Boolean (single dimensional) association rules.

Additional related information regarding the identity of patients, such as *age*, *occupation*, *sex*, *address*, *blood type*, etc., may also have a correlation to the illness of patients. Considering each data attribute as a predicate, it can therefore be interesting to mine association rules containing *multiple* predicates, such as:

Rule-2:

$$Age(X, \text{"60"}) \wedge Smk(X, \text{"yes"}) \Rightarrow Dis(X, \text{"LungCancer"}),$$

where there are three predicates, namely *Age*, *Smk* (*smoking)* and *Dis (disease)*. Association rules that involve two or more dimensions or predicates can be referred to as *multidimensional association rules*. Multidimensional association rules with no repeated predicate as given by Rule-2, are called *interdimension association rules* [1]. It may be interesting to mine multidimensional association rules with repeated predicates. These rules are called *hybrid-dimension association rules*, e.g.:

Rule-3:

978-0-7695-3328-5/08 $25.00 © 2008 IEEE
DOI 10.1109/ICHIT.2008.229

$$Age(X,"60") \wedge Smk(X,"yes") \wedge Dis(X,"Bronchitis") \Rightarrow Dis(X,"LungCancer"),$$

To provide a more meaningful association rule, it is necessary to utilize *fuzzy sets* over a given database attribute called *fuzzy association rule* as discussed in [4,5]. Formally, given a crisp domain *D*, any arbitrary fuzzy set (say, fuzzy set *A*) is defined by a membership function of the form [2,8]:

$$\mu_A : D \rightarrow [0,1]. \qquad (1)$$

A fuzzy set may be represented by a meaningful fuzzy label. For example, "*young*", "*middle-aged*" and "*old*" are fuzzy sets over *age* that is defined on the interval [0, 100] as arbitrarily given by[2]:

$$young(x) = \begin{cases} 1 & , x \le 20 \\ (35-x)/15 & , 20 < x < 35 \\ 0 & , x \ge 35 \end{cases}$$

$$middle\_aged(x) = \begin{cases} 0 & , x \le 20 \text{ or } x \ge 60 \\ (x-20)/15 & , 20 < x < 35 \\ (60-x)/15 & , 45 < x < 60 \\ 1 & , 35 \le x \le 45 \end{cases}$$

$$old(x) = \begin{cases} 0 & , x \le 45 \\ (x-45)/15 & , 45 < x < 60 \\ 1 & , x \ge 60 \end{cases}$$

.

Using the previous definition of fuzzy sets on *age*, an example of multidimensional fuzzy association rule relation among the predicates *Age*, *Smk* and *Dis* may then be represented by:

Rule-4

$$Age(X,"young") \wedge Smk(X,"yes") \Rightarrow Dis(X,"Bronchitis")$$

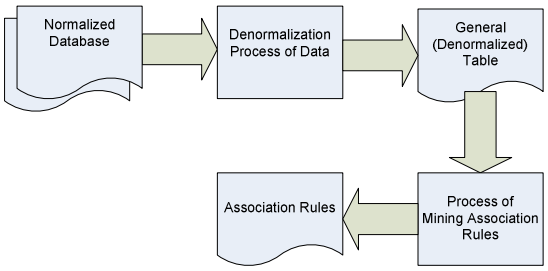

Figure 1. Process of Mining Association Rules

To generate multidimensional association rules implying fuzzy value such as given by Rule-4 from a normalized database that consists of several tables, this paper discussed two sequential processes as shown in Figure 1.

First is the process of joining tables known as Denormalization of Database. Second is the process of generating (mining) fuzzy association rules. The process of denormalization can be provided based on the relation of tables as presented in Entity Relationship Diagram (ERD) of the relational database.

For two tables that have no direct relation in ERD, they can still be joined by others transition tables (in ERD) using the transitive join process. Other solution is that we can define or create a relation function or a relation table that corresponds two distinct domains of the tables. Here, a metadata can be constructed as a data dictionary to express the relationship of tables. Result of denormalization data process is a single general (denormalized) table. The table is used as a source data for the process of mining association rules. Some parts of mining fuzzy rules has been discussed in [4,5,6] that introduced some formulations for calculating support and confidence factors. This paper also introduces a formula to calculate correlation factor as also usually used in evaluating interestingness of the rules.

The structure of the paper is the following. In Section 2, basic definition and formulation of some measures, support, correlation and confidence rule as used for determining interestingness of association rules are briefly recalled. Section 3 is devoted to propose data preparation for the further process of generation rules. Here, we will discuss a process of join table from a normalized database. Section 4 discusses a concept for mining multidimensional fuzzy association rules. Section 5 demonstrated the concept in an illustrative example. Finally a conclusion is given in Section 6.

2. Support, Confidence and Correlation

Association rules are kind of patterns representing correlation of attribute-value (items) in a given set of data provided by a process of data mining system. Generally, association rule is a conditional statement (such kind of *if-then rule*). More formally [1], association rules are the form $A \Rightarrow B$, that is, $a_1 \wedge \cdots \wedge a_m \Rightarrow b_1 \wedge \cdots \wedge b_n$, where a_i (for $i \in \{1,\ldots,m\}$) and b_j (for $j \in \{1,\ldots,n\}$) are two items (attribute-value). The association rule $A \Rightarrow B$ is interpreted as "*database tuples that satisfy the conditions in A are also likely to satisfy the conditions in B*". $A = \{a_1, \cdots, a_m\}$ and $B = \{b_1, \cdots, b_n\}$ are two distinct itemsets. Performance or interestingness of an

association rule is generally determined by three factors, namely *confidence*, *support* and *correlation* factors. Confidence is a measure of certainty to assess the validity of the rule. Given a set of relevant data tuples (or transactions in a relational database) the confidence of "$A \Rightarrow B$" is defined by:

$$\text{confidence}(A \Rightarrow B) = \frac{\#tuples(A \text{ and } B)}{\#tuples(A)}, \qquad (2)$$

where #*tuples*(*A* and *B*) means the number of tuples containing *A* and *B*.

For example, a confidence 80% for the Association Rule (for example Rule-1) means that 80% of all patients who infected bronchitis are likely to be also infected lung cancer. The support of an association rule refers to the percentage of relevant data tuples (or transactions) for which the pattern of the rule is true. For the association rule "$A \Rightarrow B$" where *A* and *B* are the sets of items, support of the rule can be defined by

$$\begin{aligned}\text{support}(A \Rightarrow B) &= \text{support}(A \cup B) \\ &= \frac{\#tuples(A \text{ and } B)}{\#tuples(all\_data)}, \qquad (3)\end{aligned}$$

where #*tuples*(*all_data*) is the number of all tuples in the relevant data tuples (or transactions).

For example, a support 30% for the association rule (e.g., Rule-1) means that 30% of all patients in the all data medical records are infected both bronchitis and lung cancer. From (3), it can be followed $\text{support}(A \Rightarrow B) = \text{support}(B \Rightarrow A)$. Also, (2) can be calculated by

$$\text{confidence}(A \Rightarrow B) = \frac{\text{support}(A \cup B)}{\text{support}(A)}, \qquad (4)$$

Correlation factor is another kind of measures to evaluate correlation between A and B. Simply, correlation factor can be calculated by:

$$\begin{aligned}\text{correlation}(A \Rightarrow B) &= \text{correlation}(B \Rightarrow A) \\ &= \frac{\text{support}(A \cup B)}{\text{support}(A) \times \text{support}(B)}, \qquad (5)\end{aligned}$$

Itemset A and B are dependent (positively correlated) iff $\text{correlation}(A \Rightarrow B) > 1$. If the correlation is equal to 1, then *A* and *B* are independent (no correlation). Otherwise, A and B are negatively correlated if the resulting value of correlation is less than 1.

A data mining system has the potential to generate a huge number of rules in which not all of the rules are interesting. Here, there are several objective measures of rule interestingness. Three of them are measure of rule support, measure of rule confidence and measure of correlation. In general, each interestingness measure is associated with a threshold, which may be controlled by the user. For example, rules that do not satisfy a confidence threshold (*minimum confidence*) of, say 50% can be considered uninteresting. Rules below the threshold (*minimum support* as well as *minimum confidence*) likely reflect noise, exceptions, or minority cases and are probably of less value. We may only consider all rules that have positive correlation between its itemsets.

3. Denormalization Data

In general, the process of mining data for discovering association rules has to be started from a single table (relation) as a source of data representing relation among item data. Formally, a relational data table [12] *R* consists of a set of tuples, where t_i represents the *i*-th tuple and if there are *n* domain attributes *D*, then $t_i = (d_{i1}, d_{i2}, \cdots, d_{in})$. Here, d_{ij} is an atomic value of tuple t_i with the restriction to the domain D_j, where $d_{ij} \in D_j$. Formally, a relational data table *R* is defined as a subset of the set of cross product $D_1 \times D_2 \times \cdots \times D_n$, where $D = \{D_1, D_2, \cdots, D_n\}$. Tuple *t* (with respect to *R*) is an element of *R*. In general, *R* can be shown in Table 1.

Tuples	D_1	D_2	$\cdots$	D_n
t_1	d_{11}	d_{12}	$\cdots$	d_{1n}
t_2	d_{21}	d_{22}	$\cdots$	d_{2n}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
t_r	d_{r1}	d_{r2}	$\cdots$	d_{rn}

Table 1. A Schema of Relational Data Table

A normalized database is assumed as a result of a process of normalization data in a certain context of data. The database may consist of several relational data tables in which they have relation one to each others. Their relation may be represented by Entities Relationship Diagram (ERD). Hence, suppose we need to process some domains (columns) data that are parts of different relational data tables, all of the involved tables have to be combined (joined) together providing a *general data table*. Since the process of joining tables

is an opposite process of normalization data by which the result of general data table is not a normalized table, simply the process is called *Denormalization*, and the general table is then called *denormalized table*. In the process of denormalization, it is not necessary that all domains (fields) of the all combined tables have to be included in the targeting table. Instead, the targeting denormalized table only consists of interesting domains data that are needed in the process of mining rules. The process of denormalization can be performed based on two kinds of data relation as follows.

3.1. Metadata of the Normalized Database

Information of relational tables can be stored in a metadata. Simply, a metadata can be stored and represented by a table. Metadata can be constructed using the information of relational data as given in Entity Relationship Diagram (ERD). For instance, given a symbolic ERD physical design is arbitrarily shown in Figure 2. From the example, it is clearly seen that there are four tables: **A**, **B**, **C** and **D**. Here, all tables are assumed to be independent for they have their own primary keys. Cardinality of relationship between Table **A** and **C** is supposed to be one to many relationships. It is similar to relationship between Table **A** and **B** as well as Table **B** and **D**.

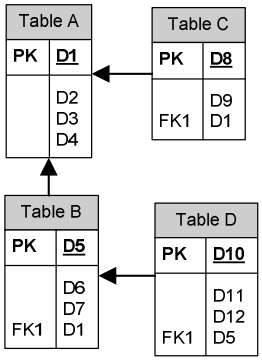

Figure 2. Example of ERD Physical Design

Table **A** consists of four domains/fields, D1, D2, D3 and D4; Table **B** also consists of four domains/fields, D1, D5, D6 and D7; Table **C** consists of three domains/fields, D1, D8 and D9; Table **D** consists of four domains/fields, D10, D11, D12 and D5. Therefore, there are totally 12 domains data as given by D={D1, D2, D3, …, D11, D12}. Relationship between **A** and **B** is conducted by domain D1. Table **A** and **C** is also connected by domain D1. On the other hand, relationship between **B** and **D** is conducted by D5. Relation among **A**, **B**, **C** and **D** can be also represented by graph as shown in Figure 3.

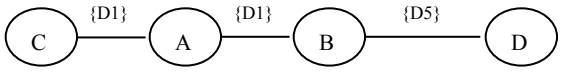

Figure 3. Graph Relation of Entities

Metadata expressing relation among four tables as given in the example can be simply seen in Table 2.

Table-1	Table-2	Relations
Table **A**	Table **B**	{D1}
Table **A**	Table **C**	{D1}
Table **B**	Table **D**	{D5}

Table 2. Example of Metadata

Through the metadata as given in the example, we may construct six possibilities of denormalized table as shown in Table 3.3.

No.	Denormalized Table
1	**CA**(D1,D2,D3,D4,D8,D9); **CA**(D1,D2,D8,D9); **CA**(D1,D3,D4,D9), etc.
2	**CAB**(D1,D2,D3,D4,D8,D9,D5,D6,D7), **CAB**(D1,D2,D4,D9,D5,D7), etc.
3	**CABD**(D1,D2,D3,D4,D5,D6,D7,D8,D9, D10,D11,D12), etc.
4	**AB**(D1,D2,D3,D4,D5,D6,D7), etc.
5	**ABD**(D1,D2,D3,D4,D5,D6,D7,D10, D11,D12), etc.
6	**BD**(D5,D6,D7,D10,D11,D12), etc.

Table 3. Possibilities of Denormalized Tables

CA(D1,D2,D3,D4,D8,D9) means that Table **A** and **C** are joined together, and all their domains are participated as a result of joining process. It is not necessary to take all domains from all joined tables to be included in the result, e.g. **CA**(D1,D2,D8,D9), **CAB**(D1,D2,D4,D9,D5,D7) and so on. In this case, what domains included as a result of the process depends on what domains are needed in the process of mining rules. For D1, D8 and D5 are primary key of Table **A**. **C** and **B**, they are mandatory included in the result, Table **CAB**.

3.2. Table and Function Relation

It is possible for user to define a mathematical function (or table) relation for connecting two or more domains from two different tables in order to perform a

relationship between their entities. Generally, the data relationship function performs a mapping process from one or more domains from an entity to one or more domains from its partner entity. Hence, considering the number of domains involved in the process of mapping, it can be verified that there are four possibility relations of mapping.

Let $A(A_1, A_2, \cdots, A_n)$ and $B(B_1, B_2, \cdots, B_m)$ be two different entities (tables). Four possibilities of function *f* performing a mapping process are given by:

- One to one relationship
  $f : A_i \rightarrow B_k$
- One to many relationship
  $f : A_i \rightarrow B_{p_1} \times B_{p_2} \times \cdots \times B_{p_k}$
- Many to one relationship
  $f : A_{r_1} \times A_{r_2} \times \cdots \times A_{r_k} \rightarrow B_k$
- Many to many relationship
  $f : A_{r_1} \times A_{r_2} \times \cdots \times A_{r_k} \rightarrow B_{p_1} \times B_{p_2} \times \cdots \times B_{p_k}$

Obviously, there is no any requirement considering type and size of data between domains in **A** and domains in **B**. All connections, types and sizes of data are absolutely dependent on function *f*. Construction of denormalization data is then performed based on the defined function.

4. Multidimensional Association Rules

As explained in Section 1, association rules that involve two or more dimensions or predicates can be referred to as *multidimensional association rules*. Multidimensional rules with no repeated predicates are called *interdimension association rules* (e.g. Rule-2)[1]. On the other hand, multidimensional association rules with repeated predicates, which contain multiple occurrences of some predicates, are called *hybrid-dimension association rules*. The rules may be also considered as combination (hybridization) between intradimension association rules and interdimension association rules. Example of such rule are shown in Rule-3, the predicate *Dis* is repeated. Here, we may firstly be interested in mining multidimensional association rules with no repeated predicates or interdimension association rules. Hybrid-dimension association rules as an extended concept of multidimensional association rules will be discussed later in our next paper.

The interdimension association rules may be generated from a relational database or data warehouse with multiple attributes by which each attribute is associated with a predicate. To generate the multidimensional association rules, we introduce an alternative method for mining the rules by searching for the predicate sets. Conceptually, a multidimensional association rule, $A \Rightarrow B$ consists of *A* and *B* as two datasets, called premise and conclusion, respectively.

Formally, *A* is a dataset consisting of several distinct data, where each data value in *A* is taken from a distinct domain attribute in *D* as given by

$$A = \{a_j \mid a_j \in D_j, \text{ for some } j \in \mathrm{N}_n\},$$

where, $D_A \subseteq D$ is a set of domain attributes in which all data values of *A* come from.

Similarly,

$$B = \{b_j \mid b_j \in D_j, \text{ for some } j \in \mathrm{N}_n\},$$

where, $D_B \subseteq D$ is a set of domain attributes in which all data values of *B* come from.

For example, from Rule-2, it can be found that *A*={60, *yes*}, *B*={Lung Cancer}, D_A={*Age*, *Smk*} and D_B={*Dis*}.

Considering $A \Rightarrow B$ is an interdimension association rule, it can be proved that $|D_A| = |A|$, $|D_B| = |B|$ and $D_A \cap D_B = \varnothing$.

Support of *A* is then defined by:

$$\text{support}(A) = \frac{|\{t_i \mid d_{ij} = a_j, \forall a_j \in A\}|}{r}, \qquad (6)$$

where *r* is the number of records or tuples (see Table 1). Alternatively, *r* in (6) may be changed to $|QD(D_A)|$ by assuming that records or tuples, involved in the process of mining association rules are records in which data values of a certain set of domain attributes, D_A, are not null data. Hence, (6) can be also defined by:

$$\text{support}(A) = \frac{|\{t_i \mid d_{ij} = a_j, \forall a_j \in A\}|}{|QD(D_A)|}, \qquad (7)$$

where $QD(D_A)$, simply called *qualified data* of D_A, is defined as a set of record numbers (t_i) in which all data values of domain attributes in D_A are not null data. Formally, $QD(D_A)$ is defined as follows.

$$QD(D_A) = \{t_i \mid t_i(D_j) \neq null, \forall D_j \in D_A\}. \qquad (8)$$

Similarly,

$$\text{support}(B) = \frac{|\{t_i \mid d_{ij} = b_j, \forall b_j \in B\}|}{|QD(D_B)|}. \qquad (9)$$

As defined in (3), $\text{support}(A \Rightarrow B)$ is given by

$$\text{support}(A \Rightarrow B) = \text{support}(A \cup B) = \frac{|\{t_i \mid d_{ij} = c_j, \forall c_j \in A \cup B\}|}{|QD(D_A \cup D_B)|} \quad (10)$$

$\text{confidence}(A \Rightarrow B)$ as a measure of certainty to assess the validity of $A \Rightarrow B$ is calculated by

$$\text{confidence}(A \Rightarrow B) = \frac{|\{t_i \mid d_{ij} = c_j, \forall c_j \in A \cup B\}|}{|\{t_i \mid d_{ij} = a_j, \forall a_j \in A\}|} \quad (11)$$

If support(A) is calculated by (6) and denominator of (10) is changed to r, clearly, (10) can be proved having relation as given by (4).

A and B in the previous discussion are datasets in which each element of A and B is an atomic crisp value. To provide a generalized multidimensional association rules, instead of an atomic crisp value, we may consider each element of the datasets to be a dataset of a certain domain attribute. Hence, A and B are sets of set of data values. For example, the rule may be represented by

Rule-5:

$$Age(X, \text{"20...60"}) \wedge Smk(X, \text{"yes"}) \Rightarrow Dis(X, \text{"bronchitis, lung cancer"}),$$

where A={{20…29}, {yes}} and B={{bronchitis, lung cancer}}.

Simply, let A be a generalized dataset. Formally, A is given by

$$A = \{A_j \mid A_j \subseteq D_j, \text{ for some } j \in \mathrm{N}_n\}.$$

Corresponding to (7), support of A is then defined by:

$$\text{support}(A) = \frac{|\{t_i \mid d_{ij} \in A_j, \forall A_j \in A\}|}{|QD(D_A)|}. \quad (12)$$

Similar to (10),

$$\text{support}(A \Rightarrow B) = \text{support}(A \cup B) = \frac{|\{t_i \mid d_{ij} \in C_j, \forall C_j \in A \cup B\}|}{|QD(D_A \cup D_B)|} \quad (13)$$

Finally, $\text{confidence}(A \Rightarrow B)$ is defined by

$$\text{confidence}(A \Rightarrow B) = \frac{|\{t_i \mid d_{ij} \in C_j, \forall C_j \in A \cup B\}|}{|\{t_i \mid d_{ij} \in A_j, \forall A_j \in A\}|} \quad (14)$$

To provide a more generalized multidimensional association rules, we may consider A and B as sets of fuzzy labels. Simply, A and B are called fuzzy datasets. Rule-4 is an example of such rules, where A={*young*, *yes*} and B={bronchitis}. A fuzzy dataset is a set of fuzzy data consisting of several distinct fuzzy labels, where each fuzzy label is represented by a fuzzy set on a certain domain attribute. Let A be a fuzzy dataset. Formally, A is given by

$$A = \{A_j \mid A_j \in \mathrm{F}(D_j), \text{ for some } j \in \mathrm{N}_n\},$$

where $\mathrm{F}(D_j)$ is a fuzzy power set of D_j, or in other words, A_j is a fuzzy set on D_j.

Corresponding to (7), support of A is then defined by:

$$\text{support}(A) = \frac{\sum_{i=1}^{r} \inf_{A_j \in A} \{\mu_{A_j}(d_{ij})\}}{|QD(D_A)|}. \quad (15)$$

Similar to (10),

$$\text{support}(A \Rightarrow B) = \text{support}(A \cup B) = \frac{\sum_{i=1}^{r} \inf_{C_j \in A \cup B} \{\mu_{C_j}(d_{ij})\}}{|QD(D_A \cup D_B)|} \quad (16)$$

$\text{Confidence}(A \Rightarrow B)$ is defined by

$$\text{confidence}(A \Rightarrow B) = \frac{\sum_{i=1}^{r} \inf_{C_j \in A \cup B} \{\mu_{C_j}(d_{ij})\}}{\sum_{i=1}^{r} \inf_{A_j \in A} \{\mu_{A_j}(d_{ij})\}} \quad (17)$$

Finally, $\text{correlation}(A \Rightarrow B)$ is defined by

$$\text{correlation}(A \Rightarrow B) = \frac{\sum_{i=1}^{r} \inf_{C_j \in A \cup B} \{\mu_{C_j}(d_{ij})\}}{\sum_{i=1}^{r} \inf_{A_j \in A} \{\mu_{A_j}(d_{ij})\} \times \inf_{B_k \in B} \{\mu_{B_k}(d_{ik})\}} \quad (18)$$

Similarly, if denominators of (15) and (16) are changed to r (the number of tuples), (17) can be proved also having relation as given by (4). Here, we may consider and prove that (16) and (17) are generalization of (13) and (14), respectively. On the other hand, (13) and (14)

are generalization of (10) and (11).

5. Illustrative Example

An illustrative example is given to understand well the concept of the proposed method and how to calculate support, confidence and correlation of the multidimensional fuzzy association rule is performed. The process is started from a given a simple medical records of patients as shown in Table 4.

Tuples	*Age*	*Smk*	*Dis*
t_1	20	yes	bronchitis
t_2	25	yes	bronchitis
t_3	22	yes	bronchitis
t_4	27	No	diarrhea
t_5	30	No	diarrhea
t_6	45	yes	lung cancer
t_7	40	yes	lung cancer
t_8	50	No	diabetes
t_9	60	yes	bronchitis
t_{10}	60	yes	lung cancer
t_{11}	*Null*	No	diarrhea

Table 4. Medical Records of Patients

Based on Table 4, support and confidence of Rule-2 are calculated using (10) and (11), respectively. Related to the conceptual form of the rule $A \Rightarrow B$, it can be followed that *A*={60, yes} and *B*={lung cancer}.

$$\text{support}(\text{Rule-2}) = \frac{|\{t_{10}\}|}{|\{t_1,...,t_{10}\}|} = 0.1,$$

where $QD(D_A \cup D_B) = \{t_1,...,t_{10}\}$. t_{11} is not included in $QD(D_A \cup D_B)$, because it has a null value in *Ages*. Confidence of Rule-2 is given by

$$\text{confidence}(\text{Rule-2}) = \frac{|\{t_{10}\}|}{|\{t_{10},t_9\}|} = 0.5.$$

Correlation of Rule-2 is given by

$$\text{correlation}(\text{Rule-2}) = \frac{\text{supp}(Rule-2)}{\text{supp}(\{60, yes\}) \times \text{supp}(\{bronchitis\})} = \frac{0.1}{0.2 \times 0.4} = 1.25$$

Note: supp(.) is defined as support(.), for short. Support, confidence and correlation of Rule-5 are calculated using (13) and (14) as follows.

$$\text{support}(\text{Rule-5}) = \frac{|\{t_1,t_2,t_3,t_6,t_7,t_9,t_{10}\}|}{|\{t_1,...,t_{10}\}|} = 0.7,$$

$$\text{confidence}(\text{Rule-5}) = \frac{|\{t_1,t_2,t_3,t_6,t_7,t_9,t_{10}\}|}{|\{t_1,t_2,t_3,t_6,t_7,t_9,t_{10}\}|} = 1.$$

$$\text{correlation}(\text{Rule-5}) = \frac{0.7}{0.7 \times 0.7} = 1.43$$

Tuples	$\mu_{yg}(ages)$ α	$\mu_{ys}(smk)$ β	$\mu_{br}(dis)$ γ	min(α,β,γ)
t_1	1	1	1	1
t_2	0.66	1	1	0.66
t_3	0.87	1	1	0.87
t_4	0.53	0	0	0
t_5	0.33	0	0	0
t_6	0	1	0	0
t_7	0	1	0	0
t_8	0	0	0	0
t_9	0	1	1	0
t_{10}	0	1	0	0
t_{11}	*null*	0	0	0
Σ	3.4	7	4	2.53

Table 5. Calculation of Fuzzy Values

Rule-4 is a fuzzy rule, where *A*={*young*, *yes*} and B={bronchitis}. *Young* (*yg*) is a fuzzy labels represented by a fuzzy sets as given in Section 1. Support of Rule-4 can be calculated by (16) as shown in Table 5. Therefore,

$$\text{support}(\text{Rule-4}) = \frac{2.53}{|\{t_1,...,t_{10}\}|} = 0.253$$

On the other hand, confidence and correlation of Rule-4 are given by

$$\text{confidence}(\text{Rule}-4) = \frac{2.53}{2.53} = 1.$$

$$\text{correlation}(\text{Rule}-4) = \frac{2.53}{2.53} = 1$$

Rule-1 and Rule-5 show positively correlated that means their conclusion and condition sides are not independent.

6. Conclusion

The paper firstly discussed a method of how to provide a denormalized table from a normalized database. Then, a concept of mining multidimensional fuzzy association rules was introduced. In general, multidimensional association rules consist of two types of rules, namely *interdimension association rules* and *hybrid-dimension association rules*. In this paper, we still restricted our proposed extended method to generate interdimension association rules. Three sets of equations were introduced to calculate support, confidence and correlation of three different kinds of generalized rules.

Acknowledgement

This work has been supported by the research grant of The Higher Education Directorate of Indonesia (Penelitian Hibah Bersaing) in the year of 2007.

References

[1] J. Han, M. Kamber, *Data Mining: Concepts and Techniques*, The Morgan Kaufmann Series, 2001.

[2] G. J. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, New Jersey: Prentice Hall, 1995.

[3] Rolly Intan, An Algorithm for Generating Single Dimensional Association Rules, Jurnal Informatika Vol. 7 No. 1 (Terakreditasi SK DIKTI No. 56/DIKTI/Kep/2005), May 2006.

[4] Rolly Intan, A Proposal of Fuzzy Multidimensional Association Rules, Jurnal Informatika Vol. 7 No. 2 (Terakreditasi SK DIKTI No. 56/DIKTI/Kep/2005), November 2006.

[5] Rolly Intan, 'A Proposal of an Algorithm for Generating Fuzzy Association Rule Mining in Market Basket Analysis', Proceeding of CIRAS (IEEE). Singapore, 2005

[6] Rolly Intan, 'Generating Multi Dimensional Association Rules Implying Fuzzy Valuse', The International Multi-Conference of Engineers and Computer Scientist 2006. Hong Kong.)

[7] O. P. Gunawan, *Perancangan dan Pembuatan Aplikasi Data Mining dengan Konsep Fuzzy c-Covering untuk Membantu Analisis Market Basket pada Swalayan X*, (in Indonesian) Final Project, 2004.

[8] L. A. Zadeh, "Fuzzy Sets and systems," *International Journal of General Systems*, Vol. 17, pp. 129-138, 1990.

[9] R. Agrawal, T. Imielimski, A.N. Swami, "Mining Association Rules between Sets of Items in Large Database", Proccedings of ACM SIGMOD International Conference Management of Data, ACM Press, pp. 207-216, 1993.

[10] R. Agrawal, R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases", Proccedings of 20th International Conference Very Large Databases, Morgan Kaufman, pp. 487-499, 1994.

[11] H. V. Pesiwarissa, *Perancangan dan Pembuatan Aplikasi Data Mining dalam Menganalisa Track Records Penyakit Pasien di DR.Haulussy Ambon Menggunakan Fuzzy Association Rule Mining*, (in Indonesian) Final Project, 2005.

[12] E.F. Codd, "A Relational Model of Data for Large Shared Data Bank", Communication of the ACM 13(6), pp. 377-387, 1970.