Classification with evidential associative rules

Ahmed Samet, Eric Lefevre, Sadok Ben Yahia

Classification with evidential associative rules

Ahmed Samet[1,2], Eric Lefèvre[2], and Sadok Ben Yahia[1]

[1] Laboratory of research in Programming, Algorithmic and Heuristic Faculty of Science of Tunis, Tunisia
{ahmed.samet, sadok.benyahia}@fst.rnu.tn
[2] Univ. Lille Nord de France UArtois, EA 3926 LGI2A, F-62400, Béthune, France
eric.lefevre@univ-artois.fr

Abstract. Mining database provides valuable information such as frequent patterns and especially associative rules. The associative rules have various applications and assets mainly data classification. The appearance of new and complex data support such as evidential databases has led to redefine new methods to extract pertinent rules. In this paper, we intend to propose a new approach for pertinent rule's extraction on the basis of *confidence* measure redefinition. The confidence measure is based on conditional probability basis and sustains previous works. We also propose a classification approach that combines evidential associative rules within information fusion system. The proposed methods are thoroughly experimented on several constructed evidential databases and showed performance improvement.

Keywords: Evidential database, Confidence, Associative classification, Evidential Apriori

1 Introduction

Data mining domain allows extracting pertinent information within databases [1]. The provided information are represented in a set of rules, where each one is associated with a pertinence measure denoted *Confidence*. Among their purposes, those *associative rules* are used for data classification [2, 3]. The classification process from those associative rules is denoted *associative classification*. Associative classification offers one of the best classification rate and measure membership [3]. Recently, new databases have appeared proposing data suffering from imperfection. Those types of data fit reality where opinions are no longer represented with Boolean values. In addition, it has added more complexity in their treatment. The imperfection is handled with several theories such as fuzzy [4] and evidential theory [5, 6]. In [7], the author introduced a new type of databases that handle both imprecise and uncertain information thanks to the evidential theory. Those types of databases were denoted as the *Evidential database*. The evidential databases were shortly studied from a data mining view [8] and not so much attention was paid to that issue. In literature, two major works [8, 9] stand by proposing new measures for itemsets' support. Indeed, in [8], Hewawasam et al. proposed a methodology to estimate itemsets' support

and modelize them in a tree representation: *Belief Itemset Tree* (BIT). The BIT representation brings easiness and rapidity for the estimation of the associative rule's confidence. In [9], the authors introduced a new approach for itemset support computing and applied on a Frequent Itemset Maintenance (FIM) problem. Only [8] paid attention to associative classification where the authors introduced *evidential associative rules*. A new measure for rule's confidence was introduced based on conditional belief [6]. In this work, evidential data mining problem is tackled by putting our focus on the associative classification. We highlight problems existing in current measure of evidential rule's confidence which are based on conditional belief. A new confidence measure is proposed based on Bayesian assumption. We also introduce a new associative classification method that reduces the overwhelming number of generated rules. The retained rules are then used for classification purposes and tested on several benchmarks. This paper is organized as follows: in section 2, the main principles of the evidential database are recalled. In section 3, several state of art works on confidence measure are scrutinized and we highlight their limits. In addition, we introduce an alternative confidence measure based on probabilistic definitions. In section 4, we introduce a new method for evidential rule generation. The provided rules are filtrated and combined through a fusion system. The performance of this algorithm is studied in section 5. Finally, we conclude and we sketch issues of future work.

2 Evidence database concept

An evidential database stores data that could be perfect or imperfect. Uncertainty in such database is expressed via the evidence theory [5, 6]. An evidential database, denoted by $\mathcal{EDB}$, with n columns and d lines where each column i ($1 \leq i \leq n$) has a domain θ_i of discrete values. Cell of line j and column i contains a normalized BBA as follows:

$$m_{ij} : 2^{\theta_i} \rightarrow [0, 1] \quad with$$

$$\begin{cases} m_{ij}(\emptyset) = 0 \\ \sum\limits_{A \subseteq \theta_i} m_{ij}(A) = 1. \end{cases} \tag{1}$$

Table 1. Evidential transaction database $\mathcal{EDB}$

Transaction	Attribute A	Attribute B
T1	$m_{11}(A_1) = 0.7$	$m_{21}(B_1) = 0.4$
	$m_{11}(\theta_A) = 0.3$	$m_{21}(B_2) = 0.2$
		$m_{21}(\theta_B) = 0.4$
T2	$m_{12}(A_2) = 0.3$	$m_{22}(B_1) = 1$
	$m_{12}(\theta_A) = 0.7$	

In an evidential database, as shown in Table 1, an item corresponds to a focal element. An itemset corresponds to a conjunction of focal elements having different domains. Two different itemsets can be related via the inclusion or the intersection operator. Indeed, the inclusion operator for evidential itemsets [9] is defined as follows, let X and Y be two evidential itemsets:

$$X \subseteq Y \iff \forall x_i \in X, x_i \subseteq y_i.$$

where x_i and y_i are the i^{th} element of X and Y. For the same evidential itemsets X and Y, the intersection operator is defined as follows:

$$X \cap Y = Z \iff \forall z_i \in Z, z_i \subseteq x_i \text{ and } z_i \subseteq y_i.$$

An *Evidential associative rule* R is a causal relationship between two itemsets that can be written in the following form $R : X \rightarrow Y$ fulfilling $X \cap Y = \emptyset$. In Table 1, A_1 is an item and $\{\theta_A\ B_1\}$ is an itemset such that $A_1 \subset \{\theta_A\ B_1\}$ and $A_1 \cap \{\theta_A\ B_1\} = A_1$. $A_1 \rightarrow B_1$ is an evidential associative rule.

Several definitions for the support estimation were defined for the evidential itemsets such as [8, 9]. Those methods assess the support based on the belief function applied on the evidential database BBA $m_{\mathcal{EDB}}$[3] :

$$Support_{\mathcal{EDB}}(X) = Bel_{\mathcal{EDB}}(X) \tag{2}$$

such that:

$$Bel : 2^{\theta} \rightarrow [0, 1] \tag{3}$$

$$Bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B). \tag{4}$$

In a previous work [10], we introduced a new metric for support estimation providing more accuracy and overcoming several limits of using the belief function. The Precise support Pr is defined by:

$$Pr : 2_i^{\theta} \rightarrow [0, 1] \tag{5}$$

$$Pr(x_i) = \sum_{x \subseteq \theta_i} \frac{|x_i \cap x|}{|x|} \times m_{ij}(x) \qquad \forall x_i \in 2^{\theta_i}. \tag{6}$$

The evidential support of an itemset $X = \prod_{i \in [1 \ldots n]} x_i$ in the transaction T_j (i.e., Pr_{T_j}) is then computed as follows:

$$Pr_{T_j}(X) = \prod_{x_i \in \theta_i, i \in [1 \ldots n]} Pr(x_i) \tag{7}$$

Thus, the evidential support $Support_{\mathcal{EDB}}$ of the itemset X becomes:

$$Support_{\mathcal{EDB}}(X) = \frac{1}{d} \sum_{j=1}^{d} Pr_{T_j}(X). \tag{8}$$

[3] A BBA constructed from Cartesian product applied on the evidential database. Interested readers may refer to [8].

3 Confidence measure for evidential associative rules

The confidence is the measure assigned to the associative rules and it represents its relevance [1]. As originally introduced in Boolean databases, the confidence measure was relying on conditional probability [1]. Indeed for a rule $R : R_a \rightarrow R_c$, such that R_c and R_a are respectively the conclusion and the antecedent (premise) part of the rule R, the confidence is expressed as follows:

$$Confidence(R) = P(R_c|R_a) = \frac{\sum_{i=1}^{d} P(R_a \cap R_c)}{\sum_{i=1}^{d} P(R_a)} \tag{9}$$

In addition, even in fuzzy data mining, the associative rule's confidence is built with conditional fuzzy measures [11]. In this respect, evidential associative rules were initially introduced in [8]. The authors defined the structure of an evidential associative rule and estimated its relevance following a confidence metric. The confidence of a rule R in the set of all rules $\mathcal{R}$, i.e., $R \in \mathcal{R}$, is computed as follows:

$$Confidence(R) = Bel(R_c|R_a) \tag{10}$$

where $Bel(\bullet|\bullet)$ is the conditional Belief. The proposed confidence metric is hard to define where several works have tackled this issue and different interpretations and formulas were proposed such as those given respectively in [5, 12]. In [5], the conditional belief is defined as follows:

$$Bel(R_c|R_a) = \frac{Bel(R_c \cup \overline{R_a}) - Bel(\overline{R_a})}{1 - Bel(\overline{R_a})} \tag{11}$$

In [8], the authors used Fagin et al.'s conditional belief such that:

$$Bel(R_c|R_a) = \frac{Bel(R_a \cap R_c)}{Bel(R_a \cap R_c) + Pl(R_a \cap \bar{R_c})}. \tag{12}$$

where $Pl()$ is the plausibility function and is defined as follows:

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B). \tag{13}$$

Example 1. Through the following example, we highlight the inadequacy of the conditional belief use. We consider the Transaction 1 of Table 1 from which we try to compute the confidence of $A_2 \rightarrow B_1$ (i.e., $Bel(B_1|A_2)$). The conditional belief introduced in [5] gives the following results:

$$Bel(B_1|A_2) = \frac{Bel(B_1 \cup \overline{A_2}) - Bel(\overline{A_2})}{1 - Bel(\overline{A_2})} = \frac{Bel(B_1)}{1} = 0.4$$

The result of the belief of B_1 knowing A_2 is true is equal to that of $Bel(B_1)$ due to the independence between A_2 and B_1. On the other hand, both hypothesis might be correlated so that the event B_1 does not occur knowing already the happening of A_2.

In the following, we propose a new metric for the confidence estimation based on our Precise support measure [10] and probability assumption:

$$Confidence(R) = \frac{\sum_{j=1}^{d} Pr_{T_j}(R_a) \times Pr_{T_j}(R_c)}{\sum_{j=1}^{d} Pr_{T_j}(R_a)} \tag{14}$$

where d is the number of transactions in the evidential database. Thanks to its probabilistic writing, the proposed metric sustains previous confidence measure such as that introduced in [1].

Example 2. Let us consider the example of the evidential database in Table 1. The confidence of the evidential associative rule $R_1 : A_1 \rightarrow B_1$ is computed as follows:

$$Confidence(R_1) = \frac{Pr_{T_1}(A_1) \times Pr_{T_1}(B_1) + Pr_{T_2}(A_1) \times Pr_{T_2}(B_1)}{Pr_{T_1}(A_1) + Pr_{T_2}(A_1)} = 0.75$$

The generated rules with their confidence could find several applications. In the following, we tackle the classification problem case and a based evidential rule classifier is introduced.

4 Associative Rule Classifier

One of the main characteristics of the evidential database is the great number of items that it integrates. The number of items depends from the frame of discernment of each column. This asset makes from the evidential database more informative but more complex than the usual binary database. In [10], we have shown the significant number of generated frequent patterns that may be drawn even from small databases. Indeed, from a frequent itemset, of size k, $2^k - 2$ potential rules are generated. In order to use the generated evidential rules for a classification purposes, we first have to reduce their number for a more realistic one. In the following, we propose two processes for classification rule's reduction.

4.1 Classification rules

From the obtained rules, we retain only the classification ones. From a rule such that $\prod_{i \in I} X_i \rightarrow \prod_{j \in J} Y_j$, we only keep those matching a class hypothesis at the conclusion part (i.e., $Y_j \in \theta_C$ and θ_C is the frame of discernment).

Example 3. Let us consider the following set of the association rules $S = \{A_1 \rightarrow C_1; A_1, B_2 \rightarrow C_1; A_1 \rightarrow B_1\}$ and the class frame of discernment $\theta_C = \{C_1, C_2\}$. After classification rule reduction, the set S becomes $S = \{A_1 \rightarrow C_1; A_1, B_2 \rightarrow C_1\}$.

4.2 Generic and Precise rules

Generic rules: the rule's reduction can assimilate the redundant rules. A rule R_1 is considered as a redundant rule if and only if it does not bring any new information having at hand a rule R_2. R_2 is considered as more informative as far as its antecedent part is included in that of R_1. The retained rules from the reduction process constitute the set of Generic rules $\mathcal{R}$ extracted from the set of frequent itemsets $\mathcal{FI}$.

Example 4. Let us consider the previous set of the association rules $S = \{A_1 \rightarrow C_1; A_1, B_2 \rightarrow C_1; A_1 \rightarrow B_1\}$. After redundant rule reduction, the set S becomes $S = \{A_1 \rightarrow C_1; A_1 \rightarrow B_1\}$.

Precise rules: A rule is considered as *precise* if the rule's premise is maximized. Thus, from the set of all possible rules, we retain only those having the size of their premise part equal to n (number of columns of $\mathcal{EDB}$).

Algorithm 1 sketches the process of rule's generation as well as rule reduction. The algorithm relies on the function $Construct\_Rule(x, \theta_C)$ (Line 10) that generates associative rules and filtrates out them by retaining only the classification ones. The function $Find\_Confidence(R, Pr\_Table)$ (Line 22) computes the confidence of the rule R following the Pr_Table that contains all transactional support of each item (for more details see [10]). Finally, the function $Redundancy(\mathcal{R}, R)$ (Line 42) builds the set of all classification rules $\mathcal{R}$ which are not redundant and having the confidence value greater than or equal to the fixed threshold $minconf$.

4.3 Classification

Let us suppose the existence of an instance X to classify represented a set of BBA belonging to the evidential database $\mathcal{EDB}$ such that:

$$X = \{m_i | m_i \in X, x_i^j \in \theta_i\} \tag{15}$$

where x_i^j is a focal element of the BBA m_i. Each retained associative rule, in the set of rules $\mathcal{R}$, is considered as a potential piece of information that could be of help for X class determination. In order to select rules that may contribute to classification, we look for rules having a non null intersection with X such that:

$$\mathcal{RI} = \{R \in \mathcal{R}, \exists x_i^j \in \theta_i,\ x_i^j \in R_a\} \tag{16}$$

Each rule found in the set $\mathcal{RI}$ constitutes a piece of information concerning the instance X membership. Since several rules can be found and fulfilling the

Algorithm 1 Evidential Associative Rule Generation algorithm

Require: $Pr\_Table, minconf, \mathcal{FI}, \theta_C$
Ensure: $\mathcal{R}$
for all $x \in \mathcal{FI}$ **do**
  $R \leftarrow Construct\_Rule(x, \theta_C)$
  if $R \neq \emptyset$ **then**
    $Conf \leftarrow Find\_Confidence(R, Pr\_Table)$
    if $Conf > minconf$ **then**
      $\mathcal{R} \leftarrow Redundancy(\mathcal{R}, R)$
    end if
  end if
end for
function CONSTRUCT_RULE(X , θ_C)
  for all $x \in X$ **do**
    if $x \notin \theta_C$ **then**
      $prem \leftarrow prem + \{x\}$
    else
      $concl \leftarrow concl + \{x\}$
    end if
  end for
  $R.premise \leftarrow prem$
  $R.conclusion \leftarrow concl$
  return R
end function
function FIND_CONFIDENCE(R, Pr)
  $numer \leftarrow 0$
  $denom \leftarrow 0$
  for j=1 to d **do**
    $num \leftarrow 1$
    $den \leftarrow 1$
    for all $i \in Pr(j).focal\_element$ **do**
      if $Pr(j).focal\_element \in R.premise$ **then**
        $num \leftarrow num \times Pr(j).val$
        $den \leftarrow den \times Pr(j).val$
      else
        if $Pr(j).focal\_element \in R.conclusion$ **then**
        end if
      end if
    end for
    $numer \leftarrow numer + num$
    $denom \leftarrow denom + den$
  end for
  return $\frac{numer}{denom}$
end function
function REDUNDANCY($\mathcal{R}$,R)
  for all $rule \in \mathcal{R}$ **do**
    if $R.premise \subset rule.premise$ & $R.conclusion = rule.conclusion$ **then**
      $\mathcal{R} \leftarrow \mathcal{R} \backslash rule$
      $\mathcal{R} \leftarrow \mathcal{R} \cup R$
    end if
  end for
  return $\mathcal{R}$
end function

intersection condition, it is of importance to benefit from them all. In our work, we assume that all information is valuable and should be handled within the information fusion problem. From the set $\mathcal{RI}$, we extract the set of generic or precise classification rules (see Subsection 4.2). Indeed, each rule from the computed set $R^l \subset \mathcal{RI}$, $l \in [1 \dots L]$ and $L < |\mathcal{RI}|$, that brings a new information (different R_a) is transformed into a BBA following the frame of discernment θ_C (frame of discernment of R_c):

$$\begin{cases} m_{R^l}^{\theta_C}(\{R_c\}) = \alpha \times confidence(R^l) \\ m_{R^l}^{\theta_C}(\theta_C) = 1 - (\alpha \times confidence(R^l)) \end{cases} \tag{17}$$

where R_c is the conclusion part of the rule R^l and $\alpha \in [0, 1]$ is a discounting factor.

The L constructed BBA are then fused following the Dempster rule of combination [5] as follows:

$$m_{\oplus} = \oplus_{l=1}^{L} m_{R^l}^{\theta_C}. \tag{18}$$

$\oplus$ is the Dempster's aggregation function where for two source's BBA m_1 and m_2:

$$\begin{cases} m_{\oplus}(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B) \cdot m_2(C) & \forall A \subseteq \Theta, A \neq \emptyset \\ m_{\oplus}(\emptyset) = 0 \end{cases} \tag{19}$$

where K is defined as:

$$K = \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C). \tag{20}$$

5 Experimentation and results

In this section, we present how we managed to conduct our experiments and we discuss comparative results.

5.1 Evidential database construction

In order to perform experimental tests, we construct our own evidential databases from UCI benchmarks [13] based upon ECM [14]. Interested reader may refer to [10] for more details on evidential database construction. The transformation was operated on *Iris*, *Vertebral Column*, *Diabetes* and *Wine* databases. The studied databases are summarized on Table 2 in terms of number of instances and attributes.

Table 2. Database characteristics

Database	**#Instances**	**#Attributes**	**#Focal elements**
Iris_EDB	150	5	40
Vertebral Column_EDB	310	7	116
Diabetes_EDB	767	9	132
Wine_EDB	178	14	196

5.2 Comparative results

In the following, we compare the classification result performance between the Generic and Precise rules. Table 3 shows the difference in classification result between the generic and the precise associative rules. The precise rules highlight better results than do the generic ones. Indeed, the larger the rule's premise is, the more pertinent the rule is. On the other hand, the generic rule based

approach fuse much more rules than do the precise one. In addition, all generic rules are considered with the same weight within the fusion process despite their pertinence difference. These characteristics with Dempster's combination behavior mislead the fusion process to errors. Indeed, as shown in Figure 1, the high number of fused rules depends highly from the *minsup* value. Unlike the generic approach, the number of precise rule is defined by number of larger premise's rule which is dependent from the treated evidential transaction.

Table 3. Comparative result between Generic and Precise classification rules

Database	Iris_EDB	Vertebral Column_EDB	Diabetes_EDB	Wine_EDB
Precise rules	80.67%	88.38%	83.20%	100%
Generic rules	78.67%	67.74%	65.10%	51.68%

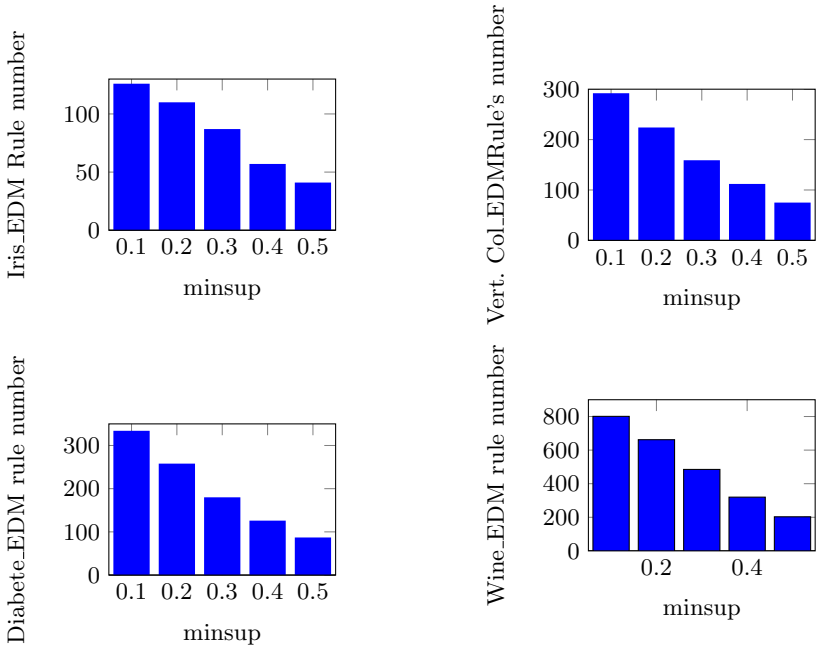

Fig. 1. Generic associative rule's number for different support values

6 Conclusion

In this paper, we tackled associative rule's extraction from evidential databases. We proposed a new confidence measure for associative rules in evidential databases.

The proposed measure is based on Precise support (i.e., probability measure) providing coherence and sustains previous work on fuzzy and binary databases. The rules are then filtrated to retain only classification and non redundant rules. A classification method based on evidential associative rules is introduced. The classification approach is based on a fusion system that represents interesting rules. As illustrated in the experimentation section, the proposed method provides an interesting performance rates. In future work, we plan to study the development of a new method to estimate the reliability of each combined associative rule. Indeed, each rule has a precision relatively to the instance to classify. The precision is measured by the intersection between the premise and the instance itself. A reliability measure for rule BBA is under study.

References

1. R. Agrawal, R. Srikant, Fast algorithm for mining association rules, In Proceedings of international conference on Very Large DataBases, VLDB, Santiago de Chile, Chile (1994) 487–499.
2. W. Li, J. Han, J. Pei, CMAR: Accurate and efficient classification based on multiple class-association rules, in Proceedings of IEEE International Conference on Data Mining (ICDM01), San Jose, CA, IEEE Computer Society (2001) 369–376.
3. I. Bouzouita, S. Elloumi, S. Ben Yahia, GARC: A new associative classification approach, In Proceedings of 8th International Conference on Data Warehousing and Knowledge Discovery, DaWaK, Krakow, Poland (2006) 554–565.
4. L. A. Zadeh, Fuzzy sets, Information and Control 8 (3) (1965) 338–353.
5. A. Dempster, Upper and lower probabilities induced by multivalued mapping, AMS-38, 1967.
6. G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
7. S. Lee, Imprecise and uncertain information in databases: an evidential approach, In Proceedings of Eighth International Conference on Data Engineering, Tempe, AZ (1992) 614–621.
8. K. K. R. Hewawasam, K. Premaratne, M.-L. Shyu, Rule mining and classification in a situation assessment application: A belief-theoretic approach for handling data imperfections, Trans. Sys. Man Cyber. Part B 37 (6) (2007) 1446–1459.
9. M. A. Bach Tobji, B. Ben Yaghlane, K. Mellouli, Incremental maintenance of frequent itemsets in evidential databases, In Proceedings of the 10th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Verona, Italy (2009) 457–468.
10. A. Samet, E. Lefevre, S. Ben Yahia, Mining frequent itemsets in evidential database, In Proceedings of the fifth International Conference on Knowledge and Systems Engeneering, Hanoi, Vietnam (2013) 377–388.
11. T.-P. Hong, C.-S. Kuo, S.-L. Wang, A fuzzy AprioriTid mining algorithm with reduced computational time, Applied Soft Computing 5 (1) (2004) 1–10.
12. R. Fagin, J. Y. Halpern, A new approach to updating beliefs, Uncertainty in Artificial Intelligence (1991) 347–374.
13. A. Frank, A. Asuncion, UCI machine learning repository (2010), URL: http://archive.ics.uci.edu/ml.
14. M.-H. Masson, T. Denœux, ECM: An evidential version of the fuzzy c-means algorithm, Pattern Recognition 41 (4) (2008) 1384–1397.