

An Approach for Quality Measure of Association Rule Based on QL-implicator

Wang Wen-qi

Department of Early Warning Intelligence, Air Force
Early Warning Academy
Air Force Early Warning Academy
Wuhan, China
wwq505@163.com

Li Qiang

Department of Early Warning Intelligence, Air Force
Early Warning Academy
Air Force Early Warning Academy
Wuhan, China
ldxy-cj@163.com

Abstract—To solve the low resolution of fuzzy association rule in airborne radar data mining using traditional quality measure, a fuzzy support degree improving approach based on QL-implicator was presented. Distinguished from The traditional approach which positive example was gained, non-counter example was gained by QL-implicator fuzzy support degree. If the former got more examples, so the number of examples in contradiction with the association was Infrequent. And this measure approach was be propitious to make a choice between two or more than two little different associations. Experimental result proves the effectiveness and feasibility of this approach.

Keywords—component; Association Rule; Quality Measure; QL-implicator; Support Degree;

I. INTRODUCTION

Similar to the common related rules, the quality metric of the fuzzy related rule refers the comprehensive metric of the legality, usefulness, novelty and briefness of rules [1]. For the research of quality metric, partial researchers have put forward various methods used for confirming degree of interest of modes, the produced large quantities of modes by using shaving related rule mining, including variance[2], chi-square related metric[2], entropy production metric, Laplace metric[3] and the metric of containing strength[4], the analysis of which are shown in literature[5], the quality metric put forward by Hüllermeier using fuzzy containing degree to build two kinds of fuzzy related rules[6,7] and De Cock and etc. have proved the effectiveness of S-implicator in the algorithm raised by Hüllermeier and the satisfying result cannot be acquired by using R-implicator.

Although numerous improved quality metric association rules have been put forward, most of them are in allusion to hard association rule algorithm and interval association rule algorithm and fewer of them can be directly applied to Fuzzy association rule field. For quantization properties waiting to be analyzed in some airborne radar intelligence database are numerous and their results are great in number, produced rules contain a lot of positive and negative rules, which shows that it is not enough to only depend on traditional quality metric methods. In actual situations, we use QL-implicator to improve Fuzzy supporting degree in the traditional quality metric so as the identification rate producing positive rules. For some airborne radar intelligence data's characteristics of dense data after

discretization, fast database updates and so on, it puts forward Fuzzy association rule mining algorithm basing on increasing quantity updates for demonstration.

II. FUZZY SUPPORTING AND CONFIDENCE COEFFICIENTS

Let $D = \{t_1, t_2, \dots, t_n\}$ be the database, $I = \{i_1, i_2, \dots, i_m\}$ be the universal set of database properties, among which n is the record sample number, $i_j (1 \leq j \leq m)$ represents classification or quantization property, m is the property number and classified property belongs to a specific situation of quantization property and it can be dealt with similar methods; $F = \{F(i_1), F(i_2), F(i_3), \dots, F(i_j)\}$ is the set of the final fuzzy classification of all properties; $F(i_j) = \{f_1^j, f_2^j, \dots, f_c^j\}$ is Fuzzy set of the single quantization property i_j and c is the number of fuzzy subsets. Then membership is determined for every fuzzy subset f_k^j of every quantization property i_j :

$$\mu_{f_k^j} : \text{dom}(i_j) \rightarrow [0,1], j=1, \dots, m, k=1, \dots, c \quad (1)$$

After obtaining Fuzzy classification of all quantization properties, it combines the membership coefficient of all property fuzzy sets, calculates the membership coefficient function of all continuous property fuzzy sets and finally uses membership coefficient functions of each fuzzy set to transfer the record set of the original record set to Fuzzy record set and $D^x = \{t^1, t^2, \dots, t^n\}$ is the finally database after the transfer.

Based on the above definition, we use the below form to represent a fuzzy association rule:

If $X = \{x_1, x_2, \dots, x_p\}$ is $A = \{a_1, a_2, \dots, a_p\}$ then $Y = \{y_1, y_2, \dots, y_q\}$ is $B = \{b_1, b_2, \dots, b_q\}$, it represents that if any record satisfies this rule antecedent, then the rule decedent has higher possibility. Among them, $a_i \in F(x_i)$, $i=1, \dots, p$, $b_j \in F(y_j)$, $j=1, \dots, q$. X and Y are designated to be the sub set of two intersecting property set, or they do not share a common property. A and B are associated with Fuzzy set associated with property X and Y .

For Fuzzy database $D^r = \{\ell^1, \ell^2, \dots, \ell^m\}$, \mathcal{N} is the record sum of the database. Define $Z = X \cup Y = \{z_1, \dots, z_{p+q}\}$ and $C = A \cup B = \{c_1, \dots, c_{p+q}\}$, the set of ordered pairs (Z, C) represents the property in Z or set the value of live fuzzy sub set c_j of fuzzy sub sets. If fuzzy association rule $(X, A) \rightarrow (Y, B)$ is the funny rule, Fuzzy supporting coefficient $FS_{(z,c)}$ satisfies the minimum supporting coefficient \min_Fsup and Fuzzy confidence coefficient $FC_{((x,A),(y,B))}$ satisfies the minimum confidence coefficient \min_Fconf .

The traditional fuzzy supporting coefficient is calculated by the following formula:

$$FS_{(z,c)} = \frac{\sum_{j=1}^n \prod_{j=1}^m \mu(\ell^j[(z_j, c_j)])}{n} = \frac{\sum_{j=1}^n \prod_{x,y \in I} (F_A(x), F_B(y))}{n} \quad (2)$$

In it, $\prod_{j=1}^m \mu(\ell^j[(z_j, c_j)])$ is the product of confidence coefficients of fuzzy sub sets associated with Fuzzy database property set ℓ^j , and $\mu(\ell^j[(z_j, c_j)]) > \varepsilon, \varepsilon \in (0,1)$, m is the number of the ordered pair (z_j, c_j) in the set of ordered pairs (Z, C) , s is the threshold value users designate. If Fuzzy mode membership coefficient $\mu(\ell^j[(z_j, c_j)])$ is smaller than s , this mode is thought to be of no interest and not considered or calculated.

Fuzzy confidence coefficient is calculated by the following formula:

$$FC_{((x,A),(y,B))} = \frac{FS_{(z,c)}}{FS_{(x,A)}} \quad (3)$$

III. THE IMPROVED METHOD OF FUZZY SUPPORTING COEFFICIENTS BASING ON QL-IMPLICATOR

We use QL-implicator to improve the traditional supporting coefficient calculation method, firstly introduce relevant definitions of t-norm, t-conorm, strict negator and fuzzy implicator.

Definition 1 T-norm is a interchangeable, associated and monotonic non-diminishing function, $\mathcal{T}: [0,1]^2 \rightarrow [0,1]$, satisfying $\mathcal{T}(x,1) = x, \forall x \in [0,1]$ at the same time.

Continuous t-conorm is a interchangeable, associated and monotonic non-diminishing function, $\mathcal{S}: [0,1]^2 \rightarrow [0,1]$, satisfying $\mathcal{S}(x,0) = x, \forall x \in [0,1]$ at the same time.

Strict negator is a diminishing continuous function, $\mathcal{N}: [0,1] \rightarrow [0,1]$, satisfying $\mathcal{N}(0) = 1$ and $\mathcal{N}(1) = 0$ at the same time.

For fuzzy set A and B, complementary operation, intersecting operation and unite operation are defined as follows:

$$co_{\mathcal{N}}A(x) = \tilde{A}(x) = \mathcal{N}(\mu_A(x)),$$

$$(A \cap_{\mathcal{T}} B)(x) = \mathcal{T}(\mu_A(x), \mu_B(x)),$$

$$(A \cup_{\mathcal{S}} B)(x) = \mathcal{S}(\mu_A(x), \mu_B(x)).$$

The base of a fuzzy set A in limited universe is defined by the following formula: $|A| = \sum_{x \in [0,1]} \mu_A(x)$

Definition 2 Fuzzy implicator is a function satisfying the following conditions: $\mathcal{I}: [0,1]^2 \rightarrow [0,1]$.

I1: if $x \leq z$ then $\mathcal{I}(x,y) \geq \mathcal{I}(z,y), \forall x, y, z \in [0,1]$.

I2: if $y \leq z$ then $\mathcal{I}(x,y) \geq \mathcal{I}(z,y), \forall x, y, z \in [0,1]$.

I3: $\mathcal{I}(1,y) = 1 \Leftrightarrow y = 1$,

I4: $\mathcal{I}(1,0) = 0, \mathcal{I}(0,0) = 0$.

By the above definitions, we give the definitions of S-implicator and QL-implicator [10,11] as follows:

S-implicator:

$$\mathcal{I}_S^{\mathcal{S}, \mathcal{N}}(x,y) = \mathcal{S}(\mathcal{N}(x), y), \forall x, y \in [0,1] \quad (4)$$

QL-implicator:

$$\mathcal{I}_{QL}^{\mathcal{T}, \mathcal{S}, \mathcal{N}}(x,y) = \mathcal{S}(\mathcal{N}(x), \mathcal{T}(x,y)), \forall x, y \in [0,1] \quad (5)$$

Hüllermeier puts forward a fuzzy association rule $(X, A) \rightarrow (Y, B)$ basing on fuzzy implicator \mathcal{I} in literature (6):

$$supp_{\mathcal{I}}((X, A) \rightarrow (Y, B)) = \sum_{x,y \in I} \mathcal{I}(F_A(x), F_B(y)) \quad (6)$$

The following demonstrates whether QL-implicator applies into formula (6):

Let D^r be a non-empty affair database only including dual property values, and D_A^r and D_B^r represent sets including affairs of A and B separately, coD_A^r and coD_B^r represent sets not including affairs of A and B separately. Then for association rule $A \rightarrow B$, its supporting number and supporting coefficient are defined separately as:

$$supp \#(A \rightarrow B) = |D_A^r \cap D_B^r| \quad (7)$$

$$supp(A \rightarrow B) = \frac{|D_A^r \cap D_B^r|}{|D^r|} \quad (8)$$

In it, $|D_A^r \cap D_B^r|$ is the affair number including A and B , $|D^r|$ represents the base of database D^r .

Definition 3 De Cock classifies and defines the affair x satisfying rule $A \rightarrow B$ in literature [8][9] as follows:

x is a positive example only when $x \in D_A^r \wedge x \in D_B^r$

x is a non positive example only when $x \notin D_A^r \vee x \notin D_B^r$

x is a counter example only when $x \in D_A^r \wedge x \notin D_B^r$

x is a non counter example only when $x \notin D_A^r \vee x \in D_B^r$

According to this definition, the obtained maximum supporting number is:

$$\begin{aligned} maxsupp \#(A \rightarrow B) &= |coD_A^r \cup D_B^r| \\ &= |(coD_A^r \cup D_B^r) \cap (coD_A^r \cup D_B^r)| \\ &= |coD_A^r \cup (D_A^r \cap D_B^r)| \end{aligned} \quad (9)$$

The maximum fuzzy supporting number of Fuzzy association rule $(X, A) \rightarrow (Y, B)$ extended by formula (9) is:

$$\begin{aligned} & \text{fmaxsup} \#((X, A) \rightarrow (Y, B)) \\ &= \sum_{x, y \in I} \mathcal{S} \left(\mathcal{N} \left(F_A(x) \right), \mathcal{T} \left(F_A(x), F_B(y) \right) \right) \\ &= \sum_{x, y \in I} \mathcal{I}_{QL}^{\mathcal{T}, \mathcal{S}, \mathcal{N}} \left(F_A(x), F_B(y) \right) \end{aligned} \quad (10)$$

Obviously, formula (10) shows that QL-implicator can be applied into formula (6), therefore Fuzzy supporting coefficient and confidence coefficient basing on QL-implicator are as follows:

$$FS_{(z, c)}^{QL} = \text{fmaxsup} \left((X, A) \rightarrow (Y, B) \right) = \frac{\sum_{x, y \in I} \mathcal{I}_{QL}^{\mathcal{T}, \mathcal{S}, \mathcal{N}} \left(F_A(x), F_B(y) \right)}{|D^r|} \quad (11)$$

$$FC_{((X, A), (Y, B))}^{QL} = \frac{FS_{(z, c)}^{QL}}{FS_{(X, A)}} \quad (12)$$

For using QL-implicator maximum fuzzy supporting coefficient gets the non-counter example in the rule, the traditional supporting coefficient only obtains the just example. If the former gets larger numbers of cases, it represents that the case number of this rule contradiction is relatively small, while the quality of this rule is higher. Meanwhile, this metric method is beneficial to selecting the rule in two or more than two rules with lower differences of supporting coefficients. For example, if the traditional fuzzy supporting coefficients gained from two rules are 0.22 and 0.23 separately, the supporting metrics obtained from the above method can reach 0.68 and 0.45 and the first rule is selected.

IV. SOME AIRBORNE RADAR INTELLIGENCE DATA FUZZY ASSOCIATION RULE MINING ALGORITHM BASING ON INCREMENTAL UPDATES

The incremental updating problem of association rule is to solve how to high-efficiently update the deduced association rule from the updated database. Incremental style updating mining algorithm is to use the mined rule as the guidance, finds out the new rule on the variable data set or threshold values and further solves the problem of knowledge updates. With the increase of some airborne radar execution missions, the intelligence information data will continually update. If every time of intelligence analysis needs to mine on all historical data repeatedly, it needs to consume too much manpower and energies. In order to cope with this type of problems that may appear in the future, it introduces the idea of incremental mining into Fuzzy association algorithm, with the introduction of the concept of negative boundaries, it needs to complete a whole scanning on the whole database only when the data update causes the negative boundary extension of large item sets and every iteration produces larger frequent item sets. The basis implement steps of algorithm are as follows:

Let DB be the original database, the minimum supporting coefficient threshold value is min_sup , the

frequent item set and negative boundary of the original database DB are L and N separately, the frequent item set and negative boundary in ab are L^{DB+db} and N^{DB+db} separately.

- Traverse the newly added database, obtain L_1^{db} and N_1^{db} ,
- L_1^{db} and N_1^{db} , item S has three possibilities: a) $s \in L_1^{db}$, b) $s \in N_1^{DB}$, c) $s \notin L_1^{DB} \cup N_1^{DB}$. Judge them, update their supporting numbers and add them into L_1^{DB+db} and N_1^{DB+db} separately,
- Repeat the above steps, find out L_k^{db} and N_k^{db} separately and add them into L_k^{DB+db} and N_k^{DB+db} until the frequent item without greater dimensions appear,
- For item sets belonging to $N^{DB} \cup L^{DB}$ but not to $N^{db} \cup L^{db}$, we judge their belongings depending on their supporting number t_{DB} in the original database and add them into L^{DB+db} or N^{DB+db} separately,
- For the s newly added in L_k^{DB+db} , or $s \notin L_k^{DB}$, $s \in N_k^{DB}$, we use association rule algorithm to produce its relevant candidate item set C_{k+1}^{DB} and prune it with L_{k+1}^{db} ,
- The data quantity in C_{k+1}^{DB} after shaved is very small or does not exist. If there is another candidate item set in C_{k+1}^{DB} , DB should be scanned again and C_{k+1}^{DB} should be merged into L_{k+1}^{DB+db} or N_{k+1}^{DB+db} separately,
- The algorithm is finished.

V. EXPERIMENT RESULTS AND ANALYSIS

In order to demonstrate the effectiveness the improved method of fuzzy metrics mentioned in the above, it selects 5000 pieces of simulation record data of some airborne radar intelligence to analyze the feasibility and effectiveness of this method. As shown in Table 1, a_1, a_2, \dots, a_5 are separately airborne inertial navigation data properties: airborne longitude, airborne latitude, airborne height (meter), airborne speed (meter/sec) and airborne angle of pitch, which are all quantization properties; a_6 represents the beam type, which is the classified property.

Table 1 Table 1 Partial records of airborne inertial navigation database

ID	a_1	a_2	a_3	a_4	a_5	a_6
1	118.74126	31.05182	8398	142.67	5.883	M
2	118.74107	31.05149	8396	141.45	6.004	M
3	118.73994	31.04358	8397	141.44	5.960	M
4	118.73852	31.04393	8397	141.43	5.982	M

Discretize the continuous and classifies properties of some airborne radar intelligence, which is shown as Table 2:

Table 2 Property and its fuzzy classification

		a_1			a_2		
0.532	0.25	0	0	0	1	0	0
0.75	0.15	0	0	0	1	0	0
0.9	0	0	0	0	1	0	0
0	0.31	0.68	0	0	0	1	0
0	0	0	0.70	0.24	0	0	1

After completing property fuzzy classification, it uses membership functions of each fuzzy set to transfer the record set of the original database to Fuzzy record set. If the corresponding property of record item is classified property, the corresponding record item will be divided on the classification number of properties, each type every sub item value belongs to is membership coefficient $\mu = \{0,1\}$. Table 3 gives fuzzy record transfer results of first five database records of “inertial navigation longitude” and “beam type”.

Table 3 Fuzzy generating record of partial property records

		a_1			a_2		
0.532	0.25	0	0	0	1	0	0
0.75	0.15	0	0	0	1	0	0
0.9	0	0	0	0	1	0	0
0	0.31	0.68	0	0	0	1	0
0	0	0	0.70	0.24	0	0	1

By the above process, the original record set D is transferred to Fuzzy record set D^f . First, it uses traditional fuzzy supporting coefficient and fuzzy confidence coefficient frame [12, 13] to mine on D^f and the minimum supporting coefficient and minimum confidence threshold value are set for 0.2 and 0.3. The produced partial fuzzy association rule is indicated as Table 4.

Table 4 Fuzzy association rule produced by the traditional quality metric

ID	rule	max fuzzy support degree	confidence
1	$a_1l_1 \diamond a_1l_1$	0.426	0.575
2	$a_1l_2 \diamond a_6l_1$	0.587	0.723
3	$a_3l_3 \diamond a_6l_2$	0.611	0.731
4	$a_2l_3 \diamond a_4l_2$	0.402	0.598
5	$a_3l_2 \diamond a_5l_1$	0.317	0.530
6	$a_4l_2, a_5l_2 \diamond a_6l_1$	0.557	0.719

In the next, it uses the maximum fuzzy supporting coefficient frame basing QL-implicator to mine on Fuzzy record set D^f . In the process of calculating QL-implicator, set $\mathcal{N}(x)$ as the standard denied implicator, or $\mathcal{N}(x) = 1 - x$, meanwhile take t-norm mould and continuing t-conorm formulas are $\mathcal{T}(x, y) = \min(x, y)$ and $\mathcal{S}(x, y) = \max(x, y)$ separately, the calculating formula obtained from formula (5) is: $\mathcal{I}_{QL}(x, y) = \max(1 - x, \min(x, y))$ and the maximum supporting coefficient and confidence coefficient can be calculated by formula (11)(12) and the minimum supporting

coefficient and minimum confidence threshold value are still 0.2 and 0.3, and the producing rule is shown as Table 5.

Table 5 Fuzzy association rule produced by new quality metric frame

ID	rule	max fuzzy support degree	confidence
1	$a_1l_1 \diamond a_1l_1$	0.426	0.575
2	$a_1l_2 \diamond a_6l_1$	0.587	0.723
3	$a_3l_3 \diamond a_6l_2$	0.611	0.731
4	$a_2l_3 \diamond a_4l_2$	0.402	0.598
5	$a_3l_2 \diamond a_5l_1$	0.317	0.530
6	$a_4l_2, a_5l_2 \diamond a_6l_1$	0.557	0.719

Table 5 gives the new metric value, which is of the same rule in Table 4. From Table 5, we can see that the maximum fuzzy support coefficients of rule 1, 4 and 5 are relatively small, which can confirm that these three conditional rules are negative association rules and should be shaved.

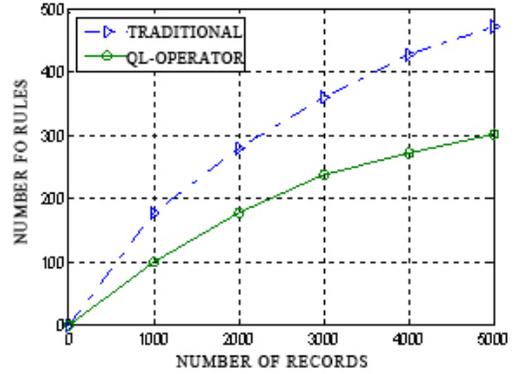


Fig. 1 Contrasting curve of the quantity rules produce

From Figure 1, we can see the contrasting situation of the produced ruling quantities by using traditional quality metric and the new quality metric frame mentioned in this paper, the latter produced rules are more concise than those of the former.

For the quality metric of the fuzzy related rule mentioned in this paper has the rule filtering function in the mining process, its calculating time greatly decreases compared to the calculating time in traditional metric frame. Just as Figure 2 shows the contrasting situation of algorithm execution time in two metrics, it can be clearly seen from the figure that using the method put forward in the paper produces useful rules and greatly decreases the time algorithm execution needs.

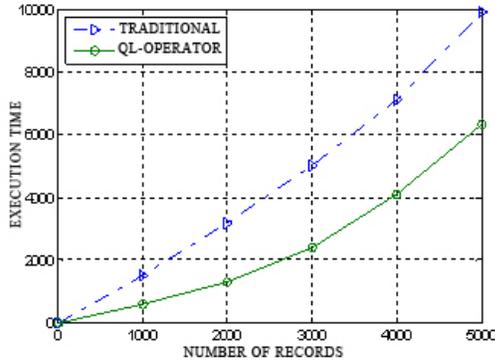


Fig. 2 Contrasting curve of algorithm execution time

VI. CONCLUSION

The chapter puts forward a calculation method of fuzzy support based on QL-operator, uses the method mentioned in the paper to produce the smaller rule quantity compared to the number acquired by the traditional method and greatly raises the mining efficiency; meanwhile for the actual need of some airborne radar information analysis, it puts forward a mining algorithm improving the fuzzy related rule and facilitating information data bases of early warning aircraft, the experiment results show that introducing QL-operator as the quality metric of the fuzzy related rule has relatively excellent performances compared to the traditional metric method. The question this paper further researches on is studying proper compressive algorithm storage items, reducing internal memory us and producing frequent items more efficiently.

REFERENCES

- [1] U. Fayyad ,G. Piantesky, "From Data Mining to Knowledge Discovery," Advances in Knowledge Discovery and Data Mining, California: AAAI Press, 1996,pp.1-36.
- [2] S.Morishita, "On Classification and Regression," In Proc of the First Int. Conf. On Discovery Science -LNAI 1532, 1998,pp.40-57.
- [3] P.Clark, P.Boswell, "Rule Induction with CN2: Some Recent Improvements, " In Machine Learning: Proc. of the Fifth European Conference, 1991,pp.151-163.
- [4] Bernadet M. "Basis of a Fuzzy Knowledge Discovery System[C]. In Proc. of the 4th European Conference on PKDD, 2000: 24-33.
- [5] J.Bayardo, "Agrawal R. Mining the Most Interesting Rules," In Proc. of the 5th ACM SIGKDD, 1999,pp.145-154.
- [6] E.Hüllermeier, "Fuzzy Association Rules: Semantic Issues and Quality Measures," In Proceedings of the International Conference, 7th Fuzzy Days on Computational Intelligence, Theory and Applications, volume 2606 of Lecture Notes in Computer Science, London, UK, Springer-Verlag, 2001,pp. 380-391.
- [7] E.Hüllermeier, "Mining Implication-based Fuzzy Association Rules in Databases," In Proc of IPMU2002, 2002,pp.101-108.
- [8] M.Cock, C.Cornelis, and E.Kerre, "Fuzzy Association Rules: A Two-sided Approach," In Proc of FIP2003, Tsinghua Univ. Press, 2003,pp. 385-390.
- [9] M.Cock, C.Cornelis, E.Kerre, "A Clear View on Quality Measures for Fuzzy Association Rules," In Proc of FSSCEF, Tsinghua Univ. Press, 2004,pp.54-61.
- [10] E.Czogala, J.Leski, "On Equivalence of Approximate Reasoning Results Using Different Interpretations of Fuzzy If-Then Rules," Fuzzy sets and Systems,vol.117,Feb. 2001,pp.279-296.
- [11] S.Weber, "A General Concept of Fuzzy Connectives, Negations and Implications Based on t-norms and t-conorms," Fuzzy Sets and Systems, vol.11,Feb.1983,pp.115-134.
- [12] H.Verlinde, M.DeCock, and R.Boute, "Fuzzy Versus Quantitative Association Rules: A Fair Data-Driven Comparison," IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, vol.36,Mar.2006,pp.679-683.
- [13] D.Dubois, E.Hüllermeier, and H.Prade, " A Systematic Approach to the Assessment of Fuzzy Association Rules," Data Min. Knowl. Discov, vol.13,Feb.2006, pp.167-192.