Mining frequent itemsets in evidential database

Ahmed Samet, Eric Lefevre, Sadok Ben Yahia

Mining frequent itemsets in evidential database

Ahmed Samet[1], 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 frequent patterns is widely used to discover knowledge from a database. It was originally applied on Market Basket Analysis (MBA) problem which represents the Boolean databases. In those databases, only the existence of an article (item) in a transaction is defined. However, in real-world application, the gathered information generally suffer from imperfections. In fact, a piece of information may contain two types of imperfection: imprecision and uncertainty. Recently, a new database representing and integrating those two types of imperfection were introduced: Evidential Database. Only few works have tackled those databases from a data mining point of view. In this work, we aim to discuss evidential itemset's support. We improve the complexity of state of art methods for support's estimation. We also introduce a new support measure gathering fastness and precision. The proposed methods are tested on several constructed evidential databases showing performance improvement.

Keywords: Evidential database, Support, Frequent evidential item, Evidential Apriori

1 Introduction

The majority of data mining algorithms were applied on precise and certain data constituting Boolean databases. This type of databases does only indicate if the considered item I exists or not. However, in real life, gathered information are suffering from imperfection due to many factors such as acquisition reliability, human errors, information absence, etc. In [1], Lee detailed the two sides of imperfection that could manifest in a database. Indeed, we may encounter databases containing *imprecise* and *uncertain* information. The imprecision is relevant to the content of an attribute value of a data object, while the concept of uncertainty is relevant to the degree of truth of its attribute value. Due to its adequate imprecision representation, the fuzzy theory [2] was largely used to extract fuzzy frequent patterns and association rules such that [3–5]. In [1], the author introduced a new type of databases that handle both imprecise and uncertain information. This database was modeled via the evidence theory [6, 7], which offers a certain level of flexibility in imperfect information representation.

Those types of databases were denoted as the *Evidential database*. The Evidential database has brought flexibility in handling those imperfect knowledge but also added complexity in their treatment. Indeed, the number of patterns increases exponentially as far as the number of attributes arises in the database. Even the literature methodologies for estimating itemset's support are time consumer, since they rely on Cartesian product. In addition to the time limit, the proposed support functions such as those introduced in [8, 9], are not that precise in their estimation. The support computing do not explore all information that exist in the *Basic Belief Assignment*. This constraint makes from literature methods limited in their manner of support estimation and do not extract all existing frequent patterns within the evidential database. In this work, evidential data mining problem is tackled by putting our focus on the support estimation. Existing methods for support estimation are highlighted and we propose a ramification that considerably improves their original performance. We also introduce a new measure for evidential pattern's support estimation. This new measure improves the support computation where all pieces of information in a Basic Belief Assignment are considered. This method also presents an interesting performance comparatively to literature methods. We introduce the Evidential Data mining Algorithm (EDMA) that mines all frequent patterns in an evidential database. This paper is organized as follows: in section 2, the main principles of the evidence theory and Smets's TBM [10] interpretation are presented. In section 3, several state of art works are scrutinized and we highlight their limits. We present a ramification for their method that improves the performance. In section 4, we introduce a new method for evidential itemsets' support computing providing more precision in its estimation. Evidential Data Mining Apriori (EDMA) algorithm for mining frequent evidential patterns is introduced in section 5. The performance of this algorithm is studied in section 6. Finally, we conclude and we sketch issues of future work.

2 Evidence database

In this section, evidential database concepts based on evidence theory formalism are presented. In the following, we define evidence theory main concepts based on a Transferable Belief Model interpretation [10].

2.1 Evidence theory

The evidence theory or Dempster-Shafer theory proposes a robust formalism for modeling uncertainty. In the following, the evidence theory from a Smets's Transferable Belief Model (TBM) interpretation is presented. The TBM model represents quantified beliefs following two distinct levels: (i) a credal level where beliefs are entertained and quantified by belief functions; (ii) a pignistic level where beliefs can be used to make decisions and are quantified by probability functions. The evidence theory is based on several fundamentals such as the

Basic Belief Assignment (BBA). A BBA m is the mapping from elements of the power set 2^θ onto [0, 1]:

$$m : 2^\theta \longrightarrow [0, 1]$$

where θ is the *frame of discernment*. It is the set of possible answers for a treated problem and is composed of N exhaustive and exclusive hypotheses:

$$\theta = \{H_1, H_2, ..., H_N\}.$$

A BBA m do have some constraints such that:

$$\sum_{A \subseteq \theta} m(A) = 1 \tag{1}$$

Each subset X of 2^θ fulfilling $m(X) > 0$ is called focal element. Constraining $m(\emptyset) = 0$ is the normalized form of a BBA and this corresponds to a closed-world assumption [11], while allowing $m(\emptyset) > 0$ corresponds to an open world assumption [12].

From a BBA another function is commonly defined from 2^θ to $[0, 1]$: $Bel(A)$ is interpreted as the degree of justified support given to the proposition A by the available evidence.

$$Bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B) \tag{2}$$

Generally, in an information fusion problem, not all considered sources share the same domain (frame of discernment). This constraint prevents from using usual combination tools [6].

In this case, the Cartesian product allows the combination. Let it be two belief function m_1 and m_2 defined respectively in θ_1 and θ_2, the Cartesian product is expressed as follows:

$$m^\theta_{1\times 2}(A \times B) = m^{\theta_1}_1(A) \times m^{\theta_2}_2(B). \tag{3}$$

After source's combination which integrates the credal stage of the TBM model, taking decision is necessarily. In [10], the pignistic probability is introduced allowing probabilistic decision from BBA following this formula:

$$BetP(H_n) = \sum_{A \subseteq \theta} \frac{|H_n \cap A|}{|A|} \times m(A) \qquad \forall H_n \in \theta \tag{4}$$

where $|\cdot|$ is the cardinality operator.

2.2 Evidence database concept

An evidential database stores data that could be perfect or imperfect. Uncertainty in such database is expressed via the evidence theory. 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:

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

$$\begin{cases} m_{ij}(\emptyset) = 0 \\ \sum\limits_{A \subseteq \theta_i} m_{ij}(A) = 1. \end{cases} \tag{5}$$

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

Transaction	Attribute A	Attribute B
T1	$m(A_1) = 0.7$	$m(B_1) = 0.4$
	$m(\theta_A) = 0.3$	$m(B_2) = 0.2$
		$m(\theta_B) = 0.4$
T2	$m(A_2) = 0.3$	$m(B_1) = 1$
	$m(\theta_A) = 0.7$	

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 intersection operator. Indeed, the inclusion operator for evidential itemsets 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:

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

Example 1. 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 $A_1 \cap \{\theta_A\ B_1\} = A_1$.

3 Evidential patterns' Cartesian based support

In this section, we present related works in support estimation for evidential databases. Afterwards, we sketch with an improvement that we propose to improve support's performance.

3.1 State of art

Evidential data mining does not grasp so much attention. In [13], 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

[8], the authors introduced a new approach for itemset support computing and applied on a Frequent Itemset Maintenance (FIM) problem. All methods [8, 9] were based on Cartesian product between BBAs.

Let's study the support of an itemset $X = \prod_{i \in [1 \ldots n]} x_i$ such that x_i is an evidential item belonging to the frame of discernment θ_i. Since the items do not share the same discernment frame, any fusion rule cannot be applied. In the following, we study the belief support introduced by [8] computed by the following equation:

$$m_j(X) = \prod_{x_i \in X} m_{ij}(x_i) \tag{6}$$

where $m_j(X)$ is the Cartesian product of all BBA in the transaction T_j. Thus, the BBA of the itemset X expressed in the entire $\mathcal{EDB}$ database becomes:

$$m_{\mathcal{EDB}}(X) = \frac{1}{d} \sum_{j=1}^{d} m_j(X). \tag{7}$$

Then, the support of X in the $\mathcal{EDB}$ database becomes:

$$Support_{\mathcal{EDB}}(X) = Bel_{\mathcal{EDB}}(X). \tag{8}$$

The Cartesian product based support, presented above, fulfills several properties such that the *anti-monotony* property. A support measure satisfying the anti-monotony property consists in the fact that an itemset that contains an infrequent itemset is also infrequent. The opposite is true, all itemsets constituting a frequent one are also frequent. With this satisfied property, the construction of an Apriori based algorithm becomes straightforward [8].

3.2 Cartesian support ramification

The support measure, proposed by [8, 9] works (shown in equation 6), relies on Cartesian product. Indeed, the Cartesian product is the suited solution in case of combining BBAs with different frame of discernment. However, such solution waste execution time because of its exponential complexity. In this section, we focus our interest in simplifying the Cartesian based method for performance requirements.

Let us consider the evidential database $\mathcal{EDB}$ and the itemset $X = x_1 \times \cdots \times x_n$ constituted by the product of items (focal elements) x_i $(1 \leq i \leq n)$ of the exclusive frame of discernment θ_i. For a transaction T_j, we have:

$$Support_{T_j}(X) = \prod_{i \in [1 \ldots n]} Support_{T_j}(x_i) = \prod_{i \in [1 \ldots n]} Bel(x_i) \tag{9}$$

$$Support_{\mathcal{EDB}}(X) = \frac{1}{d} \sum_{j=1}^{d} Support_{T_j}(X) \tag{10}$$

Proof. Let us consider two evidential items and focal elements x_1 and x_2 belonging respectively to m_1 and m_2 BBA such that $m = m_1 \times m_2$.

$$
\begin{aligned}
Bel(\prod_{x_i \in \theta_i, 1 \leq i \leq n} x_i) &= \sum_{a \subseteq x_1 \times \cdots \times x_n} m_{1 \times \cdots \times n}(a) \\
Bel(\prod_{x_i \in \theta_i, 1 \leq i \leq n} x_i) &= \sum_{y_1 \subseteq x_1, \ldots, y_2 \subseteq x_n} m_1(y_1) \times \cdots \times m_n(y_2) \\
Bel(\prod_{x_i \in \theta_i, 1 \leq i \leq n} x_i) &= \sum_{y_1 \subseteq x_1} m_1(y_1) \times \cdots \times \sum_{y_n \subseteq x_n} m_n(y_n) \\
Bel(\prod_{x_i \in \theta_i, 1 \leq i \leq n} x_i) &= Bel(x_1) \times \cdots \times Bel(x_n) = \prod_{i \in [1 \ldots n]} Bel(x_i)
\end{aligned}
$$

In this section, the support is estimated with the $Bel(.)$ function which generates several limits. In the following, we highlight those limits and we propose a new support alternative: *The precise support.*

4 Precise Evidential support estimation

The evidential database relies on representing information's imperfection with BBAs. A BBA does not only represent belief accorded to a single hypothesis but also to their disjunction. As shown in section 2, from a piece of evidence (BBA), several functions exist allowing the pertinence's estimation of each hypothesis. The $Bel(.)$ (see equation 2), used for support definition in section 3, is not the only function that estimates the degree of veracity of each hypothesis in the superset 2^θ. In addition, $Bel(.)$ estimates the belief by referring only to a small subset of the superset. This limits make from belief based support measures imprecise. In the following, we propose a new alternative to the Cartesian support estimation allowing a precise support computing and a reasonable time scale performance.

4.1 Support Definition

Let us consider an evidential database $\mathcal{EDB}$ and the itemset $X = x_1 \times \cdots \times x_n$ constituted by the product of items (focal elements) x_i $(1 \leq i \leq n)$ of the exclusive frame of discernment θ_i. The degree of presence of an item x_i in a transaction T_j (BBA) can be measured as follow:

$$Pr : 2^\theta \rightarrow [0, 1] \tag{11}$$

$$Pr(x_i) = \sum_{x \subseteq \theta_i} \frac{|x_i \cap x|}{|x|} \times m(x) \qquad \forall x_i \in 2^{\theta_i}. \tag{12}$$

As illustrated above, the $Pr(.)$ measure allows to compute x_i presence in a single BBA. The Pr measure is equal to the pignistic probability if $x_i \in \theta_i$. The evidential support of an itemset $X = \prod_{i \in [1 \ldots n]} x_i$ is then computed as follows:

$$Support_{T_j}^{Pr}(X) = \prod_{X_i \in \theta_i, i \in [1 \ldots n]} Pr(x_i) \tag{13}$$

$$Support_{\mathcal{EDB}}(X) = \frac{1}{d} \sum_{j=1}^{d} Support_{T_j}^{Pr}(X). \tag{14}$$

The presented approach for estimating itemset evidential support is similar to the support ramification given subsection 3.2. However, our approach presents several assets where our support inclusion is larger than those given in respectively [8, 9]. Indeed, in the previous cited works, the authors evaluate support of X by considering only subsets included in it. The $Pr(.)$ function does not only consider all subsets of X but also those having intersection with it. In addition, our support estimation provides an interesting performance since we get rid of the Cartesian product. The proposed support function sustains previous works on fuzzy [3] in case of dealing with consonant BBA[3]. It also sustains previous data mining works on binary databases [14] when BBA are certain[4]. Indeed, previous works have adopted a probabilistic orientation in support measure in those databases. Support generally represents the frequency of appearance of an itemset therefore it can be assimilated to an apriori probability. Interestingly enough, the precise support estimation function keeps the interesting anti-monotony property useful in infrequent itemsets removal. This fulfilled condition is proven in the proof given below.

Proof. Assuming an evidential database $\mathcal{EDB}$, let's consider two evidential itemsets A and $A \times X$ where $A \subset A \times X$ such that $\forall x \in A,\ x \in A \times X$. We aim to prove that considering this condition $Support(A \times X) \leq Support(A)$.

$Support_{T_j}(A \times X) = Pr(A) \times Pr(X)$
$Support_{T_j}(A \times X) \leq Support_{T_j}(A)$ Since $Pr(X) \in [0, 1]$ then
$Support_{\mathcal{EDB}}(A \times X) \leq Support_{\mathcal{EDB}}(A)$

4.2 Pr Table

Let us consider be the evidential database $\mathcal{EDB}$ containing n attributes and d transactions. The *Pr Table* is a table having d rows ($j \in [1, d]$) where each one contains the $Pr(.)$ measure of all items (focal elements) found in the j^{th} transaction of $\mathcal{EDB}$. Since the support function can be written as a simple product, the storage of item's Pr measure in a table became a need. Table 2 shows the Pr Table extracted from the evidential database $\mathcal{EDB}$ presented in Table 1.

[3] A BBA is said consonant if focal elements are nested.
[4] A BBA with only one focal element A is said to be certain and is denoted $m(A) = 1$.

Table 2. Pr Table deduced from the evidential database $\mathcal{EDB}$ presented in Table 1

Transaction	Transactional Support
T1	$Pr^{\theta_A}(A_1) = 0.85$
	$Pr^{\theta_A}(A_2) = 0.15$
	$Pr^{\theta_A}(\theta_A) = 1.00$
	$Pr^{\theta_B}(B_1) = 0.60$
	$Pr^{\theta_B}(B_2) = 0.40$
	$Pr^{\theta_B}(\theta_B) = 1.00$
T2	$Pr^{\theta_A}(A_1) = 0.35$
	$Pr^{\theta_A}(A_2) = 0.65$
	$Pr^{\theta_A}(\theta_A) = 1.00$
	$Pr^{\theta_B}(B_1) = 1.00$
	$Pr^{\theta_B}(B_2) = 0.00$
	$Pr^{\theta_B}(\theta_B) = 1.00$

5 Evidential Data Mining Apriori: EDMA

In this section, we introduce the Evidential Data Mining Apriori (EDMA) algorithm that allows the extraction of all frequent itemsets. Each itemset having a support greater than a threshold *minsup* is considered as frequent and is retained. The proposed algorithm relies on Apriori algorithm basics [14].

Apriori exploits this assumption by generating frequents in a level-wise manner. First of all, it generates frequent items (level 1) by removing those candidates (items) that do not fulfill the *minsup* constraint. From the generated frequent, it seeks to find the frequent of the next level by composing those of the precedent level. The treatment comes to an end when no further frequent itemset can be generated.

Algorithm 1 sketches the EDMA algorithm where it generates $\mathcal{EIFF}$ the set of all evidential frequent itemset. Their determination is based on support measure and *minsup* constraint fulfillment. This test is performed by *Frequent_itemset* function. The support is computed through the *Support_estimation* function. As it is shown, the support estimation does not rely anymore on calculating the Cartesian product of BBAs but on stored item's precise measure (Pr Table). Since the algorithm sweeps the search space in breadth first manner, EDMA generates candidates (i.e., *candidate_apriori_gen* function) from the frequent itemset of the previous level.

EDMA is an Apriori based algorithm that extract frequent evidential itemsets. It also generalizes several other known mining algorithms. In case of having only categorical items (categorical BBA), the database can be viewed as a binary transaction database. In this case, the proposed EDMA approach for evidential databases matches the original Apriori algorithm for binary databases [14]. Since Evidential database represents imprecision and uncertainty, it assimilates fuzzy databases via consonant BBA. The EDMA approach also assimilates other fuzzy Apriori algorithms as [3].

Algorithm 1 Evidential Data Mining Apriori (EDMA) algorithm

Require: $\mathcal{EDB}, minsup, PT, Size\_\mathcal{EDB}$
Ensure: $\mathcal{EIFF}$
function SUPPORT_ESTIMATION(PT , I, d)
  $Sup_I \leftarrow 0$
  for j=1 to d **do**
    $Sup_{Trans} \leftarrow 1$
    for all $i \in Pr(j).focal\_element$ **do**
      if $Pr(j).focal\_element \in I$ **then**
        $Sup_{Trans} \leftarrow Sup_{Trans} \times Pr(j).value$
      end if
    end for
    $Sup_I \leftarrow Sup_I + Sup_{Trans}$
  end for
  return $\frac{Sup_I}{d}$
end function
function FREQUENT_ITEMSET($candidate$, $minsup$, PT, $Size\_\mathcal{EDB}$)
  $frequent \leftarrow \emptyset$
  for all x in $candidate$ **do**
    if $Support\_estimation(PT, x, Size\_\mathcal{EDB}) \geq minsup$ **then**
      $frequent \leftarrow frequent \cup \{x\}$
    end if
  end for
  return $frequent$
end function
$\mathcal{EIFF} \leftarrow \emptyset$
$size \leftarrow 1$
$candidate \leftarrow candidate\_apriori\_gen(\mathcal{EDB}, size)$
While ($candidate \neq \emptyset$)
$freq \leftarrow Frequent\_itemset\ (candidate, minsup, PT, Size\_\mathcal{EDB})$
$size \leftarrow size + 1$
$\mathcal{EIFF} \leftarrow \mathcal{EIFF} \cup freq$
$candidate \leftarrow candidate\_apriori\_gen(\mathcal{EDB}, size, freq)$
End While

6 Experimentation and results

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

6.1 Evidential database construction

No doubt the evidential database is a real life need where opinions are perfectly modeled via BBAs. Despite their real contribution, evidential databases are really hard to find. In [8], tests were conducted on synthetic database. Even in [15], the constructed BBA includes only one evidential attributes. In [9], the authors worked on a simplified naval anti-surface warfare scenario. In the following, we propose a method that allows to construct an evidential database from a numerical dataset. We based our evidential database construction on the ECM [16] clustering approach. It is an FCM-like algorithm based on the concept of credal partition, extending those of hard, fuzzy, and possibilistic ones. To derive such a structure, we minimized the proposed objective function:

$$J_{ECM}(M,V) \triangleq \sum_{i=1}^{d} \sum_{\{j/A_j \neq \emptyset, A_j \subseteq \Omega\}} c_j^{\alpha} m_{ij}^{\beta} dist_{ij}^2 + \sum_{i=1}^{n} \delta^2 m_{i\emptyset}^{\beta} \tag{15}$$

subject to:

$$\sum_{\{j/A_j \neq \emptyset, A_j \subseteq \Omega\}} m_{ij} + m_{i\emptyset} = 1 \quad \forall i = 1, d \tag{16}$$

where $m_{i\emptyset}$ and m_{ij} denote respectively $m_i(\emptyset)$ and $m_i(A_j)$. M is the credal partition $M = (m_1, \ldots, m_d)$ and V is a cluster centers matrix. c_j^α is a weighting coefficient and $dist_{ij}$ is the Euclidean distance. In our case, the parameters α, β and δ were fixed to 1, 2 and 10.
In order to obtain evidential databases, this approach was applied on several UCI benchmarks [17]. The studied datasets are summarized on Table 3 in terms of number of instances and attributes. For each dataset, the number of focal elements after ECM application was addressed. The number of focal element is related to the objective function J_{ECM} that was minimized.

Table 3. Data set characteristics

Data set	**#Instances**	**#Attributes**	**#Focal elements**
Iris	150	4	32
Vertebral Column	310	6	64
Diabetes	767	9	144
Abalone	4177	9	40

6.2 Comparative results

We compared the precise support measure integrated into the EDMA algorithm to [8, 9] support metric. As it is shown in Table 4, the $EDMA - Pr$ attribute concerns the introduced precise support and $EDMA - Bel$ refers to the definition given in section 3. For our experimentation, we integrated the ramification support (subsection 3.2) into EDMA algorithm. Since EDMA relies on a table that contains all item's metric values, we created the Belief Table (BT). The Belief Table has the same structure as that of Pr Table (c.f., subsection 4.2) and in which we stored all item's belief. The different approaches were tested on the obtained evidential database as illustrated in subsection 6.1.

Table 4. Comparative results in terms of the number of frequent pattern number

Support	Iris		Diabete		Vertebral Column		Abalone	
	EDMA-Pr	EDMA-Bel	EDMA-Pr	EDMA-Bel	EDMA-Pr	EDMA-Bel	EDMA-Pr	EDMA-Bel
0.9	15	15	319	191	63	63	767	511
0.8	23	15	1503	319	95	63	767	511
0.7	47	15	5055	671	415	63	767	511
0.6	95	31	9074	1407	799	95	1919	511

Table 4 illustrates the number of extracted frequent patterns from the evidential databases. As is it demonstrated, the Precise support extracts more frequent pattern than do the based belief method. This result is expected since the precise metric study all subsets of the superset and considers those having

an intersection with the considered itemset. In addition, the number of patterns increases normally as far as the considered *minsup* threshold decreases.

We also conducted performance test on our proposed algorithm which we compared to an exhaustive approach. This approach consists in the Cartesian based algorithm (Cart-Bel in Table 5). The Cartesian algorithm computes all possible BBAs needed for support measure. The complexity of such approach is exponential with respect to the number of focal elements. Indeed, for an evidential database with k attributes each one has n focal elements and d transactions. The arithmetic complexity of a Cartesian product is: $\mathcal{C} = d \times n^k = O(n^k)$.

Table 5. Comparative results in terms of execution time (seconds)

Support	Iris			Diabete			Vertebral Column			Abalone		
	EDMA-Pr	EDMA-Bel	Cart-Bel ≈	EDMA-Pr	EDMA-Bel	Cart-Bel ≈	EDMA-Pr	EDMA-Bel	Cart-Bel ≈	EDMA-Pr	EDMA-Bel	Cart-Bel ≈
0.9	0.13	0.10	96172	4.72	1.65	6.43E+48	0.74	0.24	3.96E+24	79.71	16.35	9.13E+41
0.8	0.13	0.11	96172	175.94	3.00	6.43E+48	1.11	0.24	3.96E+24	75.77	16.18	9.13E+41
0.7	0.35	0.15	96172	21188	12.69	6.43E+48	15.21	0.24	3.96E+24	77.23	16.24	9.13E+41
0.6	1.01	0.25	96172	12.21E+4	100.56	6.43E+48	116.32	0.33	3.96E+24	337.87	16.21	9.13E+41

Table 5 illustrates a comparative performance tests between EDMA and Cartesian based algorithm. The proposed algorithm has drastically improved the results. The extraction performance of the EDMA-Bel is better than those of EDMA-Pr. This observation can be explained by the number of extracted patterns. The more frequent candidates are generated, the more time consumed is.

7 Conclusion

In this paper, we tackled data mining problem in evidential databases. We focused on evidential itemsets' support estimation. We detailed state of art evidential support metric. To drop the original complexity, we proposed a simplification for their methods by reducing the Cartesian product to a simple belief product. We also introduced a new support measure that brings precision by analyzing deeply the BBA's frame of discernment. The proposed precise measure extracts more hidden frequent patterns than the usual method. The precise measure was applied on an Apriori based algorithm and was tested on evidential databases obtained from transformed datasets. As illustrated in the experimentation section, despite the huge item's number that evidential database contains, EDMA generates all frequent itemsets in a reasonable execution time. This problem can be recovered in future works by tackling compact evidential itemset representation. Indeed, estimating the support exactly from a compact set had never been more challenging. In addition, quality test for the generated frequent patterns is a need. In future work, we plan to study the developpement of a new method

to estimate the confidence of evidential associative rules based on our support measure.

References

1. 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.
2. L. A. Zadeh, Fuzzy sets, Information and Control 8 (3) (1965) 338–353.
3. 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.
4. Y. Chen, C. Weng, Mining association rules from imprecise ordinal data, Fuzzy Set Syst 159(4) (2008) 460–474.
5. Y. Lee, T. Hong, W. Lin, Mining fuzzy association rules with multiple minimum supports using maximum constraints, Knowledge-Based Intelligent Information and Engineering Systems 3214 (2004) 1283–1290.
6. A. Dempster, Upper and lower probabilities induced by multivalued mapping, AMS-38, 1967.
7. G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.
8. 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.
9. 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.
10. P. Smets, The Transferable Belief Model and other interpretations of Dempster-Shafer's Model, In Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence, UAI'90, MIT, Cambridge, MA (1990) 375–383.
11. P. Smets, Belief functions, in Non Standard Logics for Automated Reasoning , P. Smets, A. Mamdani, D. Dubois, and H. Prade, Eds. London,U.K: Academic (1988) 253–286.
12. P. Smets, R. Kennes, The Transferable Belief Model, Artificial Intelligence 66 (2) (1994) 191–234.
13. K. K. R. Hewawasam, K. Premaratne, M.-L. Shyu, S. P. Subasingha, Rule mining and classification in the presence of feature level and class label ambiguities, in: SPIE 5803, Intelligent Computing: Theory and Applications III, 98, 2005.
14. 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.
15. M. A. Bach Tobji, B. Ben Yaghlane, K. Mellouli, Frequent itemset mining from databases including one evidential attribute, In Proceedings of second international conference on Scalable Uncertainty Management, Napoli, Italy 5291 (2008) 19–32.
16. M.-H. Masson, T. Denœux, ECM: An evidential version of the fuzzy c-means algorithm, Pattern Recognition 41 (4) (2008) 1384 – 1397.
17. A. Frank, A. Asuncion, UCI machine learning repository (2010), URL: http://archive.ics.uci.edu/ml.