The 3rd International Conference on Mathematics and Statistics (ICoMS-3)
Institut Pertanian Bogor, Indonesia, 5-6 August 2008

DEVELOPING DATA MINING SYSTEM USING FUZZY ASSOCIATION RULES

[1]Imas S. Sitanggang, [2]Arsha Mustika, and [3]Aziz Kustiyo

[1,2,3]Computer Science Department, Bogor Agricultural University
Jl. Meranti, Wing 20 Level V, Kampus IPB Darmaga, Bogor 16680 – Indonesia

e-mail : [1] imasitanggang@yahoo.com, [2] Arsha.Mustika@mitrais.com, [3] azizboo@yahoo.com

Abstract. *This research aims to develop a data mining system in order to extract association rules from 24962 records of village potential data for the year 2003 (PODES 2003) especially in the regions of Java. The algorithm used in this research named Fuzzy Quantitative Association Rules Mining is divided into three major parts including transforming the data from the original format to fuzzy sets using the Fuzzy C-Means (FCM) algorithm, generating frequent itemsets, and extracting fuzzy association rules. The results of this research shows that a considerable number of rules have high fuzzy confidence value because the value of fuzzy support for antecedents combined with their consequent are also high. The parameter that gives significant influence is minimum fuzzy support (minsup). For minsup 90% and minimum fuzzy confidence (minconf) 90%, the system generates 16 fuzzy association rules. For lift value 1.04, there are two rules which show the relation of number of family that using electricity and number of permanent building. Besides, for mincorr value 0.8, there are five rules which show the relation of number of unemployment, number of students who dropped out from elementary school, number of family that use electricity, and number of permanent building.*

Keywords: *data mining, fuzzy association rules, Fuzzy C-Means, lift*

Introduction

Nowadays many computerized activities have resulted huge amounts of data in organizations. But the data usually stored in various storage without further process to extract more valuable information. This problem can be solved by applying data mining techniques. Data mining is a process to extract information and patterns from huge databases (Han & Kamber 2006). Association rule mining as one of methods in data mining is very useful to find items relationships in databases. Some techniques in association rule mining to handle quantitative attributes have been proposed before; one of them is partition method that finds association rules by partitioning attribute domain, combining the adjacent partitions and then changing it to binary. Kuok *et al.* (1998) stated that even though partition method can solve the problems resulted from indefinite domain, it causes vague boundary domain. The method also neglects elements located in partitions boundaries.

By applying fuzzy concept, association rules are more understandable. Beside, fuzzy sets can handle numeric data better than crisp methods because fuzzy sets smooth strict boundaries. An example of association rules using fuzzy concept is "10% old married people have several cars" (Gyenesei 2000).

This research aims to develop data mining system using fuzzy association rules mining in order to find relationship between items in village potential data for year 2003 especially in the regions of Java. The steps in Knowledge Discovery in Databases (KDD) are applied including data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation, knowledge representation. Fuzzy C-Means algorithm is used to construct fuzzy sets for fives attributes in the data. Then the clusters will be evaluated using Xie-Beni index.

Literature Review

Fuzzy Association Rules

Let $D = \{t_1, t_2, \ldots, t_n\}$ is a database and t_i is the i[th] record in D. $I = \{i_1, i_2, \ldots, i_m\}$ represents all attributes in D and i_j is the j[th] attribute. I is called itemset. Each attribute or item is related to several fuzzy sets. A fuzzy association rule has the following form (Kuok *et al.* 1998):

"IF $X = \{x_1, x_2, \ldots, x_p\}$ is $A = \{fx_1, fx_2, \ldots, fx_p\}$ THEN $Y = \{y_1, y_2, \ldots, y_q\}$ is $B = \{gy_1, gy_2, \ldots, gy_q\}$"

where $fx_i \in$ {a fuzzy set related to the attribute x_i}, $gy_j \in$ { a fuzzy set related to the attribute y_j}, and X and Y are itemset. X and Y are subset of I, they are disjoint which means X and Y do not have the same attribute. A and B contain fuzzy set associated with the corresponding attributes in X dan Y. The first part of an association rule, X is A, is called antecedent and the second part, Y is B, is called consequent of the rule. The meaning of the rule is when "X is A" is satisfied then we can imply that "Y is B" is also satisfied. The term 'satisfied' means that the amount of records which contributed in the attribute-fuzzy set pair is greater than a user specified threshold (Kuok *et al.* 1998).

Fuzzy Support

To generate association rules, the first step is to find frequent itemsets i.e. itemsets that have fuzzy support above minimum fuzzy support (minsup). Fuzzy support of itemset $\langle X,A \rangle$, in transactions set D is (Kuok *et al.* 1998):

$$FS_{\langle X,A \rangle} = \frac{\sum_{t_i \in D} \prod_{x_j \in X} \{\alpha_{a_j}(t_i[x_j])\}}{total(D)} \qquad (1)$$

where

$$\alpha_{a_j}(t_i[x_j]) = \begin{cases} m_{a_j \in A}(t_i[x_j]) & \text{if} \quad m_{a_j} \geq \omega \\ 0 & \text{otherwise} \end{cases}$$

$m_{a_j \in A}(t_i[x_j])$ is the membership degree of x_j in the i[th] record.

Contribution of each record is calculated using the membership degree of each x_j in the corresponding record. Membership degrees cannot be less that the specified minimum membership degree, ω. Therefore low membership degrees are neglected.

Fuzzy Confidence

Frequent itemsets are used to generate all possible association rules. All subset of a frequent itemset are also frequent. If a combination of antecedent and consequent has the high fuzzy confidence, then the rule is called interesting. The measure used to determine an interesting rule is minimum fuzzy confidence (minconf). The value of fuzzy confidence for an association rule is calculated using the following formula (Kuok et al. 1998):

$$FC_{\langle\langle X,A\rangle\langle Y,B\rangle\rangle} = \frac{\textit{fuzzy support of } \langle Z,C \rangle}{\textit{fuzzy support of } \langle X,A \rangle} = \frac{\sum_{t_i \in D} \prod_{z_k \in Z} \{\alpha_{a_k}(t_i[z_k])\}}{\sum_{t_i \in D} \prod_{x_j \in X} \{\alpha_{a_j}(t_i[x_j])\}} \qquad (2)$$

where

$$\alpha_{a_j}(t_i[x_j]) = \begin{cases} m_{a_j \in A}(t_i[x_j]) & \text{if} \quad m_{a_j} \geq \omega \\ 0 & \text{otherwise} \end{cases}$$

$m_{a_j \in A}(t_i[x_j])$ is membership degree of x_j in the i[th] record. $Z = X \cup Y$, $C = A \cup B$.

Fuzzy Correlation

In data mining systems, the semantic of association rule X $\Rightarrow$ Y is that X implies Y. It can be assumed that Y implies X due to the data distribution of X and Y. Thus, the formula for computing expectation values of the antecedent is changed to accommodate the meaning of fuzzy association rules. Below is the equation of fuzzy correlation (Kuok et al. 1998):

$$FCorr_{\langle\langle X,A\rangle\langle Y,B\rangle\rangle} = \frac{Cov(X,Y)}{\sqrt{Var(X).Var(Y)}} \qquad (3)$$

where

$$Cov(X,Y) = E[\langle Z,C\rangle] - E[\langle X,A\rangle]E'[\langle Y,B\rangle] \quad (4)$$

$$Z = X \cup Y, C = A \cup B \quad (5)$$

$$Var(X) = E[\langle X,A\rangle^2] - E[\langle X,A\rangle]^2 \quad (6)$$

$$Var(Y) = E[\langle Y,B\rangle^2] - E[\langle Y,B\rangle]^2 \quad (7)$$

$$E[\langle X,A\rangle] = \frac{\sum_{t_i \in D} \prod_{x_j \in X} \{\alpha_{a_j}(t_i[x_j])\}}{total(D)} \quad (8)$$

$$\alpha_{a_j}(t_i[x_j]) = \begin{cases} m_{a_j \in A}(t_i[x_j]) & \text{if } m_{a_j} \geq \omega \\ 0 & \text{otherwise} \end{cases}$$

$$E'[\langle Y,B\rangle] = \frac{\sum_{t_i \in D} \prod_{y_k \in Y} \beta[t_i]}{total(D)} \quad (9)$$

$$\beta[t_i] = \begin{cases} \prod_{y_k \in Y} \{\alpha_{a_j}(t_i[x_j])\} & \text{if } \gamma \geq \omega \\ 0 & \text{otherwise} \end{cases}$$

$$\gamma = \prod_{x_j \in X} \{\alpha_{a_j}(t_i[y_k])\}$$

In the equation (8), the calculation of E[⟨X,A⟩] is similar to an ordinary expectation value as we know in statistic except in this formula the threshold ω is involved. If the product of membership degree of the antecedent of a record is less than ω, then contribution of the consequent of that record will be zero.

Fuzzy correlation values are in the range [-1,1]. If the value is positive, then we can say that antecedent and consequent of an association rule is related. The higher the value of fuzzy correlation is, the more related the antecedent and consequent of a rule. Thus, an association rule will hold if the fuzzy correlation of that rule is greater than 0 (Kuok et al. 1998). Threshold of fuzzy correlation called minimum fuzzy correlation (mincorr) can be specified by users.

Patterns Evaluation

Most association algorithms have potential for resulting association rules in large quantity. Therefore some criteria to evaluate the quality of resulted rules are needed. The values of support and confidence are used to eliminate uninteresting patterns. In some cases the high value of confidence can be misleading, because we probably consider that a rule A ⇒ B is interesing eventhough in fact the occurence of A does not imply the occurance of B. This case can happen because the formula of confidence ignores the support of itemset in consequent part of a rule.

There are some objective measures to evaluate how interesting the rules are. One of the measures is Lift which is the ratio between confidence and support of consequent. Lift of a association rule is calculated by applying the following formula (Tan *et al.* 2006):

$$Lift = \frac{c(A \to B)}{s(B)} \quad (10)$$

where A is antecedent of an association rule, B is consequent of an association rule, $c(A\to B)$ is confidence of an association rule, and $s(B)$ is support of consequent of an association rule. If the value of lift of a rule equals to 1, it means that the antecendent and the consequent of that rule are independent each others. If the value of lift is greater than 1 then the antecendent and the consequent is positive corellated. On the other hand if the value of lift is less than 1 then the antecendent and the consequent is negative corellated.

Fuzzy C-Means (FCM)

Fuzzy C-Means (FCM) is a widely used algorithm in clustering based on fuzzy logic. FCM finds out center of clusters (centroids) and then it calculates the membership degree of each object in each cluster. This algorithm will iteratively minimize an objective function. The following equation in the form of objective function in FCM (Wang 1997):

$$J_m(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m \|x_k - v_i\|^2 \quad (11)$$

where

$x_1, x_2, \ldots, x_n$

are n vectors of data sample.

U = $[u_{ik}]$ is a $c \times n$ matrix, u_{ik} is the membership degree of kth vector in ith cluster that fill the following conditions:

$$0 \le u_{ik} \le 1, \quad i=1,2,...,c; \quad k=1,2,...,n$$

$$\sum_{i=1}^{c} u_{ik} = 1, \quad k=1,2,...,n$$

$$0 < \sum_{k=1}^{n} u_{ik} < n, \quad i=1,2,...,c$$

$V = \{v_1, v_2,...,v_c\}$ is centers of clusters.

$m \in (1,\infty)$ is weighting constant.

The goal of FCM is to determine U and V such that the objective function $J_m(U,V)$ is minimum. This function is sum of squared Euclidean distance between each sample and the corresponding center of cluster in which the distance is weighted by the membership degree. Center of clusters and membership degrees are iteratively computed using the formula (12) and (13) respectively (Wang 1997):

$$v_i = \frac{\sum_{k=1}^{n} (u_{ik})^m x_k}{\sum_{k=1}^{n} (u_{ik})^m} \qquad (12)$$

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\left\| \frac{x_k - v_i}{x_k - v_j} \right\| \right)^{\frac{2}{m-1}}} \qquad (13)$$

where i = 1, 2, ..., c; k = 1, 2, ..., n.

Validity of Fuzzy Clustering

Validity of fuzzy clustering is calculated to obtain a clustering scheme in which most vectors of data set have high membership degrees in a cluster. In fuzzy clustering, a matrix $U = |u_{ij}|$ is defined where u_{ij} is the membership degree of a vector x_i in a cluster j. Some methods of validity of fuzzy clustering, for examples: the partition coefficient and the partition entropy coefficient, only use the membership degree u_{ij}. Other categories, for example: Xie-Beni Index, include both the matrix U and the object (Halkidi *et al.* 2002). Below is the formula of partition coefficient (PC) (Halkidi *et al.* 2002):

$$PC = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{n_c} u_{ij}^2 \qquad (14)$$

The range value of PC index is $[1/n_c, 1]$, where n_c is the number of clusters. The closer the value of PC index to 1, the crisper the clustering is. If the value of PC index is 1 then U consists of only 0 and 1. In the other word, the clustering is non-fuzzy. If all value of membership degree in fuzzy partition are equal, then $u_{ij} = 1/n_c$, and the value of PC index is the lowest. The closer the value of PC index to $1/n_c$, the fuzzier the clustering is. The higher the value of PC index, the more effective the partition of data set is (Halkidi et al. 2002).

According to Xie and Beni (1991), the average of compactness and separation of a fuzzy c-partition is measured using S Index as the validity function of fuzzy clustering. S is defined as ratio between compactness π and separation s.

$$S = \frac{\pi}{s} = \frac{(\sigma/n)}{(d_{\min})^2} = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^2 \|V_i - X_j\|^2}{n \min_{i,j} \|V_i - V_j\|^2} \qquad (15)$$

where σ is the total of variance of data set X, n is the number of object in data set X, c is the number of cluster, V_i is the ith center of cluster, μ_{ij} is membership degree of the jth object in the ith cluster. The lower the value of S is, the more compact the resulted clusters are and the more separated one cluster to others is.

Method

Some following steps in Knowledge Discovery in Database (KDD) are implemented in this research:

1. Data transformation
   In this step, the data PODES 2003 in .sd2 format are transformed to .mdb format using a software STATransfer for further processing.
2. Data selection
   24962 records related to Java Island are selected from 65536 records. There are five numeric attributes selected from 750 attributes in the data. They are in the following table:

Table 1 List of selected attributes

Attribute code	Name of attribute
V403A1	number of pre-wealthy I family
V406	number of unemployment
V501B1	number of family that use electricity
V507A	number of permanent building
V603	number of students who dropped out from the elementary school

3. Data Mining
   Fuzzy Quantitative Association Rules Mining algorithm proposed by Gyenesei (2000) is implemented in this step. Figure 1 shows the pseudo code of the algorithm. Below are the parameters used in this algorithm:

Input : a database D, three threshold values: minsup, mincof and mincorr
Output : a list of interesting association rules
Notations :

D: the database; D_T: the transformed database; F_k : set of frequent k-itemset (have k items); C_k: set of candidate k-itemset (have k items); I: complete itemset, minsup: support threshold, minconf: confidence threshold

```
Algorithm (minsup,minconf,D)
I = Search(D);
(C1,DT) = Transform(D,I);
k = 1;
(Ck,Fk) = Checking(Ck,DT,minsup);
while (|Ck| ≠ Ø) do
begin
    inc(k);
    if k==2 then
        Ck = Join1(Ck-1);
    else Ck = Join2(Ck-1);
    Ck = Prune(Ck);
    (Ck,Fk) = Checking(Ck,DT,minsup);
    F = F ∪ Fk;
end
Rules(F,minconf);
```

Figure 1 Fuzzy Quantitative Association Rules Mining Algorithm (Gyenesei 2000)

The subroutines in the algorithm are outlined as follows:

i. *Search(D)* finds out and returns the complete itemset $I = \{i_1, i_2, \ldots, i_m\}$ from the database.
ii. *Transform(D,I)* generates a new transformed (fuzzy) database D_T from the original database. Then candidate 1-itemset C_1 will be generated from D_T. Membership functions for fuzzy sets are determined by applying FCM algorithm.
iii. *Checking*(C_k,D_T,*minsup*): In this step D_T is scanned and fuzzy support of candidate in C_k is counted. If fuzzy support is greater than or equal to minsup, then the candidate will be stored in C_k. At the same time, frequent itemset F_k will be generated from C_k.

iv. *Join1*(C_{k-1}): this routine generates C_2 from C_1 as follows:
    insert into C_2
    select $\langle X,A \rangle$, $\langle Y,B \rangle$,
    from $\langle X,A \rangle$, $\langle Y,B \rangle$ *in* C_1
    where $X \neq Y$

v. *Join2*(C_{k-1}) generates C_k from C_{k-1}.

vi. *Prune*(C_k): During this step, itemset will be pruned if a subset of candidate itemset in C_k does not exist in C_{k-1}.

vii. *Rules*(*F,minconf*) finds association rules from frequent itemset F.

4. Pattern evaluation

In this step the two objective interesting measures, lift and fuzzy correlation, are implemented to evaluate the rules.

Results and Disccusion

Fuzzy Sets

Each attributes in the database is transformed to fuzzy sets by applying FCM algorithm using MATLAB 6.5. Input parameters for FCM include the value of maximum iteration: 100, expected minimum *error*: 10^{-5}, and weighting power: 2. We choose the value 100 for maximum iteration because the 100th iteration we have the best clustering. Based on the computation of clustering validity using Xie-Beni Index, we have the best clustering in which the number of clusters for each attribute is 2. The original table which has five attributes is transformed into a new table which has 10 attributes. Each attribute in the original table is transformed into two fuzzy sets: low and high. Ten fuzzy sets resulted from five attributes are saved in a new table. Each cell in the new table contains the membership degree of an object in a fuzzy set. Table 2 shows minimum and maximum values of objects as well as number of members in clusters of each attribute.

Table 2 Minimum value, maximum value and number of members in clusters of each attribute

No	Itemset	Mini-mum	Maxi-mum	Number of member	No	Itemset	Mini-mum	Maxi-mum	Number of member
1	V403A1, high	670	9432	6095	6	V501B1, low	0	2260	23521
2	V403A1, low	0	669	18867	7	V507A, high	1756	67610	1682
3	V406, high	888	8831	1270	8	V507A, low	0	1754	23280
4	V406, low	0	887	23692	9	V603, high	173	998	517
5	V501B1, high	2261	12026	1441	10	V603, low	0	172	24445

Frequent Itemset Generating

Frequent itemsets are generated by calculating fuzzy support of each item, beginning from candidate 1-itemset to candidate n-itemset. Itemset with the value of fuzzy support above minsup is considered as frequent itemset, while others are removed. When reaching the candidate 2-itemset or above, the pruning step is implemented before checking the fuzzy support. If a subset of candidate k-itemset does not exist in candidate (k-1)-itemset then the candidate k-itemset is pruned.

In this research, several experiments are conducted using several values of minsup in range [50% - 90%]. We use minimum threshold for membership degree (ω) 0.5 because the number of fuzzy sets of each attribute is 2. Therefore, an object is assigned to a certain fuzzy set if its membership degree is greater than or equal to 0.5. Figure 2 shows the number of frequent itemset for several values of minsup.

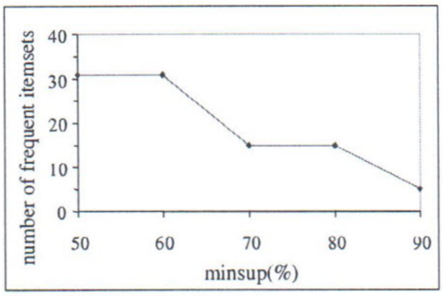

Figure 2 Number of frequent itemset for several values of minsup.

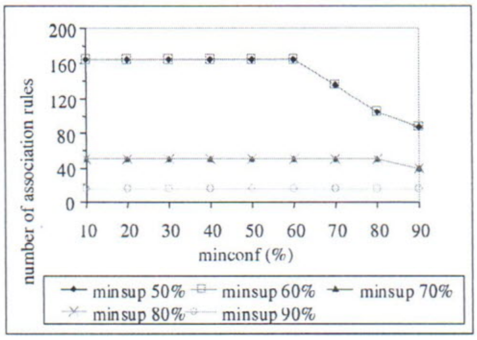

Figure 3 Number of rules for several values of minsup and minconf.

e as tl … of frequent itemset from frequent 1-itemset to frequent 3-itemset is still complete. When the minsup reach 90%, there is no frequent 3-itemset or above generated.

Association Rules Generating

All possible association rules are generated from frequent itemsets. Then we compute fuzzy confidence values of the generated rules. The rules with fuzzy confidence above minconf are kept, others are removed. The experiments use some values of fuzzy confidence in the interval [10% - 90%] and some values of fuzzy support in the interval [50% - 90%]. In general, for all experiments the number of association rules tends to decline when minsup is increased as shown in Figure 3. But the number of rules for minsup 50% is the same as it for minsup 60% for all values of mincof. The same case occurs for the value of minsup 70% and 80%. For that two pairs of minsup values, the number of rules is equal because the number of frequent itemsets for each pair is also the same (Figure 2). Figure 3 shows that the number of generated rules is steady for mincof values 10% to 60%. And then it tends to decrease for mincof above 60% except for minsup 90%, where the number of rules is constant. It is because most of generated frequent itemset has high fuzzy confidence value. Some rules and the correspond values of fuzzy confidence, fuzzy correlation and lift for minsup 90% and minconf 90% are in Table 3.

Table 3 Association rules for minsup 90% and minconf 90%

No	Antecedent	Consequent	Fuzzy confidence	Fuzzy correlation	Lift
1	V406, low	V501B1, low	0.933188885	0.649607647	1.020615986
2	V406, low	V507A, low	0.918503544	0.582314345	1.019293943
3	V406, low	V603, low	0.974717565	0.814934493	1.005390509
4	V501B1, low	V406, low	0.947369898	0.727801383	1.020615986
5	V501B1, low	V507A, low	0.951385864	0.835040786	1.055784547
6	V501B1, low	V603, low	0.973550552	0.817622622	1.004186772
7	V507A, low	V406, low	0.946142734	0.741948026	1.019293943
8	V507A, low	V501B1, low	0.965344869	0.892270689	1.055784547
9	V507A, low	V603, low	0.97356811	0.834998086	1.004204883
10	V603, low	V406, low	0.933237101	0.336352768	1.005390509
11	V603, low	V501B1, low	0.918167016	0.270054863	1.004186772
12	V603, low	V507A, low	0.904906529	0.198036778	1.004204883
13	V501B1, low	V406, low; V603, low	0.924924843	0.648426361	1.022281349
14	V507A, low	V406, low; V603, low	0.923870028	0.668586224	1.021115505
15	V406, low; V603, low	V501B1, low	0.934711591	0.717740444	1.022281349
16	V406, low; V603, low	V507A, low	0.920144986	0.663107455	1.021115505

Pattern Evaluation

The rules are evaluated using both lift and fuzzy correlation. Figure 4 illustrates some values of lift and the number of rules for varied minsup values and mincof 90%. Beside using lift, we can use fuzzy correlation to evaluate association rules. In Figure 5 some values of fuzzy correlation (mincorr) from 0.1 to 0.9 are applied to rules with varied minsup values and mincof 90%.

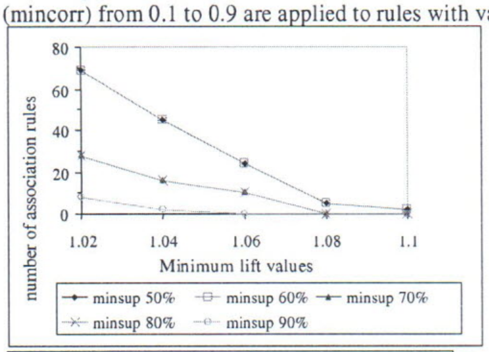

Figure 4 Some values of lift and the number of rules for some minsup values and mincof 90%.

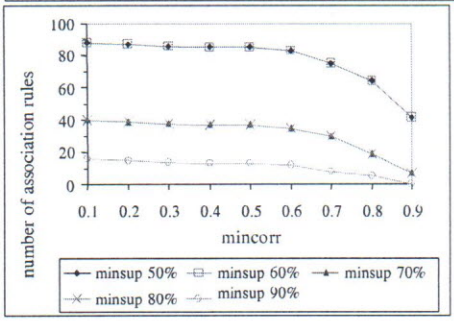

Figure 5 Some values of fuzzy correlation and the number of rules for some minsup values and mincof 90%.

For the highest values of minsup, minconf and lift that are 90%, 90% and 1.04 respectively, we have two association rules: 1) If number of family that uses electricity is low then number of permanent building is low; 2) If number of permanent building is low then number of family that uses electricity is low. We have different rules if the thresholds used for minsup, minconf, and mincorr are 90%, 90% and 0.8 respectively. The rules are:

- If number of unemployment is low then number of students who dropped out from elementary school is low
- If number of family that use electricity is low then number of permanent building is low
- If number of family that use electricity is low then number of students who dropped out from elementary school is low
- If number of permanent building is low then number of family that use electricity is low
- If number of permanent building is low then number of students who dropped out from elementary school is low.

Conclusion

In this paper we present the results of the application of fuzzy quantitative association rules mining algorithm to village potential data for the year 2003 (PODES 2003) especially in the regions of Java. We use varied values of minimum confidence and minimum support in several experiments. The measures lift and fuzzy correlation are applied to evaluate the rules. Most rules have high fuzzy confidence because fuzzy support values for the combinations of the antecedent and the consequent are high. The parameter that gives the most significant influence on the number of generated association rules is minimum support (minsup). For minsup 90%, minimum confidence (minconf) 90% and lift 1.04, there are four rules related to two attributes: number of family that use electricity and number of permanent building. Besides, for mincorr value 0.8, there are five rules which show the relation of number of unemployment, number of students who dropped out from elementary school, number of family that use electricity, and number of permanent building.

References

Gyenesei A. 2000. A Fuzzy Approach for Mining Quantitative Association Rules. Turku Centre for Computer Science [TUCS] Technical Reports No. 336. http://citeseer.ist.psu.edu/cache/papers/cs/19318/http:zSzzSzwww.tucs.fizSzpublicationszSztechreportszSzTR336.pdf/gyenesei00fuzzy.pdf [12 Desember 2005].

Halkidi M, Batistakis Y, Vazirgiannis M. Clustering Validity Checking Methods: Part II. www.db-net.aueb.gr/mhalk/ papers/PARTII_clvalidity.pdf [31 Mei 2006]

Han J, Kamber M. 2006. *Data Mining: Concepts and Techniques*. San Diego, USA: Morgan-Kaufmann.

Kuok CM, Fu A, Wong MH. 1998. Mining Fuzzy Association Rules in Databases. SIGMOD Record, 27(1):41-46. www.cs.wpi.edu/~ifc/disc/disc99/disc/record/issues/9803/kuok.pdf [29 Mei 2005].

Tan PN, Steinbach M, Kumar V. *Introduction to Data Mining*. 2006. Boston, USA: Pearson Educations.

Wang LX. 1997. *A Course in Fuzzy Systems and Control*. New Jersey, USA: Prentice-Hall International, Inc.

Xie XL, Beni G. 1991. *A Validity Measure for Fuzzy Clustering*. IEEE Trans on Pattern Anal Machine Intell, 13(8):841-847.

ISBN 979-19256-0-0

9 789791 925600