



Meta-association rules for mining interesting associations in multiple datasets



M.D. Ruiz^{a,b,*}, J. Gómez-Romero^{a,b}, M. Molina-Solana^{a,b}, J.R. Campaña^{a,b},
M.J. Martín-Bautista^{a,b}

^a CITIC-UGR, University of Granada, Spain

^b Department of Computer Science and A.I., University of Granada, Spain

ARTICLE INFO

Article history:

Received 26 February 2016

Received in revised form 4 August 2016

Accepted 5 August 2016

Available online 11 August 2016

Keywords:

Meta-association rules

Multiple datasets

Higher order mining

Fuzzy association rules

RL-theory

ABSTRACT

Association rules have been widely used in many application areas to extract new and useful information expressed in a comprehensive way for decision makers from raw data. However, raw data may not always be available, it can be distributed in multiple datasets and therefore there resulting number of association rules to be inspected is overwhelming. In the light of these observations, we propose meta-association rules, a new framework for mining association rules over previously discovered rules in multiple databases. Meta-association rules are a new tool that convey new information from the patterns extracted from multiple datasets and give a “summarized” representation about most frequent patterns. We propose and compare two different algorithms based respectively on crisp rules and fuzzy rules, concluding that fuzzy meta-association rules are suitable to incorporate to the meta-mining procedure the obtained quality assessment provided by the rules in the first step of the process, although it consumes more time than the crisp approach. In addition, fuzzy meta-rules give a more manageable set of rules for its posterior analysis and they allow the use of fuzzy items to express additional knowledge about the original databases. The proposed framework is illustrated with real-life data about crime incidents in the city of Chicago. Issues such as the difference with traditional approaches are discussed using synthetic data.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Association rules are a well-established technique for mining information from structured databases. They provide support for the identification of novel, potentially useful, and comprehensive knowledge in the form of implications $X \rightarrow Y$, which represent the joint co-occurrence of X and Y in the database. The data used for data mining is usually assumed to be *primary data* captured by some application, cleaned and preprocessed according to the necessities of the mining algorithm. However, in other applications, primary data is only available for a short time. That is the case of stream data which is not usually stored, but processed in real time and then deleted after storing the results. Beside this, primary data might not be available in several scenarios, or the owners may not be

authorized to share it. For instance, this is the case of law enforcing agencies facing security threats and other institutions that cannot disclose personal data. However, these organizations may be able to publish summarized or extracted knowledge. Beside this, the well known overwhelming number of association rules extracted by traditional algorithms prevent their inspection in a feasible time.

Therefore, the motivation of this work is twofold. On one hand, we aim to provide a meaningful tool for discovering associations among data which sometimes are not completely available at the beginning of the process. And, on the other hand, in many of these scenarios, the number of extracted rules exceeds the quantity of rules that can be inspected by a user in a reasonable amount of time. Then, the proposed tool has to provide a manageable set of rules to be analysed. In order to bring into play the advantages of knowledge discovery to these scenarios, we need to change our perspective to address pattern analysis instead of data analysis.

Our proposal is thus to combine association rule mining techniques in a pattern-analysis process for being able to find new information from different sets of rules previously extracted from multiple databases. In this case, the analyst has to inspect a different set of rules for each of the databases. Our proposal is to

* Corresponding author at: CITIC-UGR, University of Granada, Spain.

E-mail addresses: mdruiz@decsai.ugr.es

(M.D. Ruiz), jgomez@decsai.ugr.es (J. Gómez-Romero),

miguems@decsai.ugr.es (M. Molina-Solana), jesuscg@decsai.ugr.es (J.R. Campaña),

mbautis@decsai.ugr.es (M.J. Martín-Bautista).

employ meta-association rules for discovering the commonalities of the different sets of results recovered from multiple datasets. Meta-association rules are then extracted when some association rules frequently occur in a sufficient high percentage of databases, helping thus in the inspection process, since the user need only to analyse the final set of meta-association rules.

We further describe the problem and the proposal in Section 2. We continue reviewing the existent literature about this topic in Section 3. In Section 4, we summarized the concepts of association rule and fuzzy association rule, as well as the derived assessment measures needed for the comprehension of the paper. Section 5 describes our proposals for crisp and fuzzy meta-association rule mining. Section 6 shows the experimentation with synthetic and real-life data and discusses the obtained results. Finally, Section 7 points out some conclusions and prospective directions for future research.

2. Research problem statement and contributions

Roddick et al. defined in [32] higher order mining (HOM) as the *sub-field of knowledge discovery concerned with mining over patterns/models derived from one or more large and/or complex datasets*. These authors expounded the fundamentals of HOM and established a classification based on the mining techniques used for primary data and for processed data. Accordingly, we can distinguish several combinations of association discovery, clustering, classification, trend analysis, and rule extraction techniques. Our proposal combines techniques for association rule mining in both the primary and the processed data, thus producing what we have named *meta-association rules*. Meta-association rules are rules about rules; i.e. meta-association rules are built from regular association rules previously extracted from a group of databases about the same topic or with a similar structure. Therefore, meta-association rules can contain rules in their antecedent and/or consequent. Generally speaking, meta-association rule mining is a second order mining technique that can be employed when the available information is in the form of association rules. Our proposal also supports the incorporation of contextual information about the original databases to enrich the HOM process.

In a recent work [34], we presented an early approach for meta-association rule mining based on crisp rule mining. Crisp meta-rules are discovered by taking into account only whether an association has been previously mined from the original dataset or not, and not its degrees of fulfilment. This means that, to build meta-rules, a rule mined with a confidence of 0.9 has the same importance as another rule with a confidence of 0.5. Consequently, it is convenient to consider the available measures, like the support and the confidence, to quantify the importance of the rule within the HOM process. Fuzzy set theory allows a direct handling of the degrees of rules fulfilment, and by extension, to any attribute considered in the meta-association rule extraction process. A first approach was introduced in [33], where the value of the confidence was considered as the assessment degree of the extracted primary rules in what we called the fuzzy meta-database. Afterwards, *meta-association rules* are discovered from the compiled meta-database, obtaining associations between the more confident associations obtained in the first step of the process.

In this work, we go beyond this idea by considering other measures that enable the creation of a fuzzy meta-database according to different criteria such as the frequency of primary rules or their reliability. We compare and discuss the different approaches for mining meta-association rules, namely those taking into account only the presence/absence of the rule in the primary databases (A), and utilizing the degree of satisfiability of the rule in the primary databases (B). Depending on the measure employed in (B), we will discover meta-association rules with a slightly different meaning.

Additionally, we propose the use of the RL-theory for mining fuzzy meta-rules [15]. This framework enables the parallelization of the process by setting a set of levels, entailing thus a more scalable algorithm to extract fuzzy rules. We also compared the mentioned approaches using a synthetic dataset, and we illustrate the proposal with some experiments using an open dataset about criminal events in the city of Chicago.

The experiments and discussion will show that the proposed methods for mining meta-association rules offer some important benefits compared to previous works:

- meta-association rules convey new relevant information that cannot be obtained by classic association rules,
- proposed algorithm produces a more manageable set of rules for human inspection, and
- it allows the incorporation of contextual information to the mining process expressed in a more human-friendly format.

3. Related work

Mining meta-association rules is a higher order mining technique since it discovers associations between associations. To the best of our knowledge, in the literature, there is only one work that also develops the idea of meta-rules. In [3], fuzzy meta-rules are mined to discover temporal changes in rules that have been previously mined in different time periods. For that, the authors mine fuzzy rules over a set of association rules with a sequence of supports and a sequence of confidences in different time periods. Then, the fuzzy meta-rules capture the regularities governing how each association rule changes over time by using appropriate linguistic labels and obtaining for each rule a fuzzy meta-rule of the type *Change in support in a period t_1 = Fairly decrease \rightarrow Change in support next period = Highly decrease*. Compared to this approach, we do not consider sequences of supports/confidences, and conversely, we use the certainty factor. Despite this difference, we can configure our algorithms to work with rules obtained over the same database in different time periods. In this way, the meaning of our meta-rules would be the joint co-occurrence of rules over a percentage of the time-periods (measured by the FSupp and the FCF).

The terminology of meta-rules has been also employed in association rule mining but with different meanings. In [18], a meta-rule is considered as a template for guiding the association rule extraction process. In higher order mining, we have found two different approaches using also the concept of meta-rule. In [11], the authors called temporal meta-rules to those rules found from temporal rules (first-order rules) by a second inference process. However, the meta-rules obtained have the traditional formulation ($A \rightarrow B$) but with a small variability of the estimated confidence among different pre-established time structures. Needless to say, our concept of meta-rule is different to those proposed in these works [18,3,11], because we may have association rules in the antecedent and/or in the consequent of the meta-rule and its meaning does not correspond to the same concepts.

There are also other approaches of HOM that use association mining over a set of clusters [29–31]. In [29], the clustering process is used for grouping the items with a large number of intra-group correlations and a small number of inter-group correlations. This significantly helps to prune the search space for then mining the association rules. This kind of approaches is very interesting when the database is very large, because they scale very well when the number of rows and columns increases. A similar idea is developed in [31]. In this paper, the authors also use a soft clustering process and then, a concept hierarchy generation to obtain a more interpretable set of association rules. In [30], the variables are firstly clustered to obtain homogeneous groups of attributes, and

afterwards, association rules are mined in each cluster. With this approach, rules that are normally invisible for a typical association rule mining algorithm are now revealed after extracting them independently in each cluster.

The work of Chandanan and Shukla [7] is intended to remove redundant rules to improve the quality of rules. This is done by removing hierarchical duplicates in multi-level datasets. Contrary to this work, meta-association rules represent the frequent patterns already extracted from multiple datasets, that can be seen as a “summary” of patterns containing the most important association rules extracted in the majority of datasets. However, work in [7] and our approach are intended to reduce the set of rules that have to be inspected when dealing with multiple datasets.

There are some works dealing with the discovery of valuable information (e.g. changes, connections) from multiple datasets. When the temporal dimension comes into play, there exists approaches like the one developed in [24], for discovering changes and connections in two temporal information systems by employing different types of association rules using rough set theory. However, our approach is intended for the discovery of the existent commonalities from multiple datasets where a set of association rules have been already discovered, obtaining a set of meta-association rules which contains the more frequent rules in the majority of datasets. In addition to this, the meta-association rules may contain supplementary information about the original datasets by including appropriate attributes into the mining process. It is also important to highlight recent proposals for extracting associations in multidimensional datasets often stored as data warehouses. Many of these approaches have concentrated on adapting existent data mining techniques for discovering information taking into account the dimensionality of the data cubes in the extraction process [42,28]. Other proposals aim to employ the discovered associations to predict future trends. This is the case of [40] where they first identify the interesting information in the datacube by measuring the most informative dimensions to finally detect meaningful associations [28].

From other perspective, our approach can be also related to those works focusing on mining association rules in distributed databases; i.e., they extract a single set of rules from the distributed data [9,22,21,36]. Cheung et al. [9] developed a distributed algorithm for mining association rules. Jin and Agrawal [22] proposed new operators for querying frequent patterns in multiple datasets. Our proposal differs from these works because it does not extract a set of rules directly from the distributed datasets. Instead, it uses this output (different sets of rules previously extracted from every dataset) and then it finds the commonalities in the rulesets in form of meta-association rules.

There are other works [23,43] proposing new data mining approaches for multiple datasets. Zhu and Wu [43] developed an algorithm for constructing a hybrid frequent pattern tree in order to discover patterns in multiple databases. Jin et al. [23] improved several data mining algorithms, in particular the fp-tree-based for association meaning, by their parallelization on shared memory that could be used for multiple datasets mining. In both proposals, the output is a set of patterns relating the frequent items in the whole set of databases. However, meta-association rule mining find

associations between the frequent patterns previously extracted from the multiple datasets.

4. Background concepts

4.1. Association rule mining

Association rules is a well-established technique in data mining proposed by Agrawal [1]. We introduce hereinafter the necessary concepts for the understanding of the paper.

An association rule can be seen as an “implication” of the form $A \rightarrow B$. It represents the joint presence of itemsets A and B in transactions of D , assuming that: I is the set of items; $A, B \subseteq I$; $A \cap B = \emptyset$ and $A, B \neq \emptyset$. The left part of the rule is known as the antecedent of the rule, and the right part, as the consequent.

The intensity of an association rule is usually measured by the ordinary measures of support and confidence, also proposed in [1]. The *support* of a rule is the percentage of transactions from D that satisfy both parts of the rule (i.e. the joint probability $P(A \cup B)$):

$$\text{Supp}(A \rightarrow B) = \frac{|\{t \in D \mid A \cup B \subseteq t\}|}{|D|} \tag{1}$$

and the *confidence* of a rule measures the proportion of transactions that satisfy the consequent given that they also satisfy the antecedent (i.e. the conditional probability $P(B|A)$):

$$\text{Conf}(A \rightarrow B) = \frac{|\{t \in D \mid A \cup B \subseteq t\}|}{|\{t \in D \mid A \subseteq t\}|} \tag{2}$$

The general framework for the extraction of association rules consists in finding those rules whose support and confidence exceed the minimum thresholds *minsupp* and *minconf* fixed by the user. In such case, we say that $A \rightarrow B$ is *frequent* if $\text{Supp}(A \rightarrow B) \geq \text{minsupp}$, and *confident* if $\text{Conf}(A \rightarrow B) \geq \text{minconf}$.

Let us remember that association rules can be discovered from a transactional database or from a Boolean database which contains the same information by considering the items as pairs *(attribute, value/interval)* where v_i represents different values for attribute A_i or an interval when the range of values of A_i is discretized (see Table 1). In this case, 1 represents the presence of the item in the transaction and 0, its absence.

However, in the literature some drawbacks have been pointed out when using the confidence, proposing therefore new alternatives for assessing the validity of rules. A review of some of these measures can be found in [39] and an interesting study about the quality of these measures is contained in [4]. Particularly, in [5], the properties of confidence, conviction, and interest were studied concluding that they have some drawbacks such as the symmetry of the interest which gives the same measure in both directions of the rule, and the non-bounded domain of the conviction which makes the choice for the threshold and also to compare rules very difficult.

To avoid this kind of disadvantages, the *certainty factor* was proposed in [6] as an alternative of the mentioned measures using the Shortliffe and Buchanan’s theory of certainty factors [37].

Definition 1 ([6]). Let $\text{supp}(B) = |\{t \in D \mid B \subseteq t\}|/|D|$ be the support of the itemset B , i.e. the proportion of transactions in D satisfying

Table 1
Transactional (left) vs. Boolean database (right).

D	A_1	A_2	...	A_m	(A_1, v_1)	...	(A_1, v_p)	(A_2, v_2)	...	(A_m, v_s)
t_1	v_{11}	v_{21}	...	v_{m1}	1	...	0	1	...	1
t_2	v_{12}	v_{22}	...	v_{m2}	0	...	1	1	...	0
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\ddots	\vdots	\vdots	\ddots	\vdots
t_n	v_{1n}	v_{2n}	...	v_{mn}	1	...	0	0	...	0

B , and let $\text{Conf}(A \rightarrow B)$ be the confidence of the rule. The *certainty factor* of the rule, denoted as $CF(A \rightarrow B)$, is defined as:

$$\begin{cases} \frac{\text{Conf}(A \rightarrow B) - \text{supp}(B)}{1 - \text{supp}(B)} & \text{if } \text{Conf}(A \rightarrow B) > \text{supp}(B) \\ \frac{\text{Conf}(A \rightarrow B) - \text{supp}(B)}{\text{supp}(B)} & \text{if } \text{Conf}(A \rightarrow B) < \text{supp}(B) \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

The certainty factor yields a value in the interval $[-1, 1]$ that measures how the belief that B is in a transaction changes when it is known that A is in that transaction. Positive values indicate that the belief increases, negative values mean that the belief decreases, and 0 means no change. The certainty factor measure avoids the previously mentioned drawbacks [6,12] and in particular, it reduces the number of rules obtained by filtering rules corresponding to statistical independence or negative dependence. A theoretical study of its properties was also conducted in [13], supporting the previous studies [6,12].

Similarly to the confidence measure, we say that $A \rightarrow B$ is *certain* if $\text{Supp}(A \rightarrow B) \geq \text{minCF}$, where minCF is the minimum threshold for the certainty factor given by the user.

4.2. Fuzzy association rules

Since the late nineties, the use of fuzzy sets in association analysis has been proposed in different ways. The most representative proposals were split in two trends: (1) those based on the extraction of fuzzy associations in crisp relational databases [26,25,19] and (2) those based on the concept of fuzzy transaction [12]. In general, fuzzy sets in proposal (1) were provided in an attribute level by allowing a “soft” definition of interval boundaries, identifying fuzzy sets with linguistic labels and having therefore a more user-friendly representation of rules, and sometimes, providing a different granularity of the data. In this paper, we focus on proposal (2), where the provided data can be represented by a fuzzy transactional database, i.e. the items are satisfied to some extent in the transaction. A fuzzy transaction is then defined as a fuzzy subset of items.

Definition 2. [12] Let $I = \{i_1, \dots, i_m\}$ be a finite set of items. A fuzzy transaction is a non empty fuzzy subset $t \subseteq I$.

For every item $i \in I$ and every fuzzy transaction t , i belongs to t with degree¹ $t(i)$, where $t(i)$ is a real number in the interval $[0, 1]$. This definition is also extended to itemsets, $A \subseteq I$, being its membership degree to the fuzzy transaction t defined as $t(A) = \min_{i \in A} t(i)$. Not surprisingly, a crisp transaction is a special case of fuzzy transaction where every item in the transaction has a membership degree equal to 1 or 0, depending on whether they are in the transaction or not. Since crisp transactions are a special case of fuzzy transactions, a crisp association rule is a special case of fuzzy association rule. It is worth noting that the appearance of a fuzzy rule may look the same as that of a crisp rule, for instance when the description of attributes in the fuzzy transactional database does not involve fuzzy description of them. However, we want to stress that fuzzy rules are extracted from a fuzzy transactional database, instead of from a crisp or Boolean one, using adapted assessment measures. In this regard, several approaches have been studied for assessing the validity of fuzzy associations. We can highlight some approaches. In [20,16], a semantic approach using multiple-valued implication operators was studied. A method based on the evaluation of quantified sentences was proposed in [12]. In [38] the use of

a partition of the fuzzy transactions into examples, counterexamples, and irrelevant examples was proposed, for then defining the assessment measures according to the membership degree of the transactions to each set of the partition. A similar partition into positive and negative examples was used in [10] to study the semantics associated to the fuzzy rule depending on the validation measure used. In this paper, we employ a recent formal model called *Representation by Levels* theory that allows a consistent generalization of crisp assessment measures to the fuzzy case [35].

4.3. Fuzzy association rules using Representation by Levels

The *Representation by Levels* theory (RL for short) has been developed for managing imprecision in fuzzy transactional databases [35]. A distinguishing feature of the RL theory is that it satisfies all natural properties of Boolean algebras. Fuzziness is represented by means of RL-sets, which propose a more general structure that has fuzzy sets as a particular case.

Formally, a RL-set Λ is a finite set of levels $\Lambda = \{\alpha_1, \dots, \alpha_p\}$ verifying that $1 = \alpha_1 > \alpha_2 > \dots > \alpha_p > \alpha_{p+1} = 0, p \geq 1$. The levels α_i are values in the unit interval that mean possible levels of relaxation of a property. Value 1 corresponds to the most restrictive interpretation, i.e. the property must be fully satisfied; 0 means no restriction at all; and 0.5 is halfway between being totally strict and not strict at all. Then, a *RL-representation* associated to a property with some kind of fuzziness defined over X is represented by a pair (Λ, ρ) , where Λ is a RL-set and $\rho: \Lambda \rightarrow \mathcal{P}(X)$ is a function which applies each level into a crisp realization in that level. The operations of \wedge, \vee and \neg between two RL-sets are defined by levels using the usual intersection, union, and complement of crisp sets. It is worth highlighting that the representation by levels has some differences with respect to the classical notion of α -cut of a fuzzy set. Firstly, the representation by levels does not impose the levels to be nested, contrary to α -cuts. Secondly, the operations in a representation by levels are defined in each level independently for satisfying properties such as $\rho(\alpha)_{A \cap B} = \rho(\alpha)_A \cap \rho(\alpha)_B$.

In a previous work, we proposed a new framework for fuzzy association rule mining based on the RL theory [15]. We include here some basic concepts of this framework for the sake of completeness. The RL variant allows (1) extending the accuracy measures used for assessing rules in the crisp case, (2) keeping the main properties of the crisp model and (3) a parallelization of the extraction process by setting a fixed set of levels.

Formally, let $I = \{i_1, \dots, i_m\}$ be a finite set of items defined over a fuzzy database \tilde{D} . Let $A \subseteq I$ be an itemset in \tilde{D} , we denote by \tilde{A} the fuzzy set defined by $\tilde{A}(\tau) = t(A)$ which gives the degree of satisfiability of A in a given transaction. Let us remark that, in this framework, fuzzy sets (and therefore, the items in a fuzzy transaction) are interpreted from a conjunctive perspective.

Given two itemsets A, B in \tilde{D} , we consider their associated RL-sets \tilde{A} and \tilde{B} . For every $\alpha \in \Lambda_{\tilde{A}} \cup \Lambda_{\tilde{B}}$, we define a fourfold table (4f-table for short) noted by \mathcal{M}_{α_i} containing the cardinalities associated to the partition of \tilde{D} generated by the itemsets A and B as follows:

\mathcal{M}_{α_i}	\tilde{B}	$\neg\tilde{B}$
\tilde{A}	a_i	b_i
$\neg\tilde{A}$	c_i	d_i

where a_i, b_i, c_i and d_i are non-negative integers so that $a_i = |\rho_{\tilde{A} \wedge \tilde{B}}(\alpha_i)|, b_i = |\rho_{\tilde{A} \wedge \neg\tilde{B}}(\alpha_i)|$ and analogous expressions for c_i and d_i . Notice that for all $\alpha_i \in \Lambda_{\tilde{A}} \cup \Lambda_{\tilde{B}}, |\tilde{D}| = a_i + b_i + c_i + d_i$.

Based on this formulation, we can generalize every interest measure from the crisp to the fuzzy case [15]. In particular, the support FSupp , the confidence FConf , and the certainty factor FCF of a fuzzy rule $A \rightarrow B$ are defined by using this probability distribution in (4)–(6), respectively. In these functions, the RL-representations

¹ For the sake of simplicity we note by $t(i)$ the membership function associated to the fuzzy set t on the referential I .

Table 2
Set of fuzzy transactions \tilde{D} .

	i_1	i_2	i_3	i_4	i_5
t_1	1	0.2	1	0.8	0.9
t_2	1	1	0.8	0	0
t_3	0.4	0.1	0.7	0.6	0
t_4	0.6	0	0.4	0.4	0.5
t_5	0.4	0.1	0	0	0
t_6	0	1	0	0	0

Table 3
Fuzzy sets \tilde{A} and \tilde{B} .

	t_1	t_2	t_3	t_4	t_5	t_6
\tilde{A}	1	0.8	0.4	0.4	0	0
\tilde{B}	0.8	0	0.6	0.4	0	0

of both itemsets (A, B) must be normalized. $CF(a_i, b_i, c_i, d_i)$ is defined in (7), which actually corresponds to the crisp certainty factor of (3) expressed in terms of the 4f-table.

$$\begin{aligned}
 FSupp(A \rightarrow B) &= \sum_{\alpha_i \in \Lambda_{\tilde{A}} \cup \Lambda_{\tilde{B}}} (\alpha_i - \alpha_{i+1}) \left(\frac{|\rho_{\tilde{A} \wedge \tilde{B}}(\alpha_i)|}{|\tilde{D}|} \right) \\
 &= \sum_{\alpha_i \in \Lambda_{\tilde{A}} \cup \Lambda_{\tilde{B}}} (\alpha_i - \alpha_{i+1}) \left(\frac{a_i}{a_i + b_i + c_i + d_i} \right) \quad (4)
 \end{aligned}$$

$$\begin{aligned}
 FConf(A \rightarrow B) &= \sum_{\alpha_i \in \Lambda_{\tilde{A}} \cup \Lambda_{\tilde{B}}} (\alpha_i - \alpha_{i+1}) \left(\frac{|\rho_{\tilde{A} \wedge \tilde{B}}(\alpha_i)|}{|\rho_{\tilde{A}}(\alpha_i)|} \right) \\
 &= \sum_{\alpha_i \in \Lambda_{\tilde{A}} \cup \Lambda_{\tilde{B}}} (\alpha_i - \alpha_{i+1}) \left(\frac{a_i}{a_i + b_i} \right) \quad (5)
 \end{aligned}$$

$$FCF(A \rightarrow B) = \sum_{\alpha_i \in \Lambda_{\tilde{A}} \cup \Lambda_{\tilde{B}}} (\alpha_i - \alpha_{i+1}) CF(a_i, b_i, c_i, d_i) \quad (6)$$

$$CF(a, b, c, d) = \begin{cases} \frac{ad - bc}{(a + b)(b + d)} & \text{if } ad > bc \\ 0 & \text{if } ad = bc \\ \frac{ad - bc}{(a + b)(a + c)} & \text{if } ad < bc. \end{cases} \quad (7)$$

The calculation of the confidence may have indeterminations of the form “ $\frac{0}{0}$ ” when $|\rho_{\tilde{A}}(\alpha_i)| = 0$. This happens when there are no transactions satisfying the antecedent. To preserve the definition of fuzzy rule, we take the value 1 for that indetermination.

In the following section, we present an example to show the process of computing the support, the confidence, and the certainty factor of fuzzy rules using the proposed formal model.

4.4. Example

Let $I = \{i_1, i_2, \dots, i_5\}$ and \tilde{D} be the fuzzy database in Table 2. We take the itemsets $A = \{i_1, i_3\}$ and $B = \{i_4\}$, whose associated fuzzy sets \tilde{A} and \tilde{B} are shown in Table 3 with RL-representations $(\Lambda_{\tilde{A}}, \rho_{\tilde{A}})$ and $(\Lambda_{\tilde{B}}, \rho_{\tilde{B}})$. Remember that $t(A) = \min_{i \in A}(i)$. Tables 4 and 5 show the resulting RL-representations from applying the negation to both fuzzy sets and the partition of \tilde{D} using the four set conjunctions of \tilde{A}, \tilde{B} and their negations.²

² Since A is a fuzzy set, its representation by levels coincides with its α -cut representation, and therefore, the representation by levels is nested. However, when

Table 4
RL-representations of $\tilde{A}, \tilde{B}, \neg\tilde{A}$ and $\neg\tilde{B}$.

α_i	$\rho_{\tilde{A}}$	$\rho_{\tilde{B}}$	$\rho_{\neg\tilde{A}}$	$\rho_{\neg\tilde{B}}$
1	$\{t_1\}$	\emptyset	$\tilde{D} \setminus t_1$	\tilde{D}
0.8	$\{t_1, t_2\}$	$\{t_1\}$	$\{t_3, t_4, t_5, t_6\}$	$\tilde{D} \setminus t_1$
0.6	$\{t_1, t_2\}$	$\{t_1, t_3\}$	$\{t_3, t_4, t_5, t_6\}$	$\{t_2, t_4, t_5, t_6\}$
0.4	$\{t_1, t_2, t_3, t_4\}$	$\{t_1, t_3, t_4\}$	$\{t_5, t_6\}$	$\{t_2, t_5, t_6\}$
0.2	$\{t_1, t_2, t_3, t_4\}$	$\{t_1, t_3, t_4\}$	$\{t_5, t_6\}$	$\{t_2, t_5, t_6\}$

Table 5
RL-representations of the possible conjunctions among the fuzzy sets $\tilde{A}, \tilde{B}, \neg\tilde{A}$, and $\neg\tilde{B}$.

α_i	$\rho_{\tilde{A} \wedge \tilde{B}}$	$\rho_{\tilde{A} \wedge \neg\tilde{B}}$	$\rho_{\neg\tilde{A} \wedge \tilde{B}}$	$\rho_{\neg\tilde{A} \wedge \neg\tilde{B}}$
1	\emptyset	$\{t_1\}$	\emptyset	$\{t_2, t_3, t_4, t_5, t_6\}$
0.8	$\{t_1\}$	$\{t_2\}$	\emptyset	$\{t_3, t_4, t_5, t_6\}$
0.6	$\{t_1\}$	$\{t_2\}$	$\{t_3\}$	$\{t_4, t_5, t_6\}$
0.4	$\{t_1, t_3, t_4\}$	$\{t_2\}$	\emptyset	$\{t_5, t_6\}$
0.2	$\{t_1, t_3, t_4\}$	$\{t_2\}$	\emptyset	$\{t_5, t_6\}$

Table 6
4f($\mathcal{M}_{\alpha_i}, A, B, \tilde{D}$) for $A = \{i_1, i_3\}$ and $B = \{i_4\}$ where $\alpha_i \in \Lambda$.

	a_i	b_i	c_i	d_i
\mathcal{M}_1	0	1	0	5
$\mathcal{M}_{0.8}$	1	1	0	4
$\mathcal{M}_{0.6}$	1	1	1	3
$\mathcal{M}_{0.4}$	3	1	0	2
$\mathcal{M}_{0.2}$	3	1	0	2

To calculate the assessment measures, for every level in Λ , the 4f-tables \mathcal{M}_{α_i} are obtained³ (see Table 6), and from them, the values for the assessment measures are computed. Particularly, the support, confidence and certainty factor for $A \rightarrow B$ are $FSupp = 0.266$, $FConf = 0.5$ and $FCF = 0.33$. If we compute the support of $C \rightarrow E$, with the itemsets $C = \{i_1, i_5\}$ and $E = \{i_4\}$, we obtain $FSupp = 0.2$ and $FConf = 1$ (because the indetermination “ $\frac{0}{0}$ ” appears in the calculation).

In the following section, we describe the proposal emphasizing the different approaches for mining meta-association rules.

5. Meta-association rules for higher order mining

5.1. Problem statement

Availability of primary data may not be always possible due to different restrictions or only produced results are available for their inspection. Moreover, the data may be allocated in different places, usually because they are used independently in the place of origin. Isolated analyses in the data are often implemented in this kind of situations, but sometimes it is necessary to carry out a different analysis at a higher scale. Besides, the high size of obtained results complicates the inspection of the obtained rule sets. At this respect, having a tool for reducing the discovered rule sets in different datasets may help the user to obtain a more comprehensive analysis of what kind of associations are predominant in such set of databases.

The objective of meta-association rules is then to extract globally valid additional knowledge based on previously extracted information, in the form of association rules, that have

operating with other itemsets – in particular when the intersection of the complement are applied – the obtained result does not compulsorily need to be nested (see for instance Table 5)

³ Theoretically, we can compute the 4f-tables for each level in $\Lambda_A \cup \Lambda_B$ but in practice, we fix Λ into a finite set of levels in the unit interval because the possible different levels that can appear in a database could be very large.

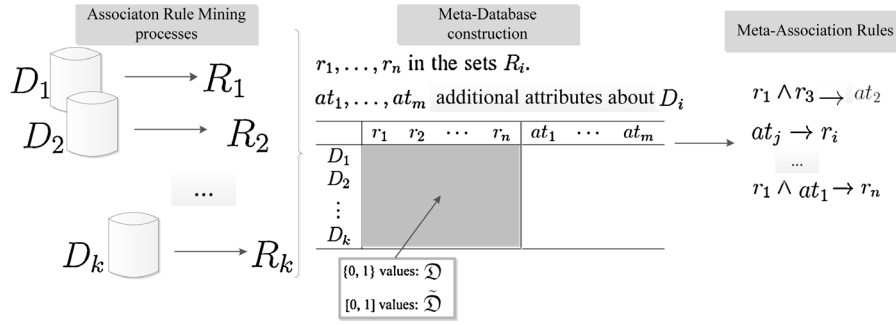


Fig. 1. Process flow: from original datasets to final meta-association rules.

similar structure and/or semantics, reducing thus the quantity of information to be examined.

5.2. Meta-association rule mining

The meta-association rule mining process is depicted in Fig. 1. Let D_1, D_2, \dots, D_k be k databases that may share some of their attributes. In the example described in Section 6, each database includes the crime incidents happened in one district of the city of Chicago; however, databases will generally correspond to different datasets with shared semantics but probably different structure. After applying a rule extraction procedure, we obtain k sets of association rules R_1, R_2, \dots, R_k , each R_i corresponding to a different D_i . We assume that these rules are crisp, but the procedure is analogous for other types of rules. The number of rules in R_1, R_2, \dots, R_k can be different, as well as the number of items in the antecedent or in the consequent of each rule. Interestingly, there may be some common rules in R_1, R_2, \dots, R_k . Without loss of generality, we also assume that the same threshold for the support and the confidence/certainty factor have been used for each dataset. We are interested in searching associations between the already extracted rules in the sets R_i . The associations mined from the primary databases D_i have at least two different values indicating their frequency of appearance (support) and also their validity (confidence, certainty factor, or other assessment measure).

For the purpose of being able to discover such kind of associations among rules, we create a structured database collecting this information. In addition, we can add new attributes that may influence or help to explain the new discovered associations among the rules extracted from D_i .

Definition 3. Let R_1, R_2, \dots, R_k be k sets of association rules extracted from the databases D_1, \dots, D_k , and at_1, \dots, at_m attributes about these databases. A *meta-database*, noted by \mathcal{D} , is a structured database where the transactions (rows) are associated to original databases and the attributes (columns) are associated to each different rule $r_i \in R_j$ for some $j \in \{1, \dots, k\}$ plus the attributes at_1, \dots, at_m .

An example of meta-database is shown in Fig. 1. Following the crime data example, the additional information can be attributes about district characteristics provided by a statistics office, such as the number of residents or the security perception index in each district of the city of Chicago. Once the meta-database is built, we can extract association rules from it. Since some of its attributes are rules, the rules extracted from the meta-database, called meta-association rules, may contain rules in the antecedent and/or the consequent of the meta-rule.

We have developed various strategies based on crisp rules and fuzzy rules, which are described below. A comparison between them is included in Section 6.

Table 7

Boolean meta-database obtained by compiling the association rules obtained from each dataset and the additional attributes.

\mathcal{D}	r_1	r_2	\dots	r_n	at_1	\dots	at_m
D_1	1	1	\dots	0	1	\dots	1
D_2	0	1	\dots	0	0	\dots	1
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots	\ddots	\vdots
D_k	1	0	\dots	1	1	\dots	0

5.2.1. Crisp meta-association rules

A first proposal for mining meta-association rules was developed in [34] by using crisp rules. In this work, we created a binary meta-database \mathcal{D} collecting the rules plus the extra information about the original datasets as shown in Table 7: if the rule r_i is found in D_j , then the value is 1; otherwise, it is 0. Afterwards, *crisp meta-association rules* are extracted from \mathcal{D} by employing a classical association rule mining algorithm. The obtained crisp meta-association rules are association rules that represent the frequent co-occurrence of items in the meta-database. Since items in the meta-database are rules (r) and attributes (at), the meta-association rules represent the co-occurrence of rules, rules and attributes, or attributes. Formally, we can obtain three types of meta-association rules:

- $r_i \rightarrow r_j$ where in the antecedent and/or the consequent we may have a conjunction of rules. An example of meta-rule could be: $r_1 \wedge r_2 \rightarrow r_3$.
- $at_i \rightarrow at_j$ where in the antecedent and/or the consequent we may have a conjunction of attributes, e.g. $at_1 \wedge at_2 \rightarrow at_3$.
- $r_i \rightarrow at_j$ or $at_j \rightarrow r_i$ where in the antecedent and/or the consequent we may have conjunctions of rules and attributes; i.e. meta-rules of the form $r_1 \wedge r_2 \wedge at_3 \rightarrow r_4$ or $r_1 \wedge r_2 \rightarrow at_3 \wedge r_4$ can be found.

Notice that the third type of rules is very interesting, because it relates *the appearance* of rules and attributes together in a percentage of the primary databases – given by the support of the meta-rule–, with a high reliability – provided by the assessment measure used.

It can be easily seen that this procedure for extracting crisp meta-rules has limitations, because it only takes into account whether a rule has been previously mined from a dataset or not. This implies that, in the Boolean meta-database, “regular” rules found with $\text{Supp}=0.1$ and with $\text{Supp}=0.3$ have the same importance. The same problem appears with rules found with $\text{CF}=0.5$ and $\text{CF}=1$. These measurements can be incorporated into the meta-database by using continuous items with values in the interval $[\text{minSupp}, 1]$ or $[\text{minCF}, 1]$. If we want to extract crisp meta-rules, the most direct alternative is to discretize the intervals by using items of the type $\langle r_i, (\text{CF}_1, \text{CF}_2) \rangle$, where $\text{CF}_j \in [\text{minCF}, 1]$ (respectively for the support). However, this approach has some problems

Table 8
Fuzzy meta-database compiling the obtained association rules and the additional attributes.

\hat{D}	r_1	r_2	...	r_n	at_1	...	at_m
D_1	0.3	1	...	0	0.9	...	1
D_2	0	1	...	0.7	0	...	0.2
...
D_k	0.8	0	...	0.9	1	...	0

due to the crisp boundaries of intervals. For example, given the values $CF(r_i)=0.75$ and $CF(r_j)=0.76$, and the intervals $(0.5, 0.75]$ and $(0.75, 1]$, the CF values would lie in different intervals even though they are very similar. This issue motivates a different representation for continuous values. Fuzzy sets are an adequate option, since fuzzy meta-rules can be mined from a meta-database where the items are satisfied to some extent.

5.2.2. Fuzzy meta-association rules

Fuzzy transactional databases support the extraction of association rules from a continuous representation of values by means of the so-called fuzzy association rules. Following the same notation used in the previous section, we can create two different fuzzy meta-databases following the support/CF framework: \hat{D}_S based on the support, and \hat{D}_{CF} based on the certainty factor. The structure of these databases is depicted in Table 8. If we consider the certainty factor, the value in column r_i and row D_j represents the certainty factor of the rule r_i in database D_j , which is a value in the interval $[minCF, 1]$ (analogously for the support). Likewise, we can use fuzzy sets for representing the additional attributes; e.g., a “high security perception” with a degree of 0.8, representing that there is a high percentage of population with a high security perception.

Once the fuzzy meta-database has been created, a fuzzy association rule mining algorithm based on the RL theory is applied to obtain the fuzzy meta-rules. We can also distinguish three different types of meta-rules; namely, meta-rules that relate only rules, meta-rules that relate only attributes, and meta-rules that relate rules and attributes. Depending on the measure used to build the fuzzy meta-database, meta-rules have semantics with subtle differences. When the support is employed to create \hat{D}_S , the fuzzy meta-rules represent the relation among rules and/or attributes that have been extracted with a *high support*. Whereas if the certainty factor is considered, \hat{D}_{CF} , the extracted fuzzy meta-rules represent the relation among rules and/or attributes that have a *high reliability* in the primary datasets.

This opens a variety of possibilities when mining meta-association rules. One may be interested in only considering meta-rules containing regular rules with a high support or containing regular rules with a high reliability. If we want to obtain meta-rules imposing both aspects, one possibility could be to consider only those meta-rules that have been discovered in both cases, i.e. the common set of meta-rules obtained using \hat{D}_S and \hat{D}_{CF} .

One important consideration about the use of the support in \hat{D}_S is that the values collected in the meta-database are often very low. This implies that very few meta-rules will be obtained from \hat{D}_S . In order to solve this problem, we normalize the support to the range of the maximum support value. To distinguish this case, we use \hat{D}_{NS} when the support is normalized in the meta-database.

6. Algorithm and experimental evaluation

6.1. Algorithm and implementation issues

Many papers were devoted to adapt ordinary association rule algorithms like Apriori or FP-Growth to the fuzzy case. This is the case for instance of [41], where a weighted FP-Growth was

proposed for mining fuzzy rules. In [27], the authors proposed a new algorithm called FAR-Miner with a faster performance of the fuzzy Apriori version on very large datasets. In [17], the FCB and PFCB algorithms were developed to obtain fuzzy itemsets using cluster tables for then obtaining the fuzzy rules. In these cases, the algorithms used adapted fuzzy versions of the support-confidence framework. However, in [2], the proposed algorithm used the adjusted difference, an assessment measure that does not need to supply any thresholds. There are also other works following different strategies like in [8], where a taxonomy of linguistic labels and their definitions is used for then mining the fuzzy rules.

In contrast to the mentioned approaches, we use the developed algorithm in [15], based on the Representation Level theory presented in Section 4.3. One of the advantages of this algorithm is that it allows us to compare crisp and fuzzy approaches, since using the RL theory we have generalized in a straightforward way the crisp assessment measures to the case of fuzzy transactional databases. For the sake of completeness, we provide a brief description of the algorithms used for mining crisp and fuzzy association rules and for crisp and fuzzy meta-rules which are based on the previous ones. Both algorithms which extract crisp association rules and fuzzy association rules are detailed in Algorithms 1 and 2, respectively. The latter is based on the RL model presented in Section 4.3. The representation by levels allows us to reduce the problem of mining fuzzy rules to that of mining crisp rules in each level. The fuzzy assessment values FSupp and FCF obtained in each level are aggregated by means of the formulas (4) and (6). Both algorithms use an itemset representation based on bit sets [15,14] that speeds up logical operations with Boolean data.

Algorithm 1. Crisp association rule mining procedure

Input: Boolean D , $minsupp$, $minCF$
Output: R (Set of crisp association rules obtained from D)

```

1: #  $D$  preprocessing
2: Read  $D$  and store the items  $I$ 
3: Transform  $D$  into a Boolean database
4: Encode database into a vector of BitSets

5: # Mining crisp very strong rules
6: Compute the candidate set  $C$  of frequent itemsets  $Supp(X) \geq minsupp$ 
7: Store the indexes in the BitSet of  $X \in C$  and the  $Supp(X)$  value
8: Compose the rule with  $X, Y \in C$ 

9: if  $Supp(X \rightarrow Y) \geq minsupp$  and  $CF(X \rightarrow Y) \geq minCF$  then
10:   The rule is a very strong rule
11: end if

```

Algorithm 2. Fuzzy association rule mining procedure

Input: Fuzzy \hat{D} , $minFSupp$, $minFCF$, Λ
Output: FR (Set of fuzzy association rules)

```

1: #  $\hat{D}$  preprocessing
2: Read  $\hat{D}$  and store the items  $I$ 
3: Transform  $\hat{D}$  into  $p$  Boolean databases ( $p$  is the number of levels in  $\Lambda$ )
4: Encode databases into  $p$  vectors of BitSets (one vector for each  $\alpha_i \in \Lambda$ )

5: # Mining fuzzy rules
6: for all  $\alpha_i \in \Lambda$  such that  $1 \leq i \leq p$  do
7:   Repeat steps 1–11 of Algorithm 1 to obtain the set of crisp rules in each level.
8:   Read all the found rules in level  $\alpha_i$ 
9:   Compute progressively the measures FSupp and FCF using formulas (4) and (6)
10: end for
11: Collect those rules exceeding the imposed thresholds ( $minFSupp$  and  $minFCF$ )

```

Algorithms 3 and 4 respectively implement the complete processes for the discovery of crisp and fuzzy meta-rules. From the original databases D_1, \dots, D_k , crisp association rules⁴ are extracted with Algorithm 1. Next, the meta-database \mathfrak{D} (Algorithm 3) or the fuzzy meta-database $\hat{\mathfrak{D}}$ (Algorithm 4) is created. In this step, additional features may be added as attributes to the meta-database. As previously explained, in the latter case, the attributes of the meta-database can be modelled as fuzzy sets. Finally, crisp or fuzzy meta-association rules are respectively extracted using again Algorithm 1 or Algorithm 2. Obviously, when the initial datasets (primary data) are not available, the meta-rule extraction procedure would start at step 4 (from the initial sets of rules).

Algorithm 3. Crisp meta-association rule mining procedure

Input: $D_1, \dots, D_k, at_1, \dots, at_m, minsupp, minCF$
Output: MR (Set of crisp meta-association rules)

```

1:   for all  $D_i$  such that  $1 \leq i \leq k$  do
2:     Repeat steps 1–11 of Algorithm 1 to obtain from each  $D_i$  the set
       of crisp association rules  $R_i$ 
3:   end for

4:   #  $\mathfrak{D}$  creation
5:   Compile all different rules from  $R_1, \dots, R_k \rightarrow \{r_1, \dots, r_n\}$ 
6:   Create the Boolean meta-database  $\mathfrak{D}$  using  $\{r_1, \dots, r_n\}$  and  $at_1, \dots,$ 
        $at_m$  (see Table 7)

7:   # Mining meta-association rules
8:   Repeat steps 1–11 of Algorithm 1 to extract meta-association rules
       from  $\mathfrak{D}$ 

```

Algorithm 4. Fuzzy meta-association rule mining procedure

Input: $D_1, \dots, D_k, at_1, \dots, at_m, minFSupp, minFCF$
Output: FMR (Set of fuzzy meta-association rules)

```

1:   for all  $D_i$  such that  $1 \leq i \leq k$  do
2:     Repeat steps 1–11 of Algorithm 1 to obtain from each  $D_i$  the set
       of crisp association rules  $R_i$ 
3:   end for

4:   #  $\hat{\mathfrak{D}}$  creation
5:   Compile all different rules from  $R_1, \dots, R_k \rightarrow \{r_1, \dots, r_n\}$ 
6:   Create the fuzzy meta-database  $\hat{\mathfrak{D}}$  using  $\{r_1, \dots, r_n\}; at_1, \dots, at_m$ 
       and their associated support/normalized support/FCF (see Table 8)

7:   # Mining fuzzy meta-association rules
8:   Repeat steps 1–11 of Algorithm 2 to mine fuzzy meta-rules from  $\hat{\mathfrak{D}}$ 

```

The computational complexity of the algorithms depends on the number of transactions and items. In Algorithm 3, the first step (lines 1–3) is $\mathcal{O}(n2^{|I|})$ for each original D (being n the number of transactions of D and $|I|$ the number of different items), whereas the second step (lines 4–8) is $\mathcal{O}(k2^{m+r})$, being k the number of databases, m the number of extended items (new items obtained from the additional attributes), and r the number of different rules obtained in the first step. Likewise, the complexity of Algorithm 4 is $\mathcal{O}(pk2^{m+r})$, where p is the number of levels considered. This complexity can be reduced to $\mathcal{O}(k2^{m+r})$ after parallelization of lines 5–10 of Algorithm 2 to process each level independently. (It goes without saying that the first step can be also parallelized in both Algorithms 3 and 4.) In practice, we have experienced quite fast executions of the experimental tests with different choices for the $minSupp$, $minCF$, $minFSupp$ and $minFCF$ (See Fig. 4). Regarding

memory requirements, the footprint is not high thanks to the use of the binary representation.

6.2. Experimental evaluation

For the experimental evaluation of our proposal, we present two different use cases. The first one uses synthetic data to analyze the differences of mining meta-rules in partitioned datasets versus mining regular rules in a single dataset that includes all the transactions. The second one employs real data collecting crime incidents reported by the police in the city of Chicago in 2012 to compare the crisp and the fuzzy approaches when mining meta-rules. This dataset is open and accessible in the URL <https://data.cityofchicago.org/>.

For the experiments, we have used a desktop computer equipped with a 2.5GHz Pentium Dual Core processor and 3GB of RAM running Java 8 on Windows 7. Without lack of generality, in order to obtain readable rules, we have limited the rules obtained in the first step of the process to have one item in the antecedent and one in the consequent. For the meta-association rules, we allow two items at most in the antecedent and the consequent.

6.2.1. Use case 1: synthetic data

In order to emphasize the differences and similarities between rules and meta-rules, we have designed two different sets of experiments carried out on a synthetic dataset $synth$ composed by 8 artificial datasets, each one with 15 transactions and 6 attributes. Each attribute has 6 possible values, having a total of 36 items of the form $(attribute, value)$. These datasets were created in order to obtain a manageable set of rules and meta-rules for being able to compare them and analyse the differences among them. Particularly,

- we mined rules in a database formed by the aggregation of the transactions in each of the 8 datasets into a single one; and
- we mined meta-rules in the meta-database resulting from the incorporation of the discovered rules in the 8 different datasets.

The intention of this is to highlight that the information obtained by the rules using the traditional approach and the one proposed here, using meta-rules, is different and that they yield different results. In both cases, we set the thresholds low enough to obtain many association rules and meta-rules in order to analyse their differences: 0.2 for the minimum support and the minimum CF.

We have observed the following facts:

- Some of the rules extracted in experiment (a) did not appear in the antecedent or the consequent of a meta-rule found in experiment (b). For instance, a rule $r_1 : A \rightarrow B$ found in (a) is not found in any meta-rule of the form $r_1 \rightarrow r_2, r_2 \rightarrow r_1, r_1 \rightarrow at_2$ or similar combinations. This is mainly due to the following: although a rule is extracted in the compiled database, it does not mean that it appears in a sufficiently high percentage of the original databases and therefore, it cannot appear in any meta-rule.
- Although one can expect that some groups of items appearing in rules extracted from the compiled dataset (a) should also appear in meta-rules obtained from (b), this is not true in general. In fact, we have discovered many meta-rules that do not have the same combination of items in any regular rule in the compiled database and vice versa. For example, if a meta-rule $(it_1 \rightarrow it_2) \rightarrow (it_3 \rightarrow it_4)$ is extracted in (b), we have not found a rule in (a) containing the same items together, that is, rules like $it_1, it_2, it_3 \rightarrow it_4$ or $it_1, it_2 \rightarrow it_3, it_4$, or any combination changing the order of appearance of the items have been extracted in (a). A possible explanation is that when aggregating the databases into a single one, some of the rules that were frequent/certain in the

⁴ We have considered that the regular rules are always crisp, but the same process can be also applied if fuzzy rules are provided, since the same kind of information is available: rule, fuzzy support, and fuzzy CF.

original datasets (i.e. they can probably appear as a part of a meta-rule) are not found in the compiled database, since they have not been extracted in a sufficiently high percentage of datasets in (a) (i.e. their support as a meta-rule is low). In addition, we want to stress that a meta-rule represents association among rules, not among the items they involve, so in principle, there is not a direct relation between the obtained meta-rules and the rules found in the compiled database.

This leads us to affirm that the results, and therefore the information, obtained when mining meta-rules differ to those obtained with regular rules when compiling together all the datasets into a single one. Beside this, it is worth noting that in some scenarios, we do not have access to the original datasets, and thus, we cannot compile the original data into a single dataset. In these cases, building the meta-database enables obtaining more information in the form of meta-associations.

6.2.2. Use case 2: City of Chicago crime data

The City of Chicago dataset consists of 22 databases according to the districts of the city. The size of the databases is described in Fig. 2, and they collect information about the quarter of the year in which the incident happened, the day period (morning, afternoon, evening, and night), the crime description according to police standard protocols, a description of the location, if there was an arrest associated to the crime, and if the crime happened in a domestic environment. We are interested in studying the possible relation among rules involving crime attributes in a neighbourhood and the educational system. For that aim, we have considered the following additional attributes of schools aggregated by district in the meta-database: number of students in the district, number of misconducts notified in the district and perceived safety index. Note that, in total, we are dealing with more than 330,000 transactions and around 6400 items.



We have conducted diverse experiments using different values for pairs of $minSupp - minCF$, and $minFSupp - minFCF$ thresholds using Algorithms 3 and 4 in order to analyse the following aspects of our proposal:

- Comparison of the number of crisp and fuzzy meta-rules.
- Comparison of the number of fuzzy meta-rules obtained using the support and the normalized support for the fuzzy meta-database.
- Comparison of execution times for the different approaches.

In Fig. 3, the number of meta-rules obtained for $minCF$ and $minFCF$ values $\{0.2, 0.3, \dots, 0.8\}$ (in logarithmic scale) is depicted. The left chart shows the results for $minSupp = minFSupp = 0.05$, and the right chart, for $minSupp = minFSupp = 0.1$. From these graphs, we can see the following:

- The number of crisp meta-rules is generally larger than the number of fuzzy meta-rules. In particular, the order of magnitude is very high since the graph is in a logarithmic scale.
- The number of crisp meta-rules is drastically reduced when the threshold of the certainty factor is slightly increased; decreasing, in some cases, several orders of magnitude.
- In the fuzzy case, the reduction of meta-rules is not so high because of the use of the rule validation measures in the meta-databases $\hat{\mathcal{D}}_S$, $\hat{\mathcal{D}}_{NS}$, and $\hat{\mathcal{D}}_{CF}$.

In this regard, the conducted experiments result in a more manageable set of meta-rules for user inspection in the fuzzy case, e.g. for $minsupp = 0.1$ and $minCF = 0.2 \sim 0.4$ the number of crisp meta-rules in $\hat{\mathcal{D}}$ is comprised between 3000 and 500 whilst for fuzzy meta-rules is below 130. This is extremely important when analysing association rules, since very often, the amount of associations to be inspected can overwhelm the user.

District Id	# transactions	# items
01	12 160	268
02	13 481	286
03	17 730	297
04	19 847	302
05	15 296	290
06	19 301	303
07	20 195	298
08	22 493	333
09	16 704	305
10	15 071	301
11	21 873	311
12	15 930	296
14	12 580	296
15	14 434	285
16	10 851	296
17	9 702	282
18	14 247	272
19	15 689	284
20	5 694	252
22	10 797	275
24	9 538	267
25	19 726	301

Fig. 2. Left: Image from <https://commonsense.uchicago.edu/page/crime-information-and-statistics> about police districts in the city of Chicago. Right: Description of the databases.

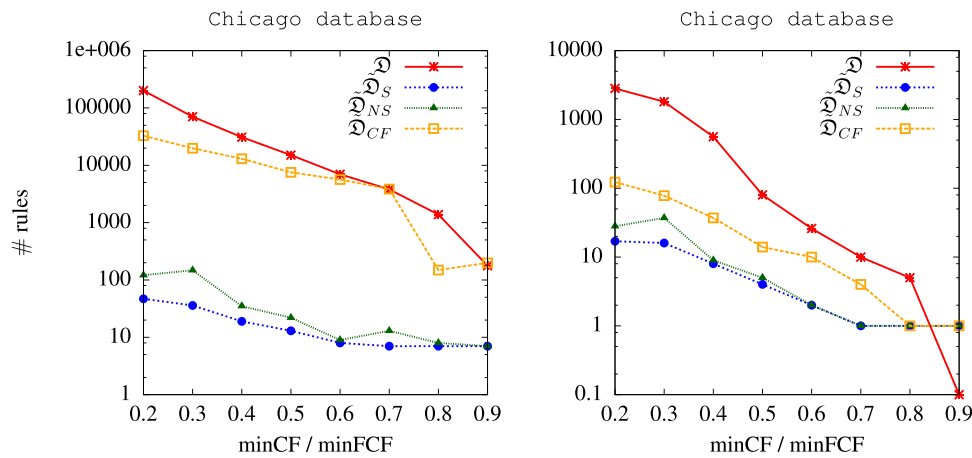


Fig. 3. Number of crisp/fuzzy meta-association rules (y -axis in a logarithmic scale) vs $minCF/minFCF$ (x -axis) when $minSupp = 0.05$ (left) and $minSupp = 0.1$ (right)

Concerning the application of normalized support, as expected, we can observe that a higher quantity of fuzzy meta-rules are mined when the normalized support is applied. We believe that normalizing the support is an appropriate choice indeed, because in several scenarios, the minimum support is set very low. Smaller values of $FSupp$ lead to higher values of the normalized support when the maximum support value is small. For instance, if we consider the following values of support $\{0.1, 0.3, 0.2\}$ and $\{0.1, 0.25, 0.2\}$, when normalizing both, we obtain $\{0.33, 1, 0.67\}$ and $\{0.4, 1, 0.8\}$ respectively. This phenomenon explains the presence of peaks in the graph for the \tilde{D}_{NS} meta-database, since the maximum support used to normalized the fuzzy meta-database changes depending on the imposed thresholds.

Regarding the execution time, in Fig. 4, the time in seconds for the same parameters as in the previous figure is shown. These times comprise the whole process of mining the rules from the original datasets, building the meta-database, and finally, mining the meta-rules. From the graphs, it can be observed that the execution times do not exceed an hour for very low thresholds, lasting in most cases less than a minute.

Table 9 contains some non-trivial meta-rules⁵ obtained when $minSupp = minFSupp = 0.05$ and $minCF = minFCF = 0.5$, which offer some interesting results. We tried to select different types of meta-association rules for showing the variety of the obtained results, since they can reveal relations among rules, among attributes, and among rules and attributes together. For instance, the first meta-association rule states that there is an association with certainty ($CF = 0.658$) between the very high number of misconducts, the rule (possession of cannabis ≤ 30 grams \rightarrow Domestic=false), and the perception of low safety-index in a district. That means that when there is a very high number of misconducts, then it is frequent and reliable to have a low perception of security and a relation between the low possession of cannabis crime and its occurrence in a non-domestic environment.

6.3. Discussion

Meta-association rules have several advantages compared to traditional association rules. The first one, and the most evident one, is that they can be built over secondary data when primary data is not available. In addition, our approach allows the incorporation of more information about the original datasets when available. Secondly, they are useful to analyse associations coming frequently

in a group of databases, and therefore, they help to obtain a higher abstraction layer that may offer some clues about why or how some phenomena (association rules) are happening.

We want to highlight that there is a subtle difference in the meaning of meta-association rules obtained depending on the meta-database employed. Crisp meta-rules represent the co-occurrence among regular rules and attributes, whilst in fuzzy meta-rules, the co-occurrence is weighted by the $CF/Supp$ of the initial rules, and therefore, when a fuzzy meta-rule is mined, it also indicates that the rules involved in the meta-rule have a high $CF/Supp$. Let us illustrate this with an example. Let us suppose that rules r_1 and r_2 are found in every original dataset with a CF around 0.1. If we mine crisp meta-rules, we will probably obtain a meta-rule of the type $r_1 \rightarrow r_2$ since it only takes into account the presence of both rules in all the datasets; on the opposite, if we mine fuzzy meta-rules using \tilde{D}_{CF} it will not be extracted due to its low CF in all the databases. That is the case for instance of the second meta-rule in Table 9, which has been extracted in \tilde{D} and not in \tilde{D}_{CF} . This also explains the overwhelming number of crisp meta-rules found in the experiments (see Fig. 3).

It is also worth mentioning that meta-rules extracted from \tilde{D}_{CF} represent a relation among rules with a high CF , while those found from $\tilde{D}_S/\tilde{D}_{NS}$ represent relations between regular rules that were extracted with a high support/normalized support. For instance, suppose that rules r_1 and r_2 are extracted in most of the datasets with a high support, but their CF is very low. The meta-rule $r_1 \rightarrow r_2$ is likely to be obtained when using \tilde{D}_S or \tilde{D}_{NS} but not employing \tilde{D}_{CF} . This is the case of the third rule in Table 9 that was found in \tilde{D}_{NS} but not in \tilde{D}_{CF} . At this respect, the information given is different and both methods, \tilde{D}_{NS} and \tilde{D}_{CF} , can be used in a complementary way in order to achieve more information about relations among rules.

Besides, the user can provide part of his/her knowledge about the original datasets by means of new fuzzy items of the form $\langle attribute, linguistic label \rangle$ and a degree of satisfiability of that item. E.g. item $\langle poverty-index, low \rangle$ is fulfilled with a degree of 0.9 in a certain district, meaning that we are 90% sure that the poverty index is low in the district. This kind of representation is more natural, and easier to interpret for humans than an item of the type $\langle attribute, value \rangle$ such as $poverty-index = 1.02$.

It is important to highlight that meta-association rules give a summarized view of what have been extracted in the initial datasets. In particular, if the initial datasets are very sparse, very different rules will be found separately in each dataset. This will result in very few meta-association rules, or at least meta-rules with very low support and CF . (This actually gives us a rough explanation of

⁵ Desc. stands for Description, f for false, and t for true.

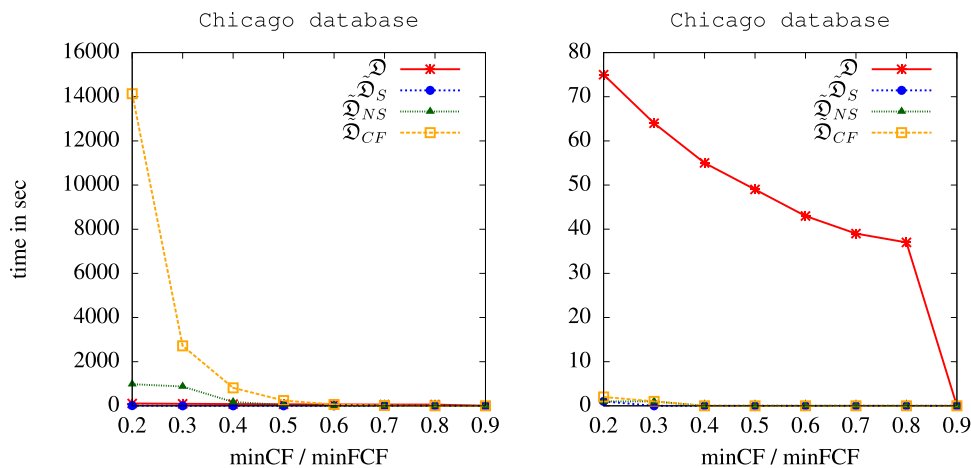


Fig. 4. Time in seconds for mining crisp/fuzzy meta-association rules (y-axis) vs $minCF/minFCF$ (x-axis) when $minSupp=0.05$ (left) and $minSupp=0.1$ (right) for the different approaches.

Table 9
Examples of meta-association rules found in the City of Chicago dataset for $minSupp = minFSupp = 0.05$ and $minCF = minFCF = 0.5$. The symbol “√” is used when the meta-rule has been found and “x” when it has not.

Antecedent	Consequent	Supp/FSupp	CF/FCF	\mathcal{D}	$\tilde{\mathcal{D}}_{CF}$	$\tilde{\mathcal{D}}_{NS}$	$\tilde{\mathcal{D}}_S$
Number-of-Misconducts= Very high	(Crime-Desc.=POSS: CANNABIS \leq 30GMS \rightarrow Domestic=f) AND Safety-Index=Low	0.136	0.658	x	√	x	x
(Crime-Desc.= \leq 500\$ \rightarrow Domestic=f) AND (Crime-Desc.=TO VEHICLE \rightarrow Arrest=false)	(Location-Description= STREET \rightarrow Domestic=f)	0.455	0.778	√	x	x	x
(Arrest=t \rightarrow Domestic=f) AND Safety-Index=Low	Number-of-Students= Medium	0.064	0.7	√	x	√	x
Number-of-Students= Low AND Number-of-Misconducts= Very high	Safety-Index=Low	0.091	1	x	√	√	√

what is happening in the initial datasets.) In this situation, it could be useful for the user to manually inspect the rules obtained from the initial datasets.

To sum up, we can stress that:

- According to the results of the crisp and fuzzy algorithms shown in the previous section, the number of fuzzy meta-rules is in general more manageable than the number of crisp meta-rules.
- The process to mine fuzzy meta-rules is more adequate, since we take into account the degree of importance of the initial rules to build the meta-database.
- The semantics of the meta-rules are different depending on the assessment measure used to build the meta-database.
- The process to mine fuzzy meta-rules allows us to use fuzzy items for expressing additional knowledge about the original databases.

7. Conclusions and future work

Meta-association rules are proposed in this paper as rules about rules that can be employed when the available information about several datasets is in the form of association rules. The main advantage of the meta-rules is that they can extract information that is not achieved by simple regular rules and that they allow the analysis of associations coming frequently in a group of databases, that in many cases, may overwhelm the user. Moreover, meta-association

rules can incorporate contextual information related to the original datasets in order to enrich the associations obtained by the meta-rules.

We have proposed different types of meta-association rules: crisp and fuzzy meta-rules. Fuzzy meta-rules take advantage of the assessment measures provided when mining rules from the original datasets. We have compared the different approaches obtaining that, in general, mining fuzzy meta-rules gives a more manageable set of rules for its posterior analysis and that they allow the use of fuzzy items to express additional knowledge about the original databases.

In the near future we plan to develop an intuitive interface for setting the input parameters as well as showing obtained meta-rules in a understandable way. Moreover, this work opens several issues to be addressed in the future. For instance, when the description of attributes is not the same in all the databases, we may have a problem, since the set of regular rules obtained in the first step of the process may contain different attribute names although they refer to the same item. A solution to this could be to have a “dictionary” or a knowledge repository that links similar concepts. Another interesting application is to consider meta-rules for analysing streaming data. In this case, the proposed process could be adapted to take into account different time granularities. We plan to advance in this line for analysing data collected from sensors in different time periods.

Acknowledgements

The research reported in this paper was partially supported by the Andalusian Government (Junta de Andalucía) under projects P11-TIC-7460 and P10-TIC-6109; by the University of Granada under the Programme “Proyectos para la incorporación de jóvenes doctores a nuevas líneas de investigación”, by the Spanish Ministry for Economy and Competitiveness by the project grant TIN2015-64776-C3-1-R; and by the ePOOLICE project FP7-SEC-2012-312651, funded from the European Union in the Seventh Framework Programme under grant agreement No. 312651 and the Energy IN TIME project under grant agreement No. 608981.

References

- [1] R. Agrawal, T. Imielinski, A. Swami, Mining associations between sets of items in massive databases, in: ACM-SIGMOD International Conference on Data, 1993, pp. 207–216.
- [2] W.H. Au, K.C.C. Chan, FARM: a data mining system for discovering fuzzy association rules, in: Fuzzy Systems Conference Proceedings, 1999 IEEE International, vol. 3, 1999, pp. 1217–1222.
- [3] W.H. Au, K.C.C. Chan, Mining changes in association rules: a fuzzy approach, *Fuzzy Sets Syst.* 149 (2004) 87–104.
- [4] J.L. Balcázar, F. Dogbey, Evaluation of association rule quality measures through feature extraction, in: A. Tucker, F. Höppner, A. Siebes, S. Swift (Eds.), *Advances in Intelligent Data Analysis XII, Lecture Notes in Computer Science*, vol. 8207, Springer, Berlin, Heidelberg, 2013, pp. 68–79.
- [5] F. Berzal, I. Blanco, D. Sánchez, M.A. Vila, A new framework to assess association rules, in: LNCS 2189, 2001, pp. 95–104.
- [6] F. Berzal, M. Delgado, D. Sánchez, M.A. Vila, Measuring accuracy and interest of association rules: a new framework, *Intell. Data Anal.* 6 (3) (2002) 221–235.
- [7] A.K. Chandanan, M.K. Shukla, Removal of duplicate rules for association rule mining from multilevel dataset, *Procedia Comput. Sci.* 45 (2015) 143–149.
- [8] G. Chen, Q. Wei, Fuzzy association rules and the extended mining algorithms, *Inform. Sci.* 147 (2002) 201–228.
- [9] D.W. Cheung, V.T. Ng, A.W. Fu, Y. Fu, Efficient mining of association rules in distributed databases, *IEEE Trans. Knowl. Data Eng.* 8 (6) (1996) 911–922.
- [10] M. De Cock, C. Cornelis, E.E. Kerre, Elicitation of fuzzy association rules from positive and negative examples, *Fuzzy Sets Syst.* 149 (1) (2005) 73–85.
- [11] P. Cotofrei, K. Stoffel, From temporal rules to temporal meta-rules, in: Kambayashi, et al. (Eds.), *Data Warehousing and Knowledge Discovery, Lecture Notes in Computer Science*, vol. 3181, Springer Berlin Heidelberg, 2004, pp. 169–178.
- [12] M. Delgado, N. Marín, D. Sánchez, M.A. Vila, Fuzzy association rules: general model and applications, *IEEE Trans. Fuzzy Syst.* 11 (2) (2003) 214–225.
- [13] M. Delgado, M.D. Ruiz, D. Sánchez, Studying interest measures for association rules through a logical model, *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* 18 (1) (2010) 87–106.
- [14] M. Delgado, M.D. Ruiz, D. Sánchez, New approaches for discovering exception and anomalous rules, *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* 19 (2) (2011) 361–399.
- [15] M. Delgado, M.D. Ruiz, D. Sánchez, J.M. Serrano, A formal model for mining fuzzy rules using the RL representation theory, *Inf. Sci.* 181 (23) (2011) 5194–5213.
- [16] D. Dubois, E. Hüllermeier, H. Prade, A systematic approach to the assessment of fuzzy association rules, *Data Mining Knowl. Discov.* 13 (2) (2006) 167–192.
- [17] A. Ebrahimzadeh, R. Sheibani, Two efficient algorithms for mining fuzzy association rules, *Int. J. Mach. Learn. Comput.* 1 (5) (2011) 510–515.
- [18] Y. Fu, J. Hah, Meta-rule-guided mining of association rules in relational databases, in: Proc. 1st Int'l Workshop on Integration of Knowledge Discovery with Deductive and Object-oriented Databases (KDOOD'95), 1995, pp. 39–46.
- [19] A. Gyenesi, A fuzzy approach for mining quantitative association rules, *Acta Cybern.* 15 (2) (2001).
- [20] E. Hüllermeier, Fuzzy association rules: semantic issues and quality measures, LNCS 2206 (2001) 380–391.
- [21] W. Jian, L. Xing ming, An efficient association rule mining algorithm in distributed databases, in: First Int. Workshop on Knowledge Discovery and Data Mining, WKDD 2008, 2008, pp. 108–113.
- [22] R. Jin, G. Agrawal, Systematic approach for optimizing complex mining tasks on multiple databases, in: Proc. of the 22nd Int. Conference on Data Engineering (ICDE 2006), 2006, p. 17.
- [23] R. Jin, G. Yang, G. Agrawal, Shared memory parallelization of data mining algorithms: techniques, programming interface and performance, *IEEE Trans. Knowl. Data Eng.* 17 (1) (2005) 71–89.
- [24] K. Kaneiwa, A rough set approach to multiple dataset analysis, *Appl. Soft Comput.* 11 (2) (2011) 2538–2547.
- [25] C.M. Kuok, A.W. Fu, M.H. Wong, Mining fuzzy association rules in databases, *SIGMOD Rec.* 27 (1) (1998) 41–46.
- [26] J.H. Lee, H.L. Kwang, An extension of association rules using fuzzy sets, in: IFSA'97, Prague, Czech Republic, 1997.
- [27] Ashish Mangalampalli, Vikram Pudi, FAR-miner: a fast and efficient algorithm for fuzzy association rule mining, *Int. J. Bus. Intell. Data Min.* 7 (4) (2012) 288–317.
- [28] N. Marin, C. Molina, J.M. Serrano, M.A. Vila, A complexity guided algorithm for association rule extraction on fuzzy datacubes, *IEEE Trans. Fuzzy Syst.* 16 (3) (2008) 693–714.
- [29] A. Nanopoulos, N. Apostolos, Y. Manolopoulos, Mining association rules in very large clustered domains, *Inf. Syst.* 32 (2007) 649–669.
- [30] M. Plasse, N. Nianga, G. Saporta, A. Villeminot, L. Leblond, Combined use of association rules mining and clustering methods to find relevant links between binary rare attributes in a large data set, *Comput. Stat. Data Anal.* 52 (1) (2007) 596–613.
- [31] T.T. Quan, L.N. Ngo, S.C. Hui, An effective clustering-based approach for conceptual association rules mining, in: International Conference on Computing and Communication Technologies, 2009. RIVF '09, 2009, pp. 1–7.
- [32] J.F. Roddick, M. Spiliopoulou, D. Lister, A. Ceglar, Higher order mining, *SIGKDD Explor.* 10 (1) (2008) 5–17.
- [33] M.D. Ruiz, J. Gómez-Romero, M.J. Martín Bautista, M.A. Vila, M. Delgado, Fuzzy meta-association rules, in: 2015 Conf. of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology (IFSAC-EUSFLAT-15), Gijón, Spain, June 30, 2015, 2015, pp. 247–254.
- [34] M.D. Ruiz, J. Gómez-Romero, M.J. Martín-Bautista, D. Sánchez, Meta-association rules for fusing regular association rules from different databases, in: Proceedings of the 17th International Conference on Information Fusion (FUSION'2014), 2014.
- [35] D. Sánchez, M. Delgado, M.A. Vila, J. Chamorro-Martínez, On a non-nested level-based representation of fuzziness, *Fuzzy Sets Syst.* 192 (2012) 159–175.
- [36] N. Saxena, R. Arora, R. Sikarwar, A. Gupta, An efficient approach of association rule mining on distributed database algorithm, *Int. J. Inf. Comput. Technol.* 3 (4) (2013) 225–234.
- [37] E. Shortliffe, B. Buchanan, A model of inexact reasoning in medicine, *Math. Biosci.* 23 (1975) 351–379.
- [38] T. Sudkamp, Examples, counterexamples, and measuring fuzzy associations, *Fuzzy Sets Syst.* 149 (1) (2005) 57–71.
- [39] P. Tan, V. Kumar, J. Srivastava, Selecting the right interestingness measure for association patterns, in: SIGKDD'02, Alberta, Canada, 2002.
- [40] M. Usman, R. Pears, A.C.M. Fong, Discovering diverse association rules from multidimensional schema, *Expert Syst. Appl.* 40 (15) (2013) 5975–5996.
- [41] C.-H. Wang, S.-H. Liu, C.-T. Pang, Mining association rules uses fuzzy weighted fp-growth, in: Soft Computing and Intelligent Systems (SCIS) and 13th Int. Symposium on Advanced Intelligent Systems (ISIS), 2012 Joint 6th Int. Conference on, 2012, pp. 983–988.
- [42] J. Wang, *Data Warehousing and Mining: Concepts, Methodologies, Tools, and Applications*, IGI Global, 2008.
- [43] X. Zhu, X. Wu, Discovering relational patterns across multiple databases, in: Proc. of the 23rd Int. Conference on Data Engineering (ICDE 2006), 2007, pp. 726–735.