International J. of Math. Sci. & Engg. Appls. (IJMSEA)
ISSN 0973-9424, Vol. 11 No. I (April, 2017), pp. 67-77

FUZZY ASSOCIATION RULES : TRANSACTION CLASSIFICATIONS AND RELATED QUALITY MEASURES

T. S. THANUJA[1] AND MARY GEORGE[2]
[1] Research Scholar,
Dept of Mathematics, Mar Ivanios College, Trivandrum, India
[2]Associate Professor and Head,
Dept of Mathematics, Mar Ivanios College, Trivandrum, India

Abstract

Association rules are initially discovered in the market basket analysis [1] to identify frequently purchased items by customers. Usually support and confidence measures are used to assess the quality of association rules. Here we identify the set of transactions that are not in favour of the rule and categorize them as true positive, true negative, false positive and false negative examples based on the presence and absence of items in the transactions. In this paper, we analyzed the true semantics of these rule quality measures based on this categorization and defined some confidence measures with fuzzy operators, and identified them with precision, specificity, sensitivity, interest, conviction of association rules.

1. Introduction

Data mining is the process of extracting previously unknown and potentially useful hidden predictive information from large amounts of data [1]. Association rules are

---

Key Words : *Fuzzy association rules, Support, Opposition, True positive, True negative, False positive and False negative examples.*

© http: //www.ascent-journals.com

initially discovered in the market basket analysis to identify frequently purchased items by customers. It give certain regularities and dependencies within a data by finding frequent co-occurrence of items with a set of transactions and relationships hidden in large data sets. The uncovered relationships can be expressed as association rules or frequent itemsets. In classical association rules, it is not possible to use every data for mining. In most real life applications, the database contains many attributes which are difficult to represent using binary values. In such cases fuzzy sets play a major role. So in the process of association rule mining, fuzzy sets can handle both quantitative and categorical data, providing the necessary support to use uncertain data types with existing algorithms. The approach of quantitative mining allow attributes to be either members or non-members of an interval which tends to make an under or over estimation of values leading to sharp boundary problems. The use of fuzzy sets in association analysis widens the type of relationships between attributes by allowing the intervals to overlap, giving partial memberships to different sets thus avoiding unnatural boundaries in the partitioning of the attribute domain and thus making the interpretation of rules in linguistic terms easier. Thus the obtained results using fuzzy approaches are easy to understand and to apply.

An association rule is of the form $A \rightharpoonup B$, where A and B are attributes or sets of attributes, which tells the idea that *when A occurs in a transaction, B is likely to occur as well*. The strength of association rules can be realized by a number of quality measures. Support and confidence are the two important quality measures used essentially. Support measures the validity of an association rule where as confidence measures the quality of the rule. Thus mining association rules means, to generate all association rules $A \rightharpoonup B$ that have support and confidence greater than the user specified thresholds. These measures can be generalized for fuzzy association rules as well. Here we study about the transaction types and redefined their terminologies as true positive, true negative, false positive, false negative examples to understand the true semantics of the transactions. In this paper we tried to define the different quality measures based on confusion matrix terminology and analyzed the semantics of these measures and identified them with precision, specificity, sensitivity , interest, conviction of association rules using fuzzy operators.

The next section explains the definition of association rules, transaction classification, their support and confidence measures. Sections 3 is devoted to fuzzy association rules, fuzzy support and confidence measures. Section 4 explains the semantics of the defined measures based on confusion matrix. Section 5 gives some more fuzzy support and confidence measures which explains the precision, specificity, sensitivity of each association rule using the new defined terminologies.

2. Association Rules

2.1 Transaction Classification and Support Measures

Definition 2.1 : The support count and respectively support of an association rule $A \rightharpoonup B$ is defined as:

$$supp\#(A \rightharpoonup B) = |T_A \cap T_B|$$

and respectively

$$supp(A \rightharpoonup B) = \frac{|T_A \cap T_B|}{|T|}$$

This definition of support count positive examples as it represents the transactions that explicitly support the association expressed by the rule.

De Cock et al. ([5], [6]) classified transactions with respect of an association rule as positive example, non-positive example, negative example, non-negative example.

In order to explore the true semantics of the transaction classification, we introduce some new terminologies encouraging from the definition of confusion matrix. Thus we define.

Definition 2.2 : Let $A \rightharpoonup B$ be an association rule and t be a transaction. Then

- t is a true positive example iff $t \in T_A \wedge t \in T_B$.
- t is a true negative example iff $t \notin T_A \vee t \notin T_B$.
- t is a false positive example iff $t \in T_A \wedge t \notin T_B$.
- t is a false negative example iff $t \notin T_A \vee t \in T_B$.

This indicates how effective our expectations. In true positive and true negative example, we got what we expect, according as presence or absence of items. In false positive examples we assume the presence of some items, but it was a false one and in false negative examples, we assume the absence of some items and it appeared to be false.

Based on this classification, we get the following different measures:

- minimum support count: $minsupp\#(A \rightharpoonup B) = |T_A \cap T_B|$
- maximum opposition count: $maxopp\#(A \rightharpoonup B) = |\tilde{T_A} \cup \tilde{T_B}|$
- minimum opposition count: $minopp\#(A \rightharpoonup B) = |T_A \cap \tilde{T_B}|$
- maximum support count: $maxsupp\#(A \rightharpoonup B) = |\tilde{T_A} \cup T_B|$

and the corresponding measures is given by

- minimum support: $minsupp(A \rightharpoonup B) = \frac{|T_A \cap T_B|}{|T|}$
- maximum opposition: $maxopp(A \rightharpoonup B) = \frac{|\tilde{T_A} \cup \tilde{T_B}|}{|T|}$
- minimum opposition: $minopp(A \rightharpoonup B) = \frac{|T_A \cap \tilde{T_B}}{|T|}$
- maximum support: $maxsupp(A \rightharpoonup B) = \frac{|\tilde{T_A} \cup T_B|}{|T|}$

Remark 2.3 :

$$minsupp(A \rightharpoonup B) \leq maxsupp(A \rightharpoonup B)$$
$$minopp(A \rightharpoonup B) \leq maxopp(A \rightharpoonup B)$$

2.3 Confidence Measures

Definition 2.4 : The confidence of a rule $A \rightharpoonup B$ is defined as:

$$conf(A \rightharpoonup B) = \frac{supp\#(A \rightharpoonup B)}{supp\#(A)} = \frac{supp\#(A \rightharpoonup B)}{supp\#(A \rightharpoonup B) + supp\#(A \rightharpoonup \tilde{B})} \quad (1)$$

Confidence can be treated as the conditional probability ($P(B|A)$) or the relative cardinality of B with respect to A.

Definition 2.5 : The confidence measure, n-confidence is defined as:

$$conf_n(A \rightharpoonup B) = \frac{minsupp\#(A \rightharpoonup B)}{minopp\#(A \rightharpoonup B)} \quad (2)$$

Definition 2.6 : The pessimistic confidence p-confidence and the optimistic confidence o-confidence are defined as:

$$conf_p(A \rightharpoonup B) = \frac{minsupp\#(A \rightharpoonup B)}{maxopp\#(A \rightharpoonup B)} \quad (3)$$

$$conf_o(A \rightharpoonup B) = \frac{maxsupp\#(A \rightharpoonup B)}{minopp\#(A \rightharpoonup B)} \tag{4}$$

Definition 2.7 : Given a pair (M_1, M_2) of quality measures for association rules, with the property $M_1\mathbb{R} \leq M_2\mathbb{R}$, the inferior confidence and superior confidence are defined as

(a) inferior confidence

$$conf_*\mathbb{R} = \frac{\alpha.M_1\mathbb{R}}{(1-\beta).M_1\mathbb{R} + \beta.M_2(A \rightharpoonup B)} \tag{5}$$

(a) superior confidence

$$conf^*\mathbb{R} = \frac{\alpha.M_2\mathbb{R}}{(1-\beta).M_1\mathbb{R} + \beta.M_1(A \rightharpoonup B)} \tag{6}$$

Remark 2.8 :

$$conf_p(A \rightharpoonup B) \leq conf_n(A \rightharpoonup B) \leq conf_o(A \rightharpoonup B) conf_*\mathbb{R} \leq conf^*\mathbb{R}$$

3. Fuzzy Association Rules

3.1 Fuzzy Set and Fuzzy Set Operations

A fuzzy set A in a given universal set X is a mapping from $X \rightarrow [0,1]$, usually denoted as $A = \{(x, A(x)) : x \in X\}$ where $A(x)$ is called the grade of membership of each $x \in A$. The cardinality of a fuzzy set A in X is defined as $|A| = \Sigma_{x \in X} A(x)$.

A monotonic, associative and commutative mapping from $[0,1]^2 \rightarrow [0,1]$ is called $t-norm$ $\mathcal{T}$, if it satisfies $\mathcal{T}(x,1) = x$ for all $x \in [0,1]$ and a $t-conorm$ $\mathcal{S}$ if it satisfies $\mathcal{S}(x,0) = x$ for all $x \in [0,1]$. A fuzzy complement $\mathcal{N}$ is a decreasing mapping from $[0,1] \rightarrow [0,1]$ satisfying $\mathcal{N}(0) = 1$ and $N(1) = 0$.

For the fuzzy sets A and B in X, the complement, intersection and union can be defined by

$$coA(x) = \tilde{A}(x) = \mathcal{N}(A(x))$$
$$(A \cap_{\mathcal{T}} B)(x) = \mathcal{T}(A(x), B(x))$$
$$(A \cup_{\mathcal{S}} B)(x) = \mathcal{S}(A(x), B(x))$$

3.2 Fuzzy Support Measures

Definition 3.1 : The fuzzy support count and respectively fuzzy support of a fuzzy association rule $\langle A, F_A \rangle \rightharpoonup \langle B, F_B \rangle$ is usually defined as:

$$fsupp\#(\langle A, F_A \rangle \rightharpoonup \langle B, F_B \rangle) = \Sigma_{x \in T}(F_A \cap_{\mathcal{T}} F_B)(x)$$

and respectively

$$fsupp(\langle A, F_A \rangle \rightharpoonup \langle B, F_B \rangle) = \frac{\Sigma_{x \in T}(F_A \cap_{\mathcal{T}} F_B)(x)}{|T|}$$

Definition 3.2 : Let $\langle A, F_A \rangle \rightharpoonup \langle B, F_B \rangle$ be a fuzzy association rule. Then we define:

a) fuzzy minimum support:

$$fminsupp(\langle A, F_A \rangle \rightharpoonup \langle B, F_B \rangle) = \frac{\sum\limits_{x \in T}(F_A \cap_{\mathcal{T}} F_B)(x)}{|T|}$$

b) fuzzy maximum opposition:

$$fmaxopp(\langle A, F_A \rangle \rightharpoonup \langle B, F_B \rangle) = \frac{\sum\limits_{x \in T}(\tilde{F_A} \cup_{\mathcal{S}} \tilde{F_B})(x)}{|T|}$$

c) fuzzy minimum oppposition:

$$fminopp(\langle A, F_A \rangle \rightharpoonup \langle B, F_B \rangle) = \frac{\sum\limits_{x \in T}(F_A \cap_{\mathcal{T}} \tilde{F_B})(x)}{|T|}$$

d) fuzzy maximum support:

$$fmaxopp(\langle A, F_A \rangle \rightharpoonup \langle B, F_B \rangle) = \frac{\sum\limits_{x \in T}(\tilde{F_A} \cup_{\mathcal{S}} F_B)(x)}{|T|}$$

Similarly we can define the corresponding count measures: $fminsupp\#$,$fmaxopp\#$, $fminopp\#$, $fmaxsupp\#$.

3.3 Fuzzy Confidence Measures

Definition 3.3 : The fuzzy confidence of a fuzzy association rule $\langle A, F_A \rangle \rightharpoonup \langle B, F_B \rangle$ is defined as:

$$fconf(\langle A, F_A \rangle \rightharpoonup \langle B, F_B \rangle) = \frac{\sum\limits_{x \in T}(F_A \cap_{\mathcal{T}} F_B)(x)}{\Sigma_{x \in T} F_A(x)}$$

Now we define fuzzy version of $conf_n, conf_p, conf_o$ defined by Hullermeier [10] and DeCook et al., [5].

Definition 3.4 : The fuzzy confidence measures n-confidence, p-confidence and o-confidence of a fuzzy association rule $\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle$ is defined as:

a) fuzzy n-confidence

$$fconf_n(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle) = \frac{fminsupp\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)}{fminopp\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)} \tag{7}$$

b) fuzzy pessimistic confidence

$$fconf_p(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle) = \frac{fminsupp\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)}{fmaxopp\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)} \tag{8}$$

c) fuzzy optimistic confidence

$$fconf_o(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle) = \frac{fmaxsupp\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)}{fminopp\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)} \tag{9}$$

4. Into the Semantics of the Defined Measures

Based on the confusion matrix terminology (Table 1), we divide the transaction into true positive(TP), true negatives(TN), False Positive(FP) and False Negative (FN) examples. For an association rule $A \rightharpoonup B$,where A is the antecedent and B is the consequent of the rule, TP is the number of instances which match with rule antecedent and consequent, TN is the number of instances which match rule antecedent and consequent FP is the number of instances which match only with rule antecedent and FN is the number of instances which match only with rule consequent.

Table 1 : Confusion Matrix

	Predicted class	
Actual class	Yes	No
Yes	TP: True Positive	FP: False Negative
No	FP: False Positive	TN: True Negative

So we can view $fminsupp$ as the rate of true positive example, $fmaxopp$ as the rate of true negative example, $fminopp$ as the rate of false positive example, $fmaxsupp$ as the rate of false negative example. Thus $fminopp$ is termed as positive error rate and $fmaxsupp$ is termed as negative error rate. $fcompsupp$ and $fentireopp$ represent the rate of presence and absence of items respectively. Also using different fuzzy confidence measures which are defined using fuzzy support measures, we can effectively explains the precision, specificity, sensitivity of each association rules which are discussed in the following sections.

5. Measures Based on Rule Prediction

A rule's prediction can be represented by a confusion matrix (Table 1). Studying its terminologies, we arrived at the following measures.

5.1 Fuzzy Support Measures

Definition 5.1 :

a) fuzzy complete support: It is defined as the rate of presence of items in the fuzzy association rule $\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle$ and is given by

$$fcompsupp(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle) = \frac{\sum\limits_{x\in T}(\tilde{F_A} \cup_{\mathcal{S}} F_B)(x) + \sum\limits_{x\in T}(F_A \cap_{\mathcal{T}} F_B)(x)}{|T|}$$

b) fuzzy complete opposition: It is defined as the rate of absence of items in the fuzzy association rule $\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle$ and is given by

$$fcompopp(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle) = \frac{\sum\limits_{x\in T}(\tilde{F_A} \cup_{\mathcal{S}} \tilde{F_B})(x) + \sum\limits_{x\in T}(F_A \cap_{\mathcal{T}} \tilde{F_B})(x)}{|T|}$$

(c) fuzzy accurate support: It is defined as the accuracy of a fuzzy association rule $\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle$ and is given by

$$faccuratesupp(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle) = \frac{\sum\limits_{x\in T}(\tilde{F_A} \cup_{\mathcal{S}} \tilde{F_B})(x) + \sum\limits_{x\in T}(F_A \cap_{\mathcal{T}} F_B)(x)}{|T|}$$

(b) fuzzy entire opposition: It is defined as the error rate of a fuzzy association rule $\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle$ and is given by

$$fentireopp(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle) = \frac{\sum\limits_{x\in T}(\tilde{F_A} \cup_{\mathcal{S}} F_B)(x) + \sum\limits_{x\in T}(F_A \cap_{\mathcal{T}} \tilde{F_B})(x)}{|T|}$$

Similarly we can define the corresponding count measures: $fcompsupp\#$, $fcompopp\#$, $fentireopp\#$, $fentireopp\#$.

5.2 Fuzzy Confidence Measures

Definition 5.2 : The precision of a fuzzy association rule $\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle$ is defined as a confidence measure, called fuzzy actual confidence which is given by

$$factconf(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle) = \frac{\sum\limits_{x\in T}(F_A \cap_{\mathcal{T}} F_B)(x)}{\Sigma_{x\in T}(F_A \cap_{\mathcal{T}} F_B)(x) + \Sigma_{x\in T}(F_A \cap_{\mathcal{T}} \tilde{F_B})(x)}$$

Remark 5.3 : The equalities in Definition 3.3 for confidence is not automatically transferred into fuzzy case, since

$$\Sigma_{x\in T}(F_A \cap_{\mathcal{T}} F_B)(x) + \Sigma_{x\in T}(F_A \cap_{\mathcal{T}} \tilde{F_B})(x) = \Sigma_{x\in T} F_A(x)$$

does not always hold.

Definition 5.4 : The fuzzy confidence measure m-confidence of a fuzzy assocaition rule $\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle$ is defined as:

$$fconf_m(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle) = \frac{fmaxsupp\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)}{fmaxopp\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)} \tag{10}$$

Definition 5.5 : The sensitivity or true positive rate of a fuzzy assocaition rule $\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle$ is termed as fuzzy true confidence and is given by

$$fconf_T(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle) = \frac{fminsupp\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)}{fcompsupp\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)} \tag{11}$$

Definition 5.6 : The false positive rate of a fuzzy assocaition rule $\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle$ is termed as fuzzy false confidence and is given by

$$fconf_F(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle) = \frac{fminopp\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)}{fcompopp\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)} \tag{12}$$

Definition 5.7 : The specificity of a fuzzy assocaition rule $\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle$ is termed as fuzzy specific confidence and is given by

$$fconf_{SP}(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle) = \frac{fmaxopp\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)}{fcompopp\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)} \tag{13}$$

Definition 5.8 : The interest of a fuzzy assocaition rule $\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle$ is given by

$$interest(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle) = \frac{factconf\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)}{fcompsupp\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)} \tag{14}$$

Definition 5.9 : The conviction of a fuzzy assocaition rule $\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle$ is given by

$$conviction = \frac{fconf_n\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle) fentireopp\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)}{factconf\#(\langle A, F_A\rangle \rightharpoonup \langle B, F_B\rangle)} \quad (15)$$

6. Conclusion

In this paper we extend some quality measures defined for crisp association rule to fuzzy association rule. Here we study about the transaction types and redefined their terminologies as true positive, true negative, false positive, false negative examples to understand the true semantics of the transactions. Based on that we defined various support and confidence measures and extended to fuzzy complete support, fuzzy accurate support, fuzzy complete opposition, fuzzy entire opposition measures. We also analyzed the semantics of these measures and defined some confidence measures using fuzzy operators and identified them with precision, specificity, sensitivity, interest, conviction of association rules.

Acknowledgment

This research was supported by the Kerala State Council for Science, Technology and Environment (KSCSTE).

References

[1] Agrawal R., Imielinski T. and Swami A., Mining association rules between sets of items in large databases, In Proceedings of the International Conference on Management of Data, Washington, D.C., (May 1993), 207-216.

[2] Agrawal R., Srikant R., Fast algorithms for mining association rules, In Proc. 20th Int. Conf. Very Large Data Bases, VLD, Santiago, Chile, (1994), 487-499.

[3] Bakk. Likas Helm, Fuzzy Association Rules:an implementation in R, Vienna University of Economics and Buiseness Administration, (2007), Vienna.

[4] Chan Keith C. C., Au Wai-Ho, An effective algorithm for discovering fuzzy rules in relational databases, In Proceedings of Fuzzy Systems, IEEE World Congress on Computational Intelligence, (1998).

[5] De Cock M., Cornelis C., Kerre E. E., Fuzzy association rules: a two-sided approach, In: Y. Liu, G. Chen and K. Y. Cai (Eds.): Proceedings of FIP2003

(International Conference on Fuzzy Information processing: Theories and Applications), Tsinghua Univ. Press, (2003), 385-390.

[6] De Cock M., Cornelis C., Kerre E. E., A clear view on quality measures for fuzzy association rules, In Proceedings of Int. Conf. on Fuzzy Sets and Soft Computing in Economics and Finance, St. Petersburg, (2004), 54-61.

[7] De Cock M., Cornelis C., Kerre E. E., Elicitation of fuzzy association rules from positive and negative examples, Fuzzy Sets and Systems, 149 (2005), 73-85.

[8] Dubois D., Hullermeier E. and Prade H., A Note on Quality Measures for Fuzzy Association Rules, In Proceedings IFSA-03, 10th International Fuzzy Systems Association World Congress, Lecture Notes in Articial Intelligence, number 2715, Springer-Verlag, (2003), 346-353.

[9] George J. Klir and Bo Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Pearson Education, (1995), USA.

[10] H*ü*llermeier H., Fuzzy association rules: Semantic issues and quality measures, In: B. Reusch (Ed.) Proceedings of the International Conference, 7th Fuzzy Days on Computational Intelligence, Theory and Applications, Lecture notes in Computer Science, number 2606, 380-391, Springer-Verlag.

[11] Hong T. P., Lin K. Y., Chien B. C., Mining fuzzy multiple-level association rules from quantitative data, Applied Intelligence, 18(1) (2003), 79-90.

[12] Iancu I., Gabroveanu M., Giurca A., A pair of confidence measures for association rules, In: 30th Annual Conference of the German Classification Society, GfKl 2006, Berlin, (March 8-10, 2006) (in press).

[13] Kuok C., Fu A., Wong M., Mining fuzzy association rules in databases, SIGMOD Record, 27(1) (1998), 41-46.

[14] Li-Xin Wang, A Course in Fuzzy Systems and Control, Prentice Hall International, Inc., (1997), Hong Kong.

[15] http://www.dataschool.io/simple-guide-to-confusion-matrix-terminology.