WCCI 2012 IEEE World Congress on Computational Intelligence
June, 10-15, 2012 - Brisbane, Australia FUZZ IEEE

Mining Fuzzy Coherent Rules from Quantitative Transactions Without Minimum Support Threshold

Chun-Hao Chen, Ai-Fang Li
Department of Computer Science and Information Engineering
Tamkang University
Taipei, Taiwan, R.O.C.
chchen@mail.tku.edu.tw,
697410172@s97.tku.edu.tw

Yeong-Chyi Lee
Department of Information Management
Cheng Shiu University
Kaohsiung, Taiwan, R. O. C.
yeongchyi@csu.edu.tw

Tzung-Pei Hong
Department of Science and Information Engineering,
National University of Kaohsiung
Kaohsiung, 811, Taiwan, R.O.C.
tphong@nuk.edu.tw

Abstract—**Many fuzzy data mining approaches have been proposed for finding fuzzy association rules with the predefined minimum support from the give quantitative transactions. However, some comment problems of those approaches are that (1) a minimum support should be predefined, and it is hard to set the appropriate one, and (2) the derived rules usually expose common-sense knowledge which may not be interested in business point of view. In this paper, we thus proposed an algorithm for mining fuzzy coherent rules to overcome those problems with the properties of propositional logic. It first transforms quantitative transactions into fuzzy sets. Then, those generated fuzzy sets are collected to generate candidate fuzzy coherent rules. Finally, contingency tables are calculated and used for checking those candidate fuzzy coherent rules satisfy four criteria or not. Experiments on the foodmart dataset are also made to show the effectiveness of the proposed algorithm.**

Keywords- fuzzy set, fuzzy association rules, fuzzy coherent rules, membership function, data mining.

I. INTRODUCTION

Data mining is most commonly used in attempts to derive useful information and extract useful patterns from large data sets or database for solving specific issue. One of the commonly used techniques is association rule mining which is an expression X→Y, where X and Y are a set of items [1]. It means in the set of transactions, if all the items in X exist in a transaction, then Y is also in the transaction with a high probability. For example, assume whenever customers in a supermarket buy bread and butter, they will also buy milk. From the transactions kept in the supermarkets, an association rule such as "Bread and Butter→Milk" will be mined out.

Lots of mining approaches are thus proposed for association rule mining [1, 2, 5], and most of them focused on binary valued transaction data. However, transaction data in real-world applications usually consist of quantitative values. Thus, by combing fuzzy theory, many mining algorithms have been proposed for deriving fuzzy rules from quantitative transaction database [3, 4, 10, 11, 12, 13, 14, 16, 17, 21, 22, 23, 25, 28].

However, there are two common problems of those fuzzy association rule mining approaches. The first one is that they all need to set parameters, namely minimum support and minimum confidence, which are hard to set. If a large minimum support value is set, only few frequent itemsets will be generated. It also means that lots of potential rules may not be mine out. On the contrary, large amount of rules will be derived such that decision makers can not use them easily to make marketing strategy when the minimum support value is set at a small value. The second problem is that some of those derived rules only expose common-sense knowledge and may not be interesting in business point of view. For example, if we derive a rule "If milk is bought, Then bread is bought," with high support and conference, it is a reliable rule according to Apriori algorithm [1]. But, it may not be valuable for business since the derived rule is common-sense knowledge.

Recently, Longbing Cao suggested the domain-driven data mining concept (D3M) [7, 8, 6], and cooperated it with industry knowledge to mine actual and useful information. Under the D3M concept, for those association rule mining algorithms on binary transaction [24], Sim et al. proposed a logical-based approach for deriving coherent rules. In that approach, by using the properties of propositional logic, relationship between items (also namely coherent rules) can be derived directly without knowing the appropriate value of minimum support.

In this paper, we thus propose an algorithm for mining fuzzy coherent rules by using the properties of propositional logic. The proposed approach first transforms quantitative transactions into fuzzy sets by utilizing the predefined membership functions. Candidate fuzzy coherent rules are then formed from the transformed fuzzy regions. For each candidate fuzzy coherent rule, the contingency table is then calculated according to the antecedent and consequent parts of that rule. At last, four criteria are used to judge the candidate fuzzy coherent rules. If it satisfies the conditions, it is then a fuzzy coherent rule. Experimental results on a foodmart dataset are made to show the effectiveness of the proposed approach.

II. RELATED WORK

In this section, related work will be introduced. In section II.A, the rule mining approaches, including binary and fuzzy data mining algorithms are described. The main issue of the minimum support threshold is stated in section II.B.

A. Binary and Fuzzy Data Mining Approaches

Data mining aims to extract useful knowledge and patterns from existing data to solve a specific issue. To date, it has been used in many different fields, such as shopping cart analysis [2], network intrusions [25], and stock market analysis [3]. The association rule which is the commonly used in shopping cart analysis, is represented by A B, where A and B are the products, and the rule expresses that if product A is purchased, product B will be purchased together with it. Two measurements are used to measure the validity of one association rule, which are support and confidence. The earliest association rule mining was suggested by Agrawal et al [1], and the main three steps can be divided into: (1) produce candidate itemsets; (2) produce frequent itemsets based on minimum support; and (3) produce frequent itemsets based on minimum confidence.

In real-world application, since transactions always have quantitative values, thus how to handle the quantitative values becomes an interesting issue. Thus, by utilizing fuzzy theory, lots of mining algorithms have been proposed for deriving fuzzy rules from quantitative transaction database. We can thus divide fuzzy data mining approaches into two kinds, namely *single-minimum-support fuzzy-mining* (*SSFM*) [3, 4, 10, 11, 12, 13,14, 21, 22, 23, 25, 28] and *multiple-minimum-support fuzzy-mining* (*MSFM*) [16, 17] approaches.

For SSFM approaches, Chan et al. proposed the F-APACS algorithm to mine fuzzy association rules [4]. They first transformed quantitative attribute values into linguistic terms and then used adjusted difference analysis to find interesting associations among the attributes. Kuok et al. proposed a fuzzy mining approach to handle numerical data in databases and to derive fuzzy association rules [14]. At nearly the same time, Hong et al. proposed a fuzzy mining algorithm to mine fuzzy rules from quantitative transaction data [11]. Basically, these fuzzy mining algorithms first used membership functions to transform each quantitative value into a fuzzy set using linguistic terms and then used a fuzzy mining process to find fuzzy association rules. Yue et al. then extended the above concept to find fuzzy association rules with weighted items from transaction data [28]. They adopted Kohonen self-organized mapping to derive fuzzy sets for numerical attributes.

Regarding mining approaches for MSFM problems, Lee et al. proposed a mining algorithm that used multiple minimum supports to mine fuzzy association rules [17]. They assumed that items had different minimum supports and the minimum support for an itemset was set was the maximum of the minimum supports of items contained in the itemset. Under this constraint, the characteristics of level-by-level processing were kept, such that the original Apriori algorithm could easily be extended to find large itemsets. In [16], Lee et al. further extended the existing approach [17] and proposed a new fuzzy association rule mining algorithm with taxonomy.

B. The Main Issue of the Minimum Support Threshold

The main issue of association rule mining approaches is how to define appropriate minimum support and minimum confidence, and the existed work has also reported that although the appropriate minimum support maybe exist, it is hard to find it [26]. As to how to find the appropriate minimum support, there are also lots of literatures have been published for this issue [24]. In general, using different minimum supports to derive association rules maybe result in different mining results. A small minimum support will generate too much frequent itemsets and association rules, which are not easily for user to make decisions. On the contrary, a larger minimum support will delete possible useful itemsets even they are infrequent and association rules [24]. In [18], Liu et al. thus proposed an algorithm that using multiple minimum supports, called Minimum Item Supports (MISs), for mining association rules. Lots of approaches have then been proposed approach for setting appropriate multiple minimum supports by using heuristics methods [15, 19, 27].

In order to solve those problems, Sim *et al.* thus proposed an association rule mining framework for association rule mining without minimum support threshold. In that approach, by using the properties of propositional logic, relationship between items can be derived directly without knowing the appropriate value of minimum support [24]. The main concept of that approach is that it maps the association rules to equivalences. And, each mapping from an association rule to an equivalence should satisfy conditions which are shown in TABLE I.

TABLE I: The four conditions for mapping rules to equivalence

Equivalences	p ≡ q	¬p ≡¬q
Association Rules	X → Y	¬X → ¬Y

True or False on Association Rules	Required Conditions	
T	X → Y	¬X → ¬Y
F	X → ¬Y	¬X → Y
F	¬X → Y	X → ¬Y
T	¬X → ¬Y	X → Y

From TABLE I, X and Y are two itemsets. It show that an association rule X→Y is mapped to p ≡ q, if and only if (1) X → Y is true; (2) ¬X →Y is false; (3)X → ¬Y is false; and (4) ¬X →¬Y is true. When used in multiple transactions, it can map association rules to implications as follows: X → Y is mapped to an implication p → q, if and only if (1) Sup(X, Y) > Sup(X, Y); (2) Sup(X, Y) > Sup(¬X, Y); (3)Sup(X, Y) > Sup(X, ¬Y); and (4) Sup(X, Y) > Sup(¬X, ¬Y). In the same way, others association rules that mapped to implications based on comparison between supports can be derived, and namely *pseudoimplications*. According to these *pseudoimplications*, Sim et al. then further defined them into a concept called coherent rules. That is the following four conditions must be satisfied for a coherent rule: (1) Sup(X, Y) > Sup(¬X, Y); (2) Sup(X, Y) > Sup(X, ¬Y); (3) Sup(¬X, ¬Y) > Sup(¬X, Y); and (4) Sup(¬X, ¬Y) > Sup(X, ¬Y). These four conditions can also be represented as the contingency table as shown in TABLE II.

By utilizing the coherent rules concept, in this paper, we thus attempt to propose a fuzzy data mining algorithm that cooperated with those four conditions into the mining process for deriving fuzzy coherent rules from quantitative transactions without minimum support threshold.

TABLE II: The contingency table of a rule

Frequency of co-occurrences		Consequence *Y*	
		Y	*¬Y*
Antecedent *X*	*X*	Q_1 = Sup(X,Y)	Q_2 = Sup(X, ¬Y)
	¬X	Q_3 = Sup(¬X,Y)	Q_4 = Sup(¬X, ¬Y)

III. THE PROPOSED FUZZY COHERENT RULE MINING ALGORITHM

Based on the fuzzy data mining algorithm [11] and the coherent rule concept [24] which is described in above section, the proposed algorithm for mining fuzzy coherent rules is described below.

The proposed fuzzy coherent rule mining algorithm:

INPUT: A body of n quantitative transaction data, a given itemset Y, and a given set of membership functions.

OUTPUT: A set of fuzzy coherent rules (*FCR*).

STEP 1: Transform the quantitative value $v_j^{(i)}$ of each transaction datum $D^{(i)}$, i=1 to n, for each item I_j, j=1 to m, into a fuzzy set $f_j^{(i)}$ represented as:

$$\left(\frac{f_{j_1}^{(i)}}{R_{j_1}} + \frac{f_{j_2}^{(i)}}{R_{j_2}} + \ldots + \frac{f_{j_l}^{(i)}}{R_{j_l}} \right)$$

using the given membership functions, where R_{jk} is the k-th fuzzy region of item I_j, $f_{jk}^{(i)}$ is $v_j^{(i)}$'s fuzzy membership value in region R_{jk}, and l (= $|I_j|$) is the number of fuzzy regions for I_j.

STEP 2: For each fuzzy region R_{jk}, calculate its complement value, and represented as follows:

$$\left(\frac{1 - f_{j_1}^{(i)}}{R_{j_1}} + \frac{1 - f_{j_2}^{(i)}}{R_{j_2}} + \ldots + \frac{1 - f_{j_l}^{(i)}}{R_{j_l}} \right).$$

STEP 3: Collect all fuzzy regions into the set A.

STEP 4: Collect all fuzzy regions of the given item y into the set L.

STEP 5: Remove the given item y's fuzzy region from the set A to from the set K.

STEP 6: Set h = 1, where h means the length of antecedent X and form candidate fuzzy coherent rule $X \rightarrow Y$, where X is an element of K, and Y is an element of L.

STEP 7: Do the following substeps to generate fuzzy coherent rules:

SUBSTEP 7.1: Calculate the contingency table for antecedent X and consequent Y. Here, four count values will be calculated, including Q_1:count$_{XY}$, Q_2:count$_{X\sim Y}$, Q_3:count$_{\sim XY}$ and Q_4:count$_{\sim X\sim Y}$. Each of them is calculated as follows:

$$count_S = \sum_{i=1}^{n} f_S^{(i)},$$

where the fuzzy value of an itemset S in each transaction is calculated as $f_S^{(i)} = f_{s_1}^{(i)} \Lambda f_{s_2}^{(i)} \Lambda \ldots \Lambda f_{s_m}^{(i)}$, and $f_{s_j}^{(i)}$ is the membership value of fuzzy item s_j in i-th transaction. If the minimum operator is used for the intersection, then:

$$f_S^{(i)} = \underset{j=1}{\overset{m}{Min}} f_{s_j}^{(i)}.$$

SUBSTEP 7.2: Check the candidate fuzzy coherent rule meets the four conditions, which are $Q_1>Q_2$, $Q_1>Q_3$, $Q_4>Q_2$ and $Q_4>Q_3$ or not. If yes, let $FCR_{all} = FCR_s \cup (X,Y)$, goto SUBSTEP 7.1 to calculate the contingency table of next candidate fuzzy coherent rule. If all candidate rules are checked, goto SUBSTEP 7.3. Otherwise, back to SUBSTEP 7.1.

SUBSTEP 7.3: Check the FCR_h is empty or not. If yes, goto STEP11. Otherwise, goto STEP 8.

STEP 8: Collect the fuzzy regions of antecedent and consequence parts of the derived fuzzy coherent rules in FCR_h to form new set K and L, respectively.

STEP 9: Set h=h+1.

STEP 10: Form candidate fuzzy coherent rule $X \rightarrow Y$ according to L and K, where the length of X is h. Note that fuzzy regions with the same item can not be used to form candidate fuzzy coherent rules.

STEP 11: If there is no candidate fuzzy coherent rule, goto STEP 12. Otherwise, goto SUBSTEP 7.1.

STEP 12: Output the derived fuzzy coherent rules FCR_{all}.

Note that the proposed approach can easily extend to mine all fuzzy coherent rules through repeating STEP 1 to STEP 12 using different itemset Y.

IV. EXPERIMENTAL RESULTS

In this section, experimental results of the proposed approach are described. They were implemented in Java on a personal computer with Intel Core i7, 2.93GHz and 4GB RAM. The dataset is described in section 5.1. The performance evolutions are given in section 5.2

A. *Dataset Descriptions*

A dataset, the foodmart database [20], was used to evaluate the performance of the algorithms under different comparisons. The foodmart database lay in the database product Microsoft SQL Server 2000. The related information of this dataset was listed as follows. There were 21,556 transactions; the total number of different items was 1,600. In the following experiments, 1000 transactions which are selected from 21,556 transactions are used to evaluate the proposed approach. The membership functions are given in Figure 1. There are three fuzzy set used in the experiments that are Low, Middle and High.

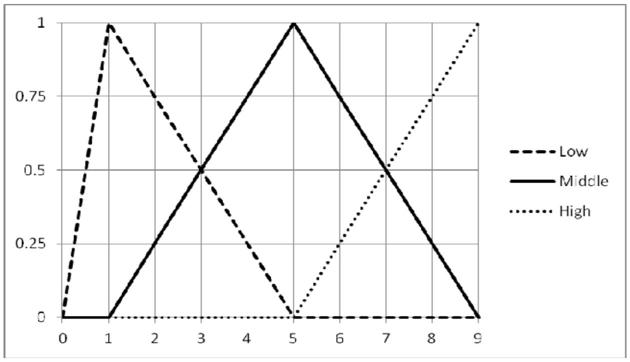

Figure 1. The membership function used in this example

B. Experimental Evaluations

Firstly, experiments were made to show the comparison results of the derived rules between the proposed approach (FCR) and the original fuzzy rule mining approach (FAR) [5]. The minimum support and minimum confidence of FAR is set at 0.000066 and 0.1, respectively. The results are shown in TABLE III.

TABLE III. Comparison results between FCR and FAR

The length of antecedent part in a rule	FCR	FAR
Number of rules (Length=2)	158	7240
Number of rules (Length=3)	42	435
Number of rules (Length=4)	8	0
Total number of rules	208	7675
The average confidence	0.870928	0.393698
The average support	0.000800	0.000417

From Table III, we can observe that the number of rules derived by FCR is 158 which is less than that by FCR when the length of a rule is 2. However, when the length of a rule is 4, we can find that the number of derived rules by FCR is larger than that by FAR. The proposed approach can derive extra eight rules than FAR. In addition, the average support and confidence of rules derived by the proposed approach is 0.0008 and 0.87, which are also larger than that by FAR. Those results show that the proposed approach is effective in finding more reliable rules. In order to explain the merits of the proposed approach more clearly, a derived fuzzy coherent rule with rule length equals 4 is given as follows:

"If 332L, 62M, and 671M, Then 349M,
sup = 0.00075, conf. = 1.0"

From the rule, we can know that the confidence value is 100%, although the support value of that rule is very small, it may be useful in terms of business. However, the rule will be pruned by using FAR if the minimum support is larger than 0.00075. From those experimental results, we thus can conclude that the proposed approach provides an interesting way to find association rules without minimum support threshold.

V. Conclusion And Future Works

In this paper, we have proposed a fuzzy coherent rule mining algorithm from quantitative transactions without minimum support threshold. The proposed algorithm first transforms quantitative transactions into fuzzy sets by the predefined membership functions. Then, the fuzzy sets are collected to form candidate fuzzy coherent rules. The contingency table is then calculated for each candidate fuzzy coherent rule, and used to cheek whether it satisfies the four conditions or not. Experiments on the foodmart dataset have also been made to show the merits of the proposed approach. Firstly, they show that the proposed approach can derive more rules than the original fuzzy rule mining approach in terms of rule length. Secondly, they also show that the proposed approach can derive useful rules in terms of high confidence rules. The main contribution of this work is that the proposed approach can derive interesting rules effectively without setting minimum support. In the future, we will continue to enhance the proposed approach to more complexity problems.

Acknowledgment

This research was supported by the National Science Council of the Republic of China under contract NSC 100-2221-E-032 -065 -.

References

[1] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules". *International Conference on Very Large Data Bases*, pp. 487-499, 1994.

[2] R. Agrawal, T. Imielinksi and A. Swami, "Mining association rules between sets of items in large database," *The 1993 ACM SIGMOD Conference,* Washington DC, USA, 1993.

[3] W. H. Au and K.C.C. Chan, "Mining fuzzy association rules in a bank-account database," *IEEE Transactions on Fuzzy Systems*, Vol. 11, No 2, pp. 238 - 248, 2003.

[4] K.C.C. Chan and W.H. Au, "An Effective Algorithm for Discovering Fuzzy Rules in Relational Databases," *The IEEE International Conference on Fuzzy Systems*, Vol.2, p.1314-1319, 1998.

[5] C. H. Cai, W. C. Fu, C. H. Cheng and W. W. Kwong, "Mining association rules with weighted items," *The International Database Engineering and Applications Symposium*, pp. 68-77, 1998.

[6] L. Cao, Y. Zhao, H. Zhang, D. Luo, C. Zhang and E. K. Park, "Flexible Frameworks for Actionable Knowledge Discovery," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 22 , No. 9, pp.1299-1312, 2010.

[7] L. Cao, "Domain Driven Data Mining (D3M)," *IEEE International Conference on Data Mining Workshops*(*ICDMW'08*), pp.74-76, 2008.

[8] L. Cao, "Domain-Driven Data Mining: Challenges and Prospects," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 22 , No. 6, pp.755-769, 2010.

[9] M. Delgado, N. Marin and D. Sanchez, M. A. Vila, "Fuzzy association rules: general model and applications," *IEEE Transactions on Fuzzy Systems*, Vol. 11, No 2, pp. 214-225, 2003.

[10] D. Dubois, H. Prade and T. Sudkamp, "On the representation, measurement, and discovery of fuzzy associations," *IEEE Transactions on Fuzzy Systems*, Vol. 13, No 2, pp. 250 - 262, 2005.

[11] T. P. Hong, C. S. Kuo and S. C. Chi, "Mining association rules from quantitative data", *Intelligent Data Analysis*, Vol. 3, No. 5, pp. 363-376, 1999.

[12] T. P. Hong and C. Y. Lee, "Induction of fuzzy rules and membership functions from training examples," *Fuzzy Sets and Systems,* Vol. 84, pp. 33-47, 1996.

[13] T. P. Hong, K. Y. Lin and B. C. Chien, "Mining Fuzzy Multiple-Level Association Rules from Quantitative Data," *Applied Intelligence*, Vol. 18, No. 1, pp. 79-90, 2003.

[14] C. M. Kuok, A. W. C. Fu and M. H. Wong, "Mining fuzzy association rules in databases," *ACM SIGMOD Record*, Vol. 27, No. 1, pp. 41–46, 1998.

[15] Y.S. Koh, N. Rountree and R.A. O'Keefe, "Finding Non-Coincidental Sporadic Rules Using Apriori-Inverse," *International Journal of Data Warehousing and Mining*, vol. 2, pp. 38-54, 2006.

[16] Y. C. Lee, T. P. Hong and T. C. Wang, "Multi-level fuzzy mining with multiple minimum supports," *Expert Systems with Applications*, Vol. 34, No. 1, pp. 459-468, 2008.

[17] Y. C. Lee, T. P. Hong and W. Y. Lin, "Mining fuzzy association rules with multiple minimum supports using maximum constraints", *Lecture Notes in Computer Science*, Vol. 3214, pp. 1283-1290, 2004.

[18] B. Liu, W. Hsu, and Y. Ma, "Mining Association Rules with Multiple Minimum Supports," *ACM SIGKDD*, pp. 337-341, 1999.

[19] W. Y. Lin, M. C. Tseng and J.H. Su, "A Confidence-Lift Support Specification for Interesting Associations Mining," *The Pacific-Asia Conf. Advances in Knowledge Discovery and Data Mining*, pp. 148-158, 2002.

[20] Microsoft Corporation, Example Database FoodMart of Microsoft Analysis Services.

[21] A. Mangalampalli and V. Pudi, "Fuzzy association rule mining algorithm for fast and efficient performance on very large datasets," *IEEE International Conference on Fuzzy Systems*, pp. 1163 - 1168, 2009.

[22] A. Mangalampalli and V. Pudi, "FPrep: Fuzzy Clustering driven Efficient Automated Pre-processing for Fuzzy Association Rule Mining", *2010 IEEE International Conference on Fuzzy Systems*, p.1-8, 2010.

[23] W. Ouyang and Q. Huang, "Mining direct and indirect weighted fuzzy association rules in large transaction databases," *International Conference on Fuzzy Systems and Knowledge Discovery*, Vol. 3, pp. 128-132, 2009.

[24] A. T. H. Sim, M. Indrawan, S. Zutshi and B. Srinivasan, "Logic-Based Pattern Discovery," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 22 , No. 6, pp.798-811, 2010.

[25] A. Tajbakhsh, M. Rahmati, A. Mirzaei, "Intrusion detection using fuzzy association rules," *Applied Soft Computing*, Vol. 9, No. 2, pp. 462-469, 2009.

[26] G.I. Webb and S. Zhang, "k-Optimal Rule Discovery," *Data Mining and Knowledge Discovery*, Vol. 10, No. 1, pp. 39-79, 2005.

[27] H. Yun, D. Ha, B. Hwang and K.H. Ryu, "Mining Association Rules on Significant Rare Data Using Relative Support," *Journal of Systems Software*, Vol. 67, pp. 181-191, 2003.

[28] S. Yue, E. Tsang, D. Yeung and D. Shi, "Mining fuzzy association rules with weighted items," *The IEEE International Conference on Systems, Man and Cybernetics*, pp. 1906-1911, 2000.