Selecting Relevant Association Rules From Imperfect Data

Cécile L 'Héritier, Sébastien Harispe, Abdelhak Imoussaten, Gilles Dusserre, Benoit Roig

Selecting Relevant Association Rules From Imperfect Data

Cécile L'Héritier[1,2], Sébastien Harispe[1], Abdelhak Imoussaten[1], Gilles Dusserre[1], and Benoît Roig[2]

[1] LGI2P, IMT Mines Ales, Univ Montpellier, Alès, France
{firstname.name}@mines-ales.fr
[2] EA7352 CHROME, Université de Nîmes, France
{firstname.name}@unimes.fr

Abstract. Association Rule Mining (ARM) in the context of imperfect data (*e.g.* imprecise data) has received little attention so far despite the prevalence of such data in a wide range of real-world applications. In this work, we present an ARM approach that can be used to handle imprecise data and derive imprecise rules. Based on evidence theory and Multiple Criteria Decision Analysis, the proposed approach relies on a selection procedure for identifying the most relevant rules while considering information characterizing their interestingness. The several measures of interestingness defined for comparing the rules as well as the selection procedure are presented. We also show how *a priori* knowledge about attribute values defined into domain taxonomies can be used to (i) ease the mining process, and to (ii) help identifying relevant rules for a domain of interest. Our approach is illustrated using a concrete simplified case study related to humanitarian projects analysis.

Keywords: Association rules · Imperfect data · Evidence theory · Multiple Criteria Decision Analysis (MCDA).

1 Introduction

Association rule mining (ARM) is a well-known data mining technique designed to extract interesting patterns in databases. It has been introduced in the context of market basket analysis [1], and has received a lot of attention since then [15]. An association rule is usually formally defined as an implication between an *antecedent* and a *consequent*, being conjunctions of attributes in a database, e.g. "People who have age-group between 20 and 30 and a monthly income greater than $2k are likely to buy product X". Such rules are interesting for extracting simple intelligible knowledge from a database; they can also further be used in several applications, e.g. recommendation, customer or patient analysis. A large literature is dedicated to the study of ARM, and numerous algorithms have been defined for efficiently extracting rules handling a large range of data types, e.g., nominal, ordinal, quantitative, sequential [15]. Nevertheless, only a few contributions of the literature study the case of ARM with imperfect data, e.g. [13, 24], even if such data is central in numerous real-world applications.

In order to extend the body of work related to ARM with imperfect data, and to answer some of the limitations of existing contributions, this paper presents a novel ARM approach that can be used to handle imprecise data and derive imprecise rules. In this study, to simplify, the proposed approach focuses on a specific case where the *antecedent* and the *consequent* are composed of predefined disjoint sets of attributes forming a partition of the whole set of attributes. This particular case is relevant, for example in classification tasks in which the label value to predict can be defined as consequent of the rules of interest. To sum up, our goal is threefold: (i) to enrich the expressivity of existing proposed frameworks, (ii) to complement them with a richer procedure for selecting relevant rules, and (iii) to present simple way to incorporate domain knowledge to ease the mining process, and to help identifying relevant rules for a domain of interest. Based on the evidence theory framework and Multiple Criteria Decision Analysis, a selection procedure for identifying the most relevant rules while considering information characterizing their interestingness is proposed. The several measures of interestingness defined for comparing the rules, as well as the selection procedure, are presented. We also show how *a priori* knowledge in the form of taxonomies about consequent and antecedent (i.e. attribute values) can be used to focus on rules of interest for a domain. We also present an illustration using a simplified case study related to humanitarian projects analysis.

The paper is structured as follows: Section 2 formally introduces traditional ARM, the theoretical notions on which our approach is based, and formally defines the problem we are considering. It also introduces related work focusing on rule selection and ARM with imperfect data. The proposed approach is detailed in Section 3, and Section 4 presents the illustration. Finally, perspectives and concluding remarks are provided in Section 5.

2 Theoretical background and related work

This section briefly presents some of the theoretical notions required to introduce our work. We next provide the problem statement of ARM with imperfect data, and our positioning w.r.t. existing contributions.

2.1 Theoretical background

Association Rule Mining (ARM): In classical ARM [1], a database $\mathcal{D} = \{d_1, \ldots, d_m\}$ to be mined consists of m observations of a set of n attributes. The set of attribute indices is denoted by $N = \{1, \ldots, n\}$. Each attribute i takes its values in a discrete -boolean, nominal or numerical- finite scale denoted Θ_i. An association rule r denoted $r : X \rightarrow Y$ links an antecedent X with a consequent Y where $X \in \prod_{i \in I} \Theta_i$, $I \subset N$ and $Y \in \prod_{j \in J} \Theta_j$, $J \subseteq N \setminus I$.

The main challenge in ARM is to extract *interesting* rules from a large search space, *e.g.*, n and m are large. In this context, defining the *interestingness* of a rule is central.

Interestingness of rules. Numerous works have studied notions related to the *interestingness* of a rule, [16, 22, 23]. No formal and widely accepted definition arose from those works, and discussing the numerous existing formulations is out of the scope of this paper. However, interestingness is generally regarded as a general concept covering several features of interest for a rule, e.g. *reliability* (how reliable is the rule?) and *conciseness* (is the rule complex?, i.e. based on numerous attribute-value pairs). Other aspects of a rule are also considered, e.g. *peculiarity*, *surprisingness*, or *actionability*, to name a few - the reader can refer to [12] for details. The literature also distinguishes objective and subjective measures, the latter being defined based on domain-dependent considerations. The two main (objective) measures used in the literature are *Support* and *Confidence* [2]. The *support* of a rule $r : X \rightarrow Y$ denoted $supp(X \rightarrow Y)$ is traditionnally defined as the proportion of the realization of X and Y in $\mathcal{D}$, and the *confidence* denoted $conf(X \rightarrow Y)$ is defined as the proportion of the realization of Y when X is observed in $\mathcal{D}$. Given support and confidence thresholds, ARM usually aims at identifying rules exceeding those thresholds [2]. In classical ARM, support and confidence are quantified using probability theory framework. When ARM involves imperfect data, this quantification requires reformulating the problem in a theoretical framework suited for handling data imperfection. In this work, we focus on contributions based on evidence theory.

Evidence theory has been introduced to represent imprecision and uncertainty [21]. We briefly introduce its main concepts. Let Θ be a finite set of elements being the most precise available information, referred to as the *frame of discernment*. A *mass function* $m : 2^\Theta \rightarrow [0, 1]$ is a set function such that $\sum_{A \subseteq \Theta} m(A) = 1$. The quantity $m(A)$, $A \subseteq \Theta$ is interpreted as the portion of belief that is exactly committed to A and to nothing smaller. The subsets of Θ having a strictly positive mass are called *focal elements*, their set is denoted $\mathcal{F}$. The total belief committed to any $A \subseteq \Theta$ is measured by the *belief function*: $Bel : 2^\Theta \rightarrow [0, 1]$ with $Bel(A) = \sum_{B \subseteq \Theta, B \subseteq A} m(B)$. In evidence theory, $Bel(\overline{A})$, where $\overline{A}$ denotes the complement of A in Θ, is characterized through the notion of *plausibility*: $Pl : 2^\Theta \rightarrow [0, 1]$, with $Pl(A) = 1 - Bel(\overline{A}) = \sum_{B \subseteq \Theta, B \cap A \neq \emptyset} m(B)$.

In order to provide a complete generalization of the probability framework, conditioning has also been defined in evidence theory. Several expressions have been proposed, none of them leading to a full consensus [7, 10]. In this paper, we will adopt the definition corresponding to the conditioning process stated by Fagin et al. [10], a natural extension of the Bayesian conditioning. We do not consider the definition proposed in Dempster [7] based on Dempster-Shafer combination rule, where a new information is interpreted as a modification of the initial belief function and used in a revision process [9]. Thus, for $A, B \subseteq \Theta$, such that $Bel(A) > 0$, we will further consider:

$$Bel(B|A) = \frac{Bel(A \cap B)}{Bel(A \cap B) + Pl(A \cap \overline{B})}, \qquad Pl(B|A) = \frac{Pl(A \cap B)}{Pl(A \cap B) + Bel(A \cap \overline{B})}$$

2.2 Problem statement and related work

Problem statement. In classical ARM, where only precise information is considered, *e.g.*, the value of attribute i is $X_i \in \Theta_i$, $i \in N$. In this paper, we consider observations as "the value of attribute i is in $A_i \subseteq \Theta_i$". The case $A_i \subset \Theta_i$ with $|A_i| > 1$ corresponds to imprecision, while $A_i = \Theta_i$ is considered when information is missing, i.e. it corresponds to the ignorance about the value of attribute i. In this setting, a rule r is defined as:

$$r : A \rightarrow B \text{ where } A = \prod_{i \in I} A_i, A_i \subseteq \Theta_i \text{ and } B = \prod_{j \in J} B_j, B_j \subseteq \Theta_j$$
$$\text{for all } I \subset N \text{ and } J \subseteq N \setminus I$$

As mentioned previously, in this paper we consider the case where antecedent A concerns only a subset $I_1 \subset N$ of attributes and consequent B concerns a subset $I_2 \subset N$ where I_1 and I_2 form partition of N, and $I_1 \neq \emptyset$. Thus:

$$r : A \rightarrow B \quad \text{where} \quad A = \prod_{i \in I_1} A_i, A_i \subseteq \Theta_i \quad \text{and} \quad B = \prod_{j \in I_2} B_j, B_j \subseteq \Theta_j \tag{1}$$

We denote by $\mathcal{R}$ the set of rules defined by Formula (1). The problem addressed here is to reduce $\mathcal{R}$ by selecting only the relevant rules.

Related work and positioning. As stated in the introduction, our goal is threefold: (i) to enrich the expressivity of existing proposed frameworks dedicated to ARM with imperfect data, (ii) to complement them with a richer procedure for selecting relevant rules (rule pruning), and (iii) to present a simple way to incorporate domain knowledge to ease the mining process, and to help identifying relevant rules for a domain of interest.[3]

Rule pruning. Most of the approaches use thresholds to select rules - only using support and confidence most often allows drastically reducing the number of rules in traditional ARM [1]. A post-mining step is generally performed to rank the remaining rules according to one specific interestingness measure -the measure used is generally selected according to the application domain and context-specific measure properties [23, 27]. Nevertheless, processing this way does not enable selecting rules when conflicting interestingness measures are used, e.g. maximizing both support and specificity of rules. This is the purpose of MCDA methods. Some works propose to take advantage of MCDA methods [3–6, 17] in the context of ARM. Those works can be divided into two categories: 1) those incorporating the end-user's preferences using Analytic Hierarchy Process (AHP) and Electre II [6], or using Electre tri [3]; and 2) those that do not incorporate such information and use Data Envelopment Analysis (DEA) [5, 26], or Choquet

[3] Note that the simplification of the mining process here refers to a reduction of complexity in terms of the number of rules analysed, i.e. search space size. Algorithmic contributions and therefore complexity analyses regarding efficient implementations of the proposed approach are left for future work.

integral [17]. Our approach is hybrid and falls within the two categories. First, selection is made based only on database information as in Bouker et al. [4]. Second, if the set of selected rules is large, a trade-off based on end-user's preferences is used within an appropriate MCDA method. As our aim is to select a subset of interesting rules, Electre I [18] seems to be the most appropriate.

ARM and imperfect data. Several frameworks have been studied to deal with imperfect data in ARM. The assumptions entailed in the approaches based on probabilistic models do not preserve imprecision and might lead to unreliable inferences [13]. Uncertainty theories have also been investigated for imperfect data in ARM using fuzzy logic [14], or using possibility theory [8]. In the case of missing and incomplete data, evidential theory seems the appropriate setting to handle ARM problem [13, 19, 24, 25]. Our approach is adopting this setting. In addition to studying a richer modelling that enables incorporating more information, we propose to combine it with a selection process taking advantage of an MCDA method, namely Electre I, to assess rules interestingness considering different viewpoints. Although some works previously mentioned tackle rule selection using MCDA, and few approaches have been addressing ARM problem using evidence theory, none of them is addressing both issues simultaneously.

We also present how to benefit from *a priori* knowledge about attribute values -organised into taxonomies- for improving the rule selection process, and reducing the increase of complexity induced by the proposed extension of modellings used so far in existing ARM approaches suited for imperfect data.

3 Proposed approach

This section presents our ARM approach for imperfect data. We first introduce how rule interestingness is evaluated by presenting the selected measures and their formalization in the evidence theory framework. Then, the main steps of the proposed approach for selecting rules based on these measures are detailed.

3.1 Assessing rule interestingness from imprecise data

In this study, we focus on important objective measures of interestingness - subjective ones, involving further interactions with final user, are most often considered context-dependent and will not be considered in this paper. We propose to evaluate rules according to (i) their support, (ii) their confidence, as well as (iii) indirect evaluations used to criticize their potential relevance. In addition, since in our context rules are imprecise, and since very imprecise rules are most often considered useless, the (iv) degree of imprecision embedded in the mined rules is also evaluated. These four notions of interest considered in the study are defined below. For convenience, we consider that we are computing measures to evaluate a rule $r : A \rightarrow B$ where $A = \prod_{i \in I_1} A_i, A_i \subseteq \Theta_i$ and $B = \prod_{j \in I_2} B_j, B_j \subseteq \Theta_j$ with $I_1 \cup I_2 = N$. In our context, since we consider $n = |N|$ attributes, the

set functions mass m, belief Bel and plausibility Pl are defined on subsets of $\Theta = \prod_{i \in N} \Theta_i$.

Support. A rule is said to be supported if observations of its realization are frequent [2]. In our context, the support of a rule relates to the masses of evidence associated to observations supporting the rule, either explicitly or implicitly. The belief function is thus used to express support:

$$supp(r : A \rightarrow B) = Bel(A \times B) \tag{2}$$

Note that the belief function is monotone, then, the rules composed of the most imprecise attribute values will necessarily be the most supported.

Confidence. A rule is said to be reliable if the relationship described by the rule is verified in a sufficiently great number of applicable cases [12]. The *Confidence* measure is traditionally evaluated as a conditional probability [1]. Its natural counterpart in evidence theory is given by the conditional belief, leading to the following expression:

$$conf(r : A \rightarrow B) = Bel(B \mid A) = \frac{Bel(A \times B)}{Bel(A \times B) + Pl(A \times \overline{B})} \tag{3}$$

The elements defining the consequent are conditioned to the elements composing the antecedent. Note that the belief and conditional belief functions have also been adopted to express support and confidence for ARM with imprecise data [13, 24]. In those cases the modelling and domain definition were different, i.e. restricted to the cartesian products of the power-sets of attribute domains.

Indirect measures of potential relevance. These measures will be introduced through an illustration. Consider humanitarian projects described by two attributes: the *transport means* with $\Theta_1 = \{truck, motorbike, helicopter\}$, and the final *coverage reached* in the project (proportion of beneficiaries), with $\Theta_2 = \{low, moderate, high\}$. To criticize the relevance of a rule $r : A \rightarrow B$, e.g. $r : \{truck\} \rightarrow \{high\}$, we propose to evaluate the following relations:

- $A \rightarrow \overline{B}$. In the example, if the rule $\{truck\} \rightarrow \{\overline{high}\}$ holds, it means that most often using *trucks* also leads to a *coverage* that is *not high*. Hence we consider that validating $A \rightarrow \overline{B}$ conveys a contradictory information w.r.t. to the rule $A \rightarrow B$ and tends to invalidate it.
- $\overline{A} \rightarrow B$. If the rule $\{\overline{truck}\} \rightarrow \{high\}$ holds, it means that in some cases, some of the *other means of transport* also allow to reach a *high coverage*. Such an information tends to decrease the interest of the rule $r : A \rightarrow B$ if we assume that B is not explained by multiple causes.
- $\overline{A} \rightarrow \overline{B}$. The rule $\{\overline{truck}\} \rightarrow \{\overline{high}\}$ means that when *trucks* are not used, a *low or moderate coverage* (not high) is obtained. We assume that most commonly, if $\{truck\} \rightarrow \{high\}$ is somehow assumed to be considered as valid, supporting $\{\overline{truck}\} \rightarrow \{\overline{high}\}$ will reinforce our interest over $\{truck\} \rightarrow \{high\}$.

In a probabilistic framework, only the relationship $\overline{A} \rightarrow \overline{B}$ would have to be studied, since the other ones do not provide additional information, i.e. $P(\overline{B}|A) = 1 - P(B|A)$, $P(B|\overline{A}) = 1 - P(\overline{B}|\overline{A})$, $P(A \times \overline{B}) = P(A)P(\overline{B}|A)$ and $P(\overline{A} \times B) = (1 - P(A))P(B|\overline{A})$. Thus, the potential relevance of a rule takes into consideration the confidence of the rule composed of the complements of the antecedent and the consequent, given by: $P(\overline{B}|\overline{A})$. Note that, in the literature, this measure is also referred to as *specificity*. When considering evidence theory, the information about the complement is provided by the plausibility function, such as $Bel(A) = 1 - Pl(\overline{A})$ and then $Bel(B|A) = 1 - Pl(\overline{B}|A)$. In this context, Table 1 introduces the relationships between the confidence of a rule (conditional belief) and the ones involving the complement of its antecedent and/or consequent.

Note that to criticize the relevance of a rule using the three rules involving its complements, we propose to consider their respective *support* and *confidence*: criticizing a rule on the basis of weakly supported rules would not be appropriate.

Table 1. Relationships between support and confidence of a rule $r : A \rightarrow B$ and rules involving its complements.

Rule	Support	Confidence	depends on quantities:
$A \rightarrow B$	$Bel(A \times B)$	$Bel(B \mid A)$	$Bel(A \times B)$ and $Pl(A \times \overline{B})$
$A \rightarrow \overline{B}$	$Bel(A \times \overline{B})$	$Bel(\overline{B} \mid A) = 1 - Pl(B \mid A)$	$Bel(A \times \overline{B})$ and $Pl(A \times B)$
$\overline{A} \rightarrow B$	$Bel(\overline{A} \times B)$	$Bel(B \mid \overline{A}) = 1 - Pl(\overline{B} \mid \overline{A})$	$Bel(\overline{A} \times B)$ and $Pl(\overline{A} \times \overline{B})$
$\overline{A} \rightarrow \overline{B}$	$Bel(\overline{A} \times \overline{B})$	$Bel(\overline{B} \mid \overline{A})$	$Bel(\overline{A} \times \overline{B})$ and $Pl(\overline{A} \times B)$

Specificity using Information Content. Finally, we propose to incorporate the specificity of a rule. Let's consider the information "the value of attribute i is in the subset A_i". This information is more specific than the information "the value of attribute i is in the subset $A_i^{'}$" where $A_i \subset A_i^{'}$. Based on the notion of Information Content (IC) defined for comparing concept specificities in ontologies [20], we propose to quantify the specificity of a rule r by:

$$IC(r : A \rightarrow B) = 1 - \frac{\log |\{X : X \subseteq A \times B\}|}{|\Theta|} \tag{4}$$

$|X|$ denotes the number of elements in the set X and $\Theta = \prod_{i \in N} \Theta_i$.

3.2 Search space reduction

Let us remind the starting set $\mathcal{R}$ -see Formula (1)- of rules from which a small subset $\mathcal{R}^*$ of interesting rules should be selected:

$$\mathcal{R} = \{r : \ A \rightarrow B \mid A = \prod_{i \in I_1} A_i, \ A_i \subseteq \Theta_i, \ B = \prod_{j \in I_2} B_j, \ B_j \subseteq \Theta_j\}$$

We assume that I_1 and I_2 are fixed before starting the ARM process.

To simplify notations in the rest of the paper, we will denote by $r_{A,B}$ the rule $r : \ A \rightarrow B$ where A and B are as in the Formula (1). Two restrictions are proposed below:

1. All rules being supported are generalizations (supersets) of focal elements $\mathcal{F}$, i.e. $\mathcal{F} = \{X : X \subseteq \Theta, m(X) > 0\}$. Since support is a prerequisite for assessing rule validity, we further consider that the evaluation will be restricted to the set:

$$\mathcal{R}_r = \{r_{A,B} \in \mathcal{R} \mid \exists X \in \mathcal{F} \text{ st. } X \subseteq A \times B\}$$

2. The search space can also be reduced using prior knowledge defined into ontologies expressing taxonomies of attribute values. Since the ontology defines the concepts of interest for a domain, a restriction can be performed only considering the attribute values defined into taxonomies. Thus, for each $i \in N$, only a subset $\mathcal{O}_i$ of 2^{Θ_i} of the information of interest for a domain is considered. We can then define the following restriction:

$$\mathcal{R}_{r,t} = \{r_{A,B} \in \mathcal{R}_r \mid A = \prod_{i \in I_1} A_i,\ A_i \in \mathcal{O}_i,\ B = \prod_{j \in I_2} B_j,\ B_j \in \mathcal{O}_j\}$$

3.3 Rules selection process

The proposed approach aims at selecting the most relevant rules $\mathcal{R}^*$ according to their evaluations on a set of interestingness measures listed in Table 2. We here consider that the evaluated rules are members of the restriction $\mathcal{R}_{r,t} \subseteq \mathcal{R}$, even if that condition could further be relaxed. We denote the set of interestingness measures by K ($|K| = 9$), and $g_k(r)$ the score of rule r for the measure $k \in K$. To simplify notations, we consider that $g_k(r)$ is to maximize[4] for all $k \in K$. A two-step pruning strategy is proposed.

Table 2. Summary of interestingness measures considered in the selection process

$k \in K$	Measures	Formulae $\forall r \in \mathcal{R}_{r,t}\ r : A \rightarrow B$	variation	weight
1	Rule Support	$supp(r) = Bel(A \times B)$	maximize	w_1
2	Rule Confidence	$conf(r) = Bel(B\|A)$	maximize	w_2
3	Rule Specificity	$IC(r)$	maximize	w_3
4	$A \rightarrow \overline{B}$	$Bel(A \times \overline{B})$	minimize	w_4
5		$Bel(\overline{B}\|A)$	minimize	w_5
6	$\overline{A} \rightarrow B$	$Bel(\overline{A} \times B)$	minimize	w_6
7		$Bel(B\|\overline{A})$	minimize	w_7
8	$\overline{A} \rightarrow \overline{B}$	$Bel(\overline{A} \times \overline{B})$	maximize	w_8
9		$Bel(\overline{B}\|\overline{A})$	maximize	w_9

[4] Indeed all the measures used in our approach take values in the interval $[0, 1]$, then a measure k to minimize can be changed to a measure to maximize by considering $1 - g_k(r)$ instead of $g_k(r)$.

Step 1: Dominance-based pruning. A reduction of the concurrent rules in $\mathcal{R}_{r,t}$ is carried out by focusing on non-dominated rules on the basis of the considered measures. A rule r_1 dominates a rule r_2, we write $r_2 \prec r_1$, iff r_1 is at least equal to r_2 on all measures and it exists a measure where r_1 is strictly superior to r_2. More formally,

$$r_2 \prec r_1 \text{ iff } g_k(r_2) \leq g_k(r_1), \forall k \in K \text{ and } \exists j \in K \text{ such that } g_j(r_2) < g_j(r_1).$$

The reduced set of rules can be stated as:

$$\mathcal{R}_{r,t,d} = \{r \in \mathcal{R}_{r,t} \mid \nexists r' \in \mathcal{R}_{r,t} : r \prec r'\}$$

Step 2: Pruning using Electre I. When $\mathcal{R}_{r,t,d}$ remains too large to be manually analyzed, a subjective pruning procedure based on the selection procedure Electre I is applied. This MCDA method enables expressing subjectivity through parameters that can be given by decision makers [18]. We use it for finding the final set of rules $\mathcal{R}^* \subseteq R_{r,t,d}$. Electre I builds an outranking relation between pairs of rules allowing to select a subset of the best rules: $\mathcal{R}^*$. This subset is such that (i) any rules excluded from $\mathcal{R}_{r,t,d}$ is outranked by at least one rule from $\mathcal{R}^*$, (ii) rules from $\mathcal{R}^*$ do not outrank each other. To do so, Electre I procedure (a) constructs outranking relationships through pairwise comparisons of rules, to further (b) exploit those relationships to build $\mathcal{R}^*$.

a) Outranking relations: the relationship "r *outranks* r'" (rSr') means that r is at least as good as r' on the set of measures K. The outranking assertion rSr' holds if: (i) a sufficient coalition of measures supports it, and (ii) none of the measures is too strongly opposed to it. These conditions are respectively referred to as concordance $c(rSr')$ and discordance indices $d(rSr')$, such that:

$$c(rSr') = \sum_{\{k\,:\,g_k(r) \geq g_k(r')\}} w_k \text{ and } d(rSr') = \max_{\{k\,:\,g_k(r) < g_k(r')\}} [g_k(r') - g_k(r)],$$

with w_k the relative importance of measure k.
From these notations, we consider rSr' if $c(rSr') \geq \widehat{c}$ and $d(rSr') \leq \widehat{d}$; with $\widehat{c}$ and $\widehat{d}$, two thresholds defining when the outranking should be considered or not.

b) Relations exploitation: a graph of outranking relationships is obtained from these pairwise comparisons. The kernel of this graph is our final reduced set of rules $\mathcal{R}^*$ to be considered, such that:

$$\begin{aligned} &- \forall r' \in \mathcal{R}_{r,t,d} \setminus \mathcal{R}^*, \exists\, r \in \mathcal{R}^* : rSr', \textit{and} \\ &- \forall (r, r') \in \mathcal{R}^* \times \mathcal{R}^*, \neg(rSr'). \end{aligned} \tag{5}$$

The set of model parameters that have to be defined for applying the subjective reduction based on Electre I are: weights $w_k, \forall k \in K$, and the concordance and discordance thresholds, $\widehat{c}$, $\widehat{d}$.[5] The choice of parameter values will be further discussed in the illustration Section 4.

[5] Evaluating support and confidence of $\overline{A} \rightarrow B$ and $\overline{A} \rightarrow \overline{B}$ can lead to undefined values, e.g. evaluating $\overline{A} \rightarrow B$, we have $Bel(\overline{A} \times B) = 0$ when $\overline{A}$ has never been

4 Illustration

As an illustration, we consider the context of humanitarian projects carried out for answering to emergency situations. A dataset of observations describes these emergency situations according to four attributes: 1) the *type of disaster* faced, 2) the *season*, 3) the *environment* in which it occurred, and 4) an evaluation of the situation w.r.t. the *human cost*. We further refer to these attributes using their number, considering that they respectively take discrete values in: $\Theta_1 = \{$ *tsunami, earthquake, epidemic, conflict, pop.displacement*$\}$, $\Theta_2 = \{$*spring, summer, autumn, winter*$\}$, $\Theta_3 = \{$*urban, rural*$\}$, $\Theta_4 = \{$*low, medium, high, veryHigh*$\}$. Besides, for each attribute, prior knowledge is defined into ontologies determining the values of interest. In this specific case study, the purpose of association rules is to highlight the influence of a situation contextual features on its evaluation according to the *Human Cost*, a useful information for project planning. Thus the searched rules $r : A \rightarrow B$ will imply the attributes in the following set $I_1 = \{1, 2, 3\}$ in the *antecedent* and in $I_2 = \{4\}$ for the *consequent*.

Table 3. Database of observations expressed using precise, imprecise or missing values.

	Disaster Type	Season	Environment	Human Cost
d_1	$\{earthquake\}$	$\{autumn\}$	$\{rural\}$	$\{medium\}$
d_2	$\{tsunami\}$	$\{autumn\}$	$\{urban\}$	$\{medium\}$
d_3	$\{epidemic\}$	-	$\{urban\}$	$\{veryHigh\}$
d_4	$\{earthquake, epidemic, tsunami\}$	$\{spring\}$	-	$\{high, veryHigh\}$
d_5	$\{epidemic\}$	$\{spring\}$	$\{urban\}$	$\{high\}$
d_6	$\{epidemic\}$	$\{spring, summer\}$	-	$\{high, veryHigh\}$
d_7	$\{epidemic\}$	$\{spring, summer\}$	$\{urban\}$	$\{high, veryHigh\}$
d_8	$\{epidemic\}$	$\{spring, summer\}$	$\{urban\}$	$\{veryHigh\}$
d_9	$\{earthquake, epidemic, tsunami\}$	$\{summer\}$	$\{rural\}$	$\{high\}$
d_{10}	$\{epidemic\}$	$\{summer\}$	$\{urban\}$	$\{high\}$
d_{11}	$\{epidemic\}$	$\{summer\}$	$\{urban\}$	$\{veryHigh\}$
d_{12}	$\{earthquake\}$	$\{winter\}$	$\{rural\}$	$\{high, medium, veryHigh\}$
d_{13}	$\{earthquake\}$	$\{winter\}$	$\{rural\}$	$\{low\}$
d_{14}	$\{earthquake, epidemic, tsunami\}$	$\{winter\}$	$\{rural\}$	$\{high\}$

Among the observations of 14 projects given in Table 3, some attribute values are expressed with imprecision, e.g. *Human cost* values may be unclear such that *"human Cost is High or VeryHigh"*. When values are missing the total ignorance is considered. In this setting, the size of the initial studied space $\mathcal{R}$ is $\prod_{i=1}^{4} 2^{|\Theta_i \setminus \emptyset|} = 20925$. Using the restrictions focusing on rules with non-null support, and involving attribute values of interest defined into ontologies (cf. Section 3), we obtain a reduced search space $\mathcal{R}_{r,t}$ composed of 484 rules.

The rule evaluation and selection process is further applied to $\mathcal{R}_{r,t}$ using the 9 interestingness measures proposed in Table 2. Using dominance-based pruning

observed, leading to $Bel(B|\overline{A})$ being undefined. However, pruning using dominance and Electre I requires the same measures to be defined. Undefined values are thus substituted by an arbitrary value that neither favor nor penalize the evaluation of the rule $A \rightarrow B$. The median of $Bel(\overline{A} \times B)$ (resp. $Bel(\overline{A} \times \overline{B})$) has been chosen. Note that $A \rightarrow \overline{B}$ is not concerned since evaluating $A \rightarrow B$ implies evidence on A.

(Step 1/2), a set of 18 non-dominated rules $\mathcal{R}_{r,t,d}$ is identified among the 484 rules initially considered. These rules are listed in Table 4, and indexed from r_0 to r_{17}. Pruning using Electre I is then applied over the set of non-dominated rules $\mathcal{R}_{r,t,d}$ (Step 2/2). Different sets of selected rules -i.e. $\mathcal{R}^*$- are given in Table 5 for different sets of model parameters. The results being sensitive to parameter values, we propose to discuss different parameter settings. We remind that these parameters are: $\forall k \in K$, w_k the weights of interestingness measures, and $\widehat{c}$ and $\widehat{d}$ the concordance and discordance thresholds. They represent end-user's preferences. They can be given directly; the weights w_k can also be elicited using Simos, a well-known weighting procedure [11].

Table 4. Set of non-dominated rules, $\mathcal{R}_{r,t,d}$.

	Disaster Type		Season		Environment		Human Cost
r_0 :	$\{earthquake\}$	$\wedge$	$\{autumn\}$	$\wedge$	$\{rural\}$	$\rightarrow$	$\{medium\}$
r_1 :	$\{earthquake, tsunami\}$	$\wedge$	$\{autumn\}$	$\wedge$	Θ_3	$\rightarrow$	$\{medium\}$
r_2 :	$\{tsunami\}$	$\wedge$	$\{autumn\}$	$\wedge$	$\{urban\}$	$\rightarrow$	$\{medium\}$
r_3 :	$\{earthquake, epidemic, tsunami\}$	$\wedge$	Θ_2	$\wedge$	Θ_3	$\rightarrow$	Θ_4
r_4 :	$\{earthquake, epidemic, tsunami\}$	$\wedge$	Θ_2	$\wedge$	Θ_3	$\rightarrow$	$\{high, medium, veryHigh\}$
r_5 :	$\{earthquake, epidemic, tsunami\}$	$\wedge$	Θ_2	$\wedge$	Θ_3	$\rightarrow$	$\{high, veryHigh\}$
r_6 :	$\{epidemic\}$	$\wedge$	Θ_2	$\wedge$	Θ_3	$\rightarrow$	$\{high, veryHigh\}$
r_7 :	$\{epidemic\}$	$\wedge$	Θ_2	$\wedge$	$\{urban\}$	$\rightarrow$	$\{veryHigh\}$
r_8 :	$\{earthquake\}$	$\wedge$	$\{autumn, winter\}$	$\wedge$	$\{rural\}$	$\rightarrow$	$\{medium\}$
r_9 :	$\{earthquake, tsunami\}$	$\wedge$	$\{autumn, winter\}$	$\wedge$	Θ_3	$\rightarrow$	$\{low, medium\}$
r_{10} :	$\{earthquake, tsunami\}$	$\wedge$	$\{autumn, winter\}$	$\wedge$	Θ_3	$\rightarrow$	$\{medium\}$
r_{11} :	$\{earthquake, epidemic, tsunami\}$	$\wedge$	$\{spring, summer\}$	$\wedge$	Θ_3	$\rightarrow$	$\{high, veryHigh\}$
r_{12} :	$\{epidemic\}$	$\wedge$	$\{spring, summer\}$	$\wedge$	Θ_3	$\rightarrow$	$\{high, veryHigh\}$
r_{13} :	$\{epidemic\}$	$\wedge$	$\{spring, summer\}$	$\wedge$	$\{urban\}$	$\rightarrow$	$\{high, veryHigh\}$
r_{14} :	$\{epidemic\}$	$\wedge$	$\{spring, summer\}$	$\wedge$	$\{urban\}$	$\rightarrow$	$\{veryHigh\}$
r_{15} :	$\{epidemic\}$	$\wedge$	$\{summer\}$	$\wedge$	$\{urban\}$	$\rightarrow$	$\{high, veryHigh\}$
r_{16} :	$\{epidemic\}$	$\wedge$	$\{summer\}$	$\wedge$	$\{urban\}$	$\rightarrow$	$\{veryHigh\}$
r_{17} :	$\{earthquake\}$	$\wedge$	$\{winter\}$	$\wedge$	$\{rural\}$	$\rightarrow$	$\{low\}$

Table 5. Final sets of rules ($\mathcal{R}^*$) obtained with Electre I pruning using four parameter settings (a to e).

Different sets of parameters, with $\widehat{c} = 0.7$											
	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	$\widehat{d}$	$\mathcal{R}^*$
a	0.27	0.15	0.1	0.08	0.08	0.08	0.08	0.08	0.08	0.3	$\{r_1, r_3, r_6, r_9, r_{11}\}$
b	0.18	0.18	0.18	0.1	0.1	0.1	0.1	0.03	0.03	0.3 0.2	$\{r_1, r_3, r_6\}$ $\{r_0, r_1, r_2, r_3, r_6, r_{13}, r_{16}, r_{17}\}$
c	0.12	0.2	0.2	0.08	0.08	0.08	0.08	0.08	0.08	0.3 0.2	$\{r_1, r_3, r_6\}$ $\{r_0, r_1, r_2, r_3, r_6, r_{13}, r_{16}, r_{17}\}$
d	0.15	0.25	0.25	0.05	0.05	0.05	0.05	0.075	0.075	0.3 0.2	$\{r_1, r_3, r_6, r_{17}\}$ $\{r_0, r_1, r_2, r_3, r_6, r_{13}, r_{16}, r_{17}\}$
e	0.33	0.33	0.34	0	0	0	0	0	0	0.3	$\mathcal{R}_{r,t,d} \setminus \{r_8, r_{10}, r_{16}, r_{17}\}$

Among the considered interestingness measures, according to the literature, we assume that *support*, *confidence* and *IC* are the most significant ones w.r.t. rule interest. They have to be associated to the most important weights. Conversely, we assume that the other measures -about rule complements- are secondary and will provide additional information for comparing and criticizing the

relevance of rules. In the first set of parameters *(a)* (cf. Table 5), the weight given to *support* and *confidence* is maximized to represent 60% of the votes required for the outranking (to exceed $\widehat{c} = 0.7$). This setting will tend to favor the rules having a high degree of imprecision, being well supported and then reliable, since $Bel(B|A) \geq Bel(A \times B)$. For example, in this setting the rules r_3, r_6, r_{11}, see Tables 5 and 4, are among the selected rules in $\mathcal{R}^*$; e.g. with r_3 involving the total imprecision on three attributes.

When restricting $\widehat{d}$ to 0.2 with the parameter settings *(b), (c), (d)*, it increases the size of the kernel, while still discarding more than half of the rules among the set of non-dominated ones. With parameters *(d)* and $\widehat{d} = 0.3$, highest importance is given to *confidence* and *IC*, providing these 2 measures with 71% of the voting power to reach the outranking condition $\widehat{c} = 0.7$. Thus, a rule with a better score on *confidence*, *IC* and on some of the other measures -except *support*- can be selected while having a low support. This is illustrated with the selection of r_{17} for example. Lastly, the parameter setting (e) is equivalent to considering only the three main measures with equal importance. Here, it enables to discard only 4 extra rules in comparison to dominance relationships. This is explained by the fact that the absence of dominance between rules is more frequent.

Finally, the parameter settings *(b), (c)* or *(d)* with $\widehat{d} = 0.2$, favoring the *support*, *confidence* and *IC* over the other measures tend to provide interesting results. This setting enables the selection of both precise and imprecise rules of interest w.r.t. the initial set of observations, such as r_{16} and r_{13}. In the initial dataset -see Table 3- the imprecise information $\{spring, summer\}$ for the *season* or $\{high, veryHigh\}$ for the *Human Cost* are frequently observed. Indeed, selecting the imprecise rule $r_{13} : \{epidemic\} \wedge \{spring, summer\} \wedge \{urban\} \rightarrow \{high, veryHigh\}$ in $\mathcal{R}^*$ is not surprising. As an interpretation of this rule, we say that the analysis of the database tends to relate the occurrence of epidemics in urban areas to a specific season, spring or summmer, and human cost. In particular, the rule seems valid at least for one the conjunction "summer and high human cost", "summer and a very High human cost", "spring and high" or "spring and veryHigh". In this illustration, different sets of parameters and their results on rule selection have been presented. However, these parameters have to be set by the end-user.

To further discuss these results, it is interesting to note that all the selected measures for rules comparison, except the *IC*, are based on observations frequency. In order to counterbalance the preponderance of this factor, it might be relevant to add subjective measures and not only data-driven ones. Subjective interestingness measures have been studied in the literature. Relying on these works, we could include here measures based for example on user expected rules or expected conjunction of attribute values. Furthermore, investigating the dependencies among frequency based measures, and considering them in the selection process will be valuable. Nevertheless, considering additional measures (especially data-driven), as the ones proposed for classical ARM, is not necessarily straightforward within the evidence theory framework. It indeed implies to define their right expression and meaning in this framework.

5 Conclusion and perspectives

Mining association rules from imperfect data is a key challenge for real-world applications dealing with imperfect data, e.g., imprecise, missing data, etc. The ARM approach introduced in this paper enables to deal with imprecise data and derive imprecise rules under specific conditions (e.g. fixing both antecedent and consequent). Relying on evidence theory and Multiple Criteria Decision Analysis, this new framework enriches expressivity of existing works while providing a novel selection procedure for identifying most interesting rules according to several viewpoints. To this aim, several interestingness measures have been proposed, and used in a two-step selection procedure based on dominance relationships and Electre I. A restriction using *a priori* knowledge has also been proposed to focus and ease the mining process by incorporating symbolic knowledge defined into domain ontologies. To further improve the approach, additional measures of interestingness could be added. Future work related to subjective measures (e.g., user-oriented) would be particularly relevant to enrich the set of frequency-based measures that are currently involved in the approach. Studying the interactions between the measures would also be of interest. Finally, only an illustration using a simplified case study related to humanitarian projects analysis has been presented in this paper. Thorough algorithmic complexity and performance evaluations of the approach have to be discussed. Difficult challenges related to algorithmic complexity and efficiency issues of the procedure also have to be addressed in order to mine rules involving numerous attributes.

References

1. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. In: Acm sigmod record. vol. 22, pp. 207–216. ACM (1993)
2. Agrawal, R., Srikant, R., et al.: Fast algorithms for mining association rules. In: Proc. 20th int. conf. very large data bases, VLDB. vol. 1215, pp. 487–499 (1994)
3. Ait-Mlouk, A., Gharnati, F., Agouti, T.: Multi-agent-based modeling for extracting relevant association rules using a multi-criteria analysis approach. Vietnam Journal of Computer Science **3**(4), 235–245 (2016)
4. Bouker, S., Saidi, R., Yahia, S.B., Nguifo, E.M.: Ranking and selecting association rules based on dominance relationship. In: 2012 IEEE 24th international conference on tools with artificial intelligence. vol. 1, pp. 658–665. IEEE (2012)
5. Chen, M.C.: Ranking discovered rules from data mining with multiple criteria by data envelopment analysis. Expert Systems with Applications **33**(4), 1110–1116 (2007)
6. Choi, D.H., Ahn, B.S., Kim, S.H.: Prioritization of association rules in data mining: Multiple criteria decision approach. Expert Systems with Applications **29**(4), 867–878 (2005)
7. Dempster, A.P.: Upper and lower probabilities induced by a multivalued mapping. The Annals of Mathematical Statistics **38**, 325–339 (1967)
8. Djouadi, Y., Redaoui, S., Amroun, K.: Mining association rules under imprecision and vagueness: towards a possibilistic approach. In: 2007 IEEE International Fuzzy Systems Conference. pp. 1–6. IEEE (2007)

9. Dubois, D., Denoeux, T.: Conditioning in dempster-shafer theory: prediction vs. revision. In: Belief Functions: Theory and Applications, pp. 385–392. Springer (2012)
10. Fagin, R., Halpern, J.Y.: A new approach to updating beliefs. In: Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence. pp. 347–374. UAI '90, Elsevier Science Inc., New York, NY, USA (1991), http://dl.acm.org/citation.cfm?id=647233.760137
11. Figueira, J., Roy, B.: Determining the weights of criteria in the electre type methods with a revised simos' procedure. European Journal of Operational Research **139**(2), 317–326 (2002)
12. Geng, L., Hamilton, H.J.: Interestingness measures for data mining: a survey. ACM Computing Surveys **38**(3), 9–es (2006)
13. Hewawasam, K., Premaratne, K., Subasingha, S., Shyu, M.L.: Rule mining and classification in imperfect databases. In: 2005 7th International Conference on Information Fusion. vol. 1, pp. 8–pp. IEEE (2005)
14. Hong, T.P., Lin, K.Y., Wang, S.L.: Fuzzy data mining for interesting generalized association rules. Fuzzy sets and systems **138**(2), 255–269 (2003)
15. Kotsiantis, S., Kanellopoulos, D.: Association rules mining: A recent overview. GESTS International Transactions on Computer Science and Engineering **32**(1), 71–82 (2006)
16. Liu, B., Hsu, W., Chen, S., Ma, Y.: Analyzing the subjective interestingness of association rules. IEEE Intelligent Systems **15**(5), 47–55 (2000). https://doi.org/10.1109/5254.889106
17. Nguyen Le, T.T., Huynh, H.X., Guillet, F.: Finding the most interesting association rules by aggregating objective interestingness measures. In: Richards, D., Kang, B.H. (eds.) Knowledge Acquisition: Approaches, Algorithms and Applications. pp. 40–49. Springer Berlin Heidelberg, Berlin, Heidelberg (2009)
18. Roy, B.: Classement et choix en présence de points de vue multiples. Revue française d'informatique et de recherche opérationnelle **2**(8), 57–75 (1968)
19. Samet, A., Lefèvre, E., Yahia, S.B.: Evidential data mining: precise support and confidence. Journal of Intelligent Information Systems **47**(1), 135–163 (2016)
20. Seco, N., Veale, T., Hayes, J.: An intrinsic information content metric for semantic similarity in wordnet. In: Ecai. vol. 16, p. 1089 (2004)
21. Shafer, G.: A mathematical theory of evidence, vol. 42. Princeton university press (1976)
22. Silberschatz, A., Tuzhilin, A.: What makes patterns interesting in knowledge discovery systems. IEEE Transactions on Knowledge and data engineering **8**(6), 970–974 (1996)
23. Tan, P.N., Kumar, V., Srivastava, J.: Selecting the right interestingness measure for association patterns. In: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 32–41. ACM (2002)
24. Tobji, M.B., Yaghlane, B.B., Mellouli, K.: A new algorithm for mining frequent itemsets from evidential databases. In: Proceedings of IPMU. vol. 8, pp. 1535–1542 (2008)
25. Tobji, M.A.B., Yaghlane, B.B., Mellouli, K.: Frequent itemset mining from databases including one evidential attribute. In: International Conference on Scalable Uncertainty Management. pp. 19–32. Springer (2008)
26. Toloo, M., Sohrabi, B., Nalchigar, S.: A new method for ranking discovered rules from data mining by dea. Expert Systems with Applications **36**(4), 8503–8508 (2009)
27. Vaillant, B., Lenca, P., Lallich, S.: A clustering of interestingness measures. In: International Conference on Discovery Science. pp. 290–297. Springer (2004)