

Inducing Fuzzy Association Rules with Multiple Minimum Supports for Time Series Data

Rakesh Rathi, Vinesh Jain and Anshuman Kumar Gautam

Abstract Technological changes have occurred at an exponential rate in recent years leading to the generation of large amount of data in various sectors. Several database and data warehouse is built to store and manage the data. As we know the data which are relevant to us should be extracted from the database for our task. Earlier different mining approaches are proposed in which items are collected at same minimum support value. In this paper we propose a fuzzy data mining algorithm which generates the fuzzy association rules from time series data having different minimum support values. The temperature varying dataset is used to generate fuzzy rules. The proposed algorithm also predicts the variation of temperature. Experiments are also performed to get the desired result.

Keywords Association rule · Data mining · Different minimum support · Fuzzy set

1 Introduction

Data mining plays a vital role in today's application. So, researchers are paying more attention toward the new tricks and techniques which can be evolved in it. It covers a large domain where it is frequently applied such as business, medical, biometrics. Fuzzy concepts have a great impact on data dredging methodology. Various data warehouses are managed to store and use the data efficiently through different domain. Time series data is a collection of data points which has some specific value

R. Rathi (✉) · V. Jain · A. K. Gautam
Government Engineering College Ajmer, Ajmer, Rajasthan, India
e-mail: rakeshrathi4@rediffmail.com

V. Jain
e-mail: vineshjain1280@gmail.com

A. K. Gautam
e-mail: er.anshuman2011@gmail.com

at that instant of time. It varies with respect to time. This paper proposes an algorithm which induce fuzzy association rule with multiple minimum support value. Earlier many algorithms have been proposed but they follow only single support value condition. Sometimes itemset has different minimum support. To explain this we applied a fuzzy concept on time series data more specifically temperature varying data. As we know that time series data comes under the category of Sequence data which has some trend or pattern in it. So algorithm would predict the near temperature using the trend analysis. The proposed algorithm has two advantages:—First, the result would be easier to understand as we are using fuzzy theory which is quite familiar with natural language. Second, it also helps to determine the sudden change in temperature of a place.

The remaining parts of this paper are assembled as follows: review of fuzzy set theory is given in Sect. 2. The related work of the paper is explained in Sect. 3. The proposed algorithm is explained in given in Sect. 4. Further experimental results are shown in Sect. 5. Finally conclusion and future work is discussed in Sect. 6.

2 Fuzzy Set Theory

Fuzzy set theory was pointed out in 1965 by Zadeh in his seminal paper entitled “Fuzzy sets” which played a vital role in human thinking, focusing in the domains of pattern recognition, communication of information and abstraction. Fuzzy set theory consists of fuzzy membership functions. Fuzzy set expresses the degree to which an element belongs to a set called as characteristic function. For a given crisp set B, the function assigns a value $\mu_B(x)$ to every $x \in X$ such that

$$\begin{aligned}\mu_B(x) &= \{1 \text{ iff } x \in B \\ \mu_B(x) &= \{0 \text{ iff } x \text{ does not } \in B\end{aligned}$$

Assume that x_1 to x_k are the elements in fuzzy set B, and μ_1 to μ_k are respectively their grades of membership function in B. B is usually represented as follows:

$$B = \mu_1/x_1 + \mu_2/x_2 + \dots + \mu_k/x_k \quad (1)$$

3 Related Work

Data mining is frequently used in inducing association rules from large itemsets. The association rules describes the effects of presence and absence of an item in a transaction with other items in terms of two measures support and confidence.

Hong proposed an algorithm which induces association rules with multiple minimum supports using maximum constraints on general items. Au and Chan

proposed a fuzzy dredging approach to find fuzzy rules for time series data. Das proposed a dredging algorithm for time series data prediction. Das used the clustering method to extract basic shapes from time series and applied Apriori method to induce the association rules on it.

4 The Proposed Algorithm with Multiple Minimum Support

Input: A time series TS with n data points, a list of m membership functions for data points, a predefined minimum support threshold for each fuzzy item $ms_i, i = 1$ to z, a predefined minimum confidence threshold λ , and a sliding window size ws.

- Step 1: Convert the time series TS into a list of subsequences W(TS) according to the sliding-window size ws. That is, $W(TS) = \{s_b | s_b = (d_b, d_{b+1}, \dots, d_{b+ws-1}), b = 1 \text{ to } n - ws + 1\}$, where d_b is the value of the b-th data point in TS.
- Step 2: Transform the k-th ($k = 1$ to ws) quantitative value v_{bk} in each subsequence s_b ($b = 1$ to $n - ws + 1$) into a fuzzy set f_{bk} represented as $(f_{bk1}/R_{k1} + f_{bk2}/R_{k2} + \dots + f_{bkn}/R_{kn})$ using the given membership functions, where R_{kl} is the l-th fuzzy region of the k-th data point in each subsequence, m is the number of fuzzy memberships, and f_{bkl} is v_{bk} 's fuzzy membership value in region R_{kl} . Each R_{kl} is called a fuzzy item.
- Step 3: Compute the scalar cardinality of each fuzzy item R_{kl} as

$$Count_{kl} = \sum_{b=1}^{n-ws+1} f_{bkl}$$

- Step 4: Check whether the support value ($= count_{kl}/n - ws + 1$) of each R_{kl} in C_1 is greater than or equal to its predefined minimum support threshold value $ms_{R_{kl}}$. If R_{kl} satisfies the above condition, collect it in the set of large 1-itemsets (L_1). That is:

$$L_1 = \{R_{kl} | count_{kl} \geq ms_{R_{kl}}, 1 \leq k \leq b + ws - 1 \text{ and } 1 \leq l \leq m\}.$$

- Step 5: IF L_1 is not null, then perform the next step; otherwise, terminate the algorithm.
- Step 6: Set $t = 1$, where t is used to represent the number of fuzzy items in the current itemsets to be processed.
- Step 7: Join the large t-itemsets L_t to obtain the candidate (t + 1)-itemsets C_{t+1} in the same way as in the Apriori algorithm provided that two items obtained from the same order of data points in subsequences cannot exist in an itemset in C_{t+1} at the same instant provided the minimum support of all the large t-itemsets must be greater than or equal to the maximum of the minimum supports of fuzzy items in theses large t-itemsets.

Step 8: Now, perform the following steps for fuzzy items in C_{t+1} :

- (a) Compute the fuzzy value of I in each subsequence s_b as $f_I^{sb} = f_{I1}^{sb} \wedge f_{I2}^{sb} \wedge \dots \wedge f_{I_{t+1}}^{sb}$ where $f_{I_k}^{sb}$ is the membership value of fuzzy item I_k in S_b . If the minimum operator is used for the intersection, then:

$$f_{Ib}^s = \text{Min}_{k=1}^{t+1} f_{I_k}^s.$$

- (b) Compute the count of I in all the subsequences as:

$$\text{Count}_I = \sum_{b=1}^{n-ws+1} f_I^{sb}$$

Step 9: If the support ($= \text{count}_I / n - ws + 1$) of I is greater than or equal to maximum of the minimum support value, put it in L_{t+1} .

$$L_{t+1} = \{I_k \mid \text{count}_I \geq \text{ms}_{I_k}, \}$$

Step 10: STEP 13: If L_{t+1} is null, then do the next step; otherwise, set $t = t + 1$ and repeat STEPs 6–9.

Step 11: Generate the association rules for each large h-itemset I with items (I_1, I_2, \dots, I_h) , $h \geq 2$, using the following substeps:

- (a) Form each possible association rule as follows: $I_1^{\wedge} \dots \wedge I_{n-1}^{\wedge} I_{n+1}^{\wedge} \dots \wedge I_h \rightarrow I_n$, $n = 1$ to h .
- (b) Calculate the confidence values of all association rules by the following formula:

$$= \sum_{b=1}^{n-ws+1} f_I^{sb} \setminus \sum_{b=1}^{n-ws+1} \left((f_I^{sb} \wedge \dots \wedge f_{I_P}^s) \right)$$

Output: A set of association rules which satisfies the condition of the maximum values of minimum supports.

5 An Example

This section explains the working of the proposed algorithm and generates fuzzy association rule (Table 1).

Assume the membership function used in the example as Fig. 1 (Table 2).

Step 1: The window size is assumed as 5. Using the formula we get $(15 - 5 + 1) = 11$ subsequences

Step 2: The data values are then converted into fuzzy item sets using the membership function shown in fig no.

Table 1 Set of data points

Time Series
3, 4, 6, 1, 8, 6, 3, 9, 3, 6, 1, 5, 8, 1, 9

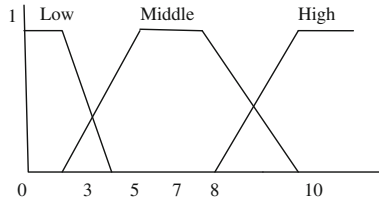


Fig. 1 Membership function used in this example

Table 2 Predefined minimum support value of all fuzzy itemset

Fuzzy item	Minimum support	Fuzzy item	Minimum support
Q1.Low	3.90	Q4.low	4.00
Q1.Middle	3.80	Q4.Middle	3.00
Q1.High	3.50	Q4.High	5.00
Q2.Low	2.00	Q5.Low	4.90
Q2.Middle	2.30	Q5.Middle	3.03
Q2.High	2.00	Q5.High	2.00
Q3.Low	2.10	Q3.High	4.00
Q3.Middle	4.50		

Table 3 Sequence generated, $ws = 5$

S_b	Subsequence	S_b	Subsequence
S_1	(3, 4, 6, 1, 8)	S_7	(3, 9, 3, 6, 1)
S_2	(4, 6, 1, 8, 6)	S_8	(9, 3, 6, 1, 5)
S_3	(6, 1, 8, 6, 3)	S_9	(3, 6, 1, 5, 8)
S_4	(1, 8, 6, 3, 9)	S_{10}	(6, 1, 5, 8, 1)
S_5	(8, 6, 3, 9, 3)	S_{11}	(1, 5, 8, 1, 9)
S_6	(6, 3, 9, 3, 6)		

- Step 3: Add all the value of fuzzy region of the subsequences called as its count. For example-Assume a fuzzy item Q_1 .Middle.count is $(0 + 0.33 + 1 + 0 + 0.33 + 1 + 0 + .2 + 0 + 1 + 0) = 3.86$
- Step 4: Now,compare the count of all fuzzy item with its individual minimum support count which is predefined. Fuzzy items whose count is greater than minimum support value of itself put the fuzzy item in the table L_1 (Table 3).
- Step 5: If L_1 consists of fuzzy item, proceed to step 6, else terminate.

Table 4 Converted fuzzy set

S _t	Q ₁			Q ₂			Q ₃			Q ₄			Q ₅		
	Low	Mid	High	Low	Mid	High	Low	Mid	High	Low	Mid	High	Low	Mid	High
S ₁	0.67	0	0	0.67	0.33	0	0	1	0	1	0	0	0	0.33	0.67
S ₂	0.67	0.33	0	0	1	0	1	0	0	0	0.33	0.67	0	1	0
S ₃	0	1	0	1	0	0	0	0.33	0.67	0	1	0	0.67	0	0
S ₄	1	0	0	0	0.33	0.67	0	1	0	0.67	0	0	0	0.2	1
S ₅	0	0.33	0.67	0	1	0	0.67	0	0	0	2	1	0.67	0	0
S ₆	0	1	0	0.67	0	0	0	0.2	1	0.67	0	0	0	1	0
S ₇	0.67	0	0	0	0.2	1	0.67	0	0	0	1	0	1	0	0
S ₈	0	0.2	1	0.67	0	0	0	1	0	1	0	0	0	1	0
S ₉	0.67	0	0	0	1	0	1	0	0	0	1	0	0.33	0.33	0.67
S ₁₀	0	1	0	1	0	0	0	1	0	0.33	0.33	0.67	1	0	0
S ₁₁	1	0	0	0	1	0	0	0.33	0.67	1	0	0	0	0.2	1
Count	4.68	3.86	1.67	4.01	4.86	1.67	3.34	4.68	2.34	4.34	3.68	2.34	2.34	2.70	3.34

Step 6: Candidate set C_{t+1} is generated from L_t. Fuzzy items in L₁ are (Q1.Low, Q1.Middle, Q2.Low, Q2.Middle, Q3.Low, Q3.Middle, Q4.Low, Q5.Low, Q5.High) .

Step 7: L₁ is joined to generate C₂. The new fuzzy items in C₂ are as follows (Q1.Low, Q2.Mid), (Q1.Low, Q3.Mid), (Q1.Low, Q5.High), (Q1.Low, Q2.Mid), (Q1.Low, Q3.Mid), (Q1.Low, Q5.Low), (Q2.Low, Q3.Low), (Q2.Low, Q4.Low), (Q2.Low, Q5.High), (Q2.Low, Q1.Mid), (Q2.Low, Q3.Mid), (Q2.Low, Q5.Low), (Q3.Low, Q4.Low), (Q3.Low, Q5.High), (Q3.Low, Q1.Mid), (Q3.Low, Q2.Mid), (Q3.Low, Q5.Low), (Q4.Low, Q1.Mid), (Q4.Low, Q2.Mid), (Q4.Low, Q5.High), (Q4.Low, Q5.Low), (Q5.High, Q1.Mid), (Q5.High, Q2.Mid), (Q5.High, Q3.High).

Table 5 Candidate set C₂

Fuzzy itemset	Count	Fuzzy itemset	Count
Q1Low, Q2Mid	4.68	Q2Low, Q5Low	2.34
Q1Low, Q3Mid	4.68	Q3Low, Q5High	3.34
Q1Low, Q5High	3.34	Q3Low, Q1Mid	3.34
Q1Low, Q2Low	4.01	Q3Low, Q2Mid	3.34
Q1Low, Q3Low	3.34	Q3Low, Q5Low	2.34
Q1Low, Q5Low	2.34	Q4Low, Q1Mid	3.86
Q2Low, Q3Low	3.34	Q4Low, Q2Mid	4.34
Q2Low, Q4Low	4.01	Q4Low, Q5High	3.86
Q2Low, Q3Low	3.34	Q4Low, Q2Mid	4.34
Q2Low, Q4Low	4.01	Q4Low, Q5High	3.86
Q2Low, Q5High	3.34	Q4Low, Q5Low	2.34
Q2Low, Q1Mid	3.86	Q5High, Q1Mid	3.34
Q2Low, Q3Mid	4.01	Q5High, Q2Mid	3.34
Q3Low, Q4Low	3.34	Q5High, Q3High	2.34

Table 6 Fuzzy itemset L_2

(Q1Low, Q2Low)(Q1Low, Q2Mid)(Q1Low, Q3Mid)(Q2Low, Q3Low)
(Q2Low, Q4Low)(Q2Low, Q5High)(Q1.Low, Q4.Low)
(Q3.Low, Q2.Mid)(Q4.Low, Q1.Mid)(Q5.High, Q2.Mid)(Q4.Low, Q5.High)

Step 8: Now compute the count of all the fuzzy items of C_2 .

Step 9: Compare the C_2 itemset count with minimum support count of Fuzzy itemset. C_2 items whose count is greater or equal to minimum support of maximum of the two itemset is stored in L_2 .

Step 10: Since L_2 is not null, repeat step no 6–9 until L_t is null (Tables 4, 5, 6).

Step 11: (a) In this example, only (Q3.Low Q2.Mid) exists. It means association rules formed are

If $Q_3 = \text{Low}$ then $Q_2 = \text{Mid}$. If $Q_2 = \text{Mid}$ then $Q_3 = \text{Low}$.

(b) Calculation of confidence of (Q3.Low Q2.Mid) = $3.34 \setminus 3.34 = 1$. It means if the value of a data point is mid at time2 then value of a data point is low at time3 with a confidence factor of 1.

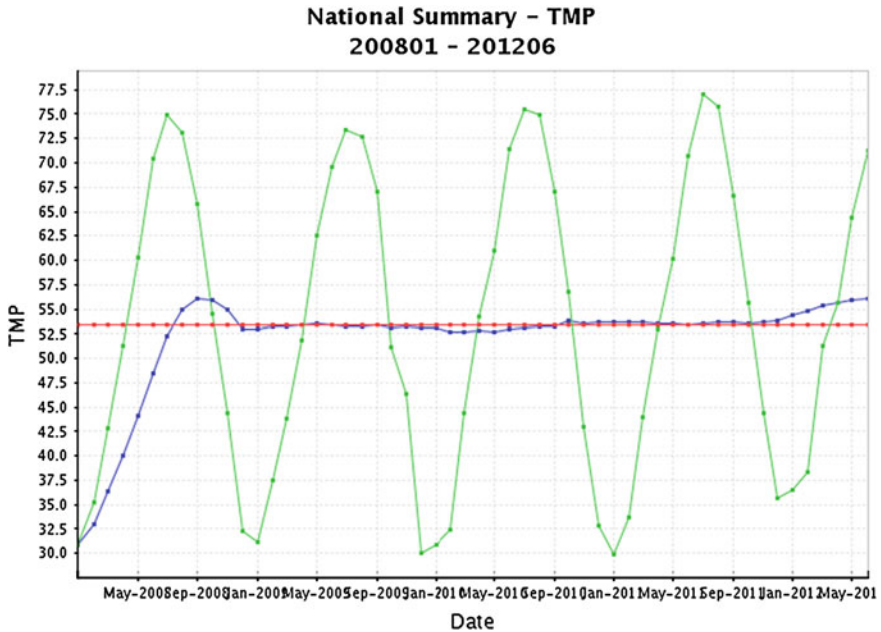


Fig. 2 Temperature varying data

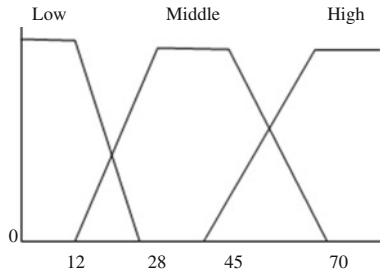


Fig. 3 Membership function used in experiment

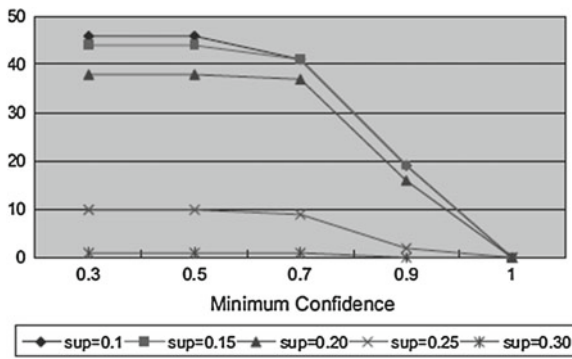


Fig. 4 Relation between support value and confidence

6 Experimental Results

The proposed algorithm is implemented in a programming language C. The dataset points consisted of temperature varying points between year 2008–2012. The dataset is taken from National Data Center (NDC) US (Figs. 2 and 3).

In Fig. 4 as the support value of fuzzy itemset is increased, number of fuzzy association rule decreased. This means change in temperature is effected with change in support value. It means that if the temperature of a day at second day of a month is moderate, then it may be high at third day of the month.

7 Conclusion and Future Work

In this paper, the proposed algorithm provided the best way to induce efficient fuzzy association rule as there is predefined minimum support for all the fuzzy items. The temperature prediction would be more accurate than earlier proposed approaches. Future work suggests that the membership function can be set dynamically. In this

paper membership functions are known in advance. More complex operations could be made in near future. It also provided us another view point for defining minimum support of fuzzy items.

References

1. Aggrawal, R.: Mining association rules between sets of items inlarge database ACM
2. Stepnicka, M.: Time series analysis and prediction based on fuzzy rules and the fuzzy transform
3. Dr.Sivatsa S.K.: Inaccuracy minimization by partitioning fuzzy data sets—validation of an analytical methodology(IJCSIS). *Int. J. Comput. Sci. Inf. Secur.* **8**(1), (2010)
4. Das, G.: Rule discovery from time series. In: Proceedings of the 4 the International Conference
5. Pongracz, R.: Application of fuzzy rule-based modeling technique to regional drought. *J. Hydrol.* **224**, 100–114 (1999)
6. Mueen, A.A.: Exact primitives for time series data mining. University of California, Riverside (2012)
7. Zhu, Y.: High performance data mining in time series: techniques and case studies. New work University, New York (2004)
8. Herrera, Francisco: Learning the membership function contexts for mining fuzzy association rules by using genetic algorithms. *Fuzzy Sets Syst.* **160**, 905–921 (2009)
9. Han, J.: Data Mining concepts and Techniques
10. Hong, T.P.: Mining association rules with multiple minimum supports. *Int. J. Approximate reasoning.* **3**, 38–42 (2005)