Minimum-Size Bases of Association Rules

José L. Balcázar

Departament de Llenguatges i Sistemes Informàtics
Laboratori d'Algorísmica Relacional, Complexitat i Aprenentatge
Universitat Politècnica de Catalunya, Barcelona
`balqui@lsi.upc.edu`

Abstract. We focus on confidence-bounded association rules; we model a rather practical situation in which the confidence threshold is fixed by the user, as usually happens in applications. Within this model, we study notions of redundancy among association rules from a fundamental perspective: we discuss several existing alternative definitions and provide new characterizations and relationships between them. We show that these alternatives correspond actually to just two variants, which differ in the special treatment of full-confidence implications. For each of these two notions of redundancy, we show how to construct complete bases of absolutely minimum size.

Keywords: association rules, redundancy, optimum bases.

1 Motivation and Related Work

Few, if any, data mining tasks have the relative importance within that field of research as association rule mining. Whereas practitioners provide some success stories in various fields, researchers have provided a wealth of algorithmic ideas related to the task. Since the publication of the first proposal of confidence- and support-bound-based association mining [2], many algorithms have been designed. The interesting FIMI competition tested a wide family of these algorithms (`http://fimi.cs.helsinki.fi`). Currently, the amount of knowledge related to association rules has grown to the extent that the tasks of creating complete surveys and websites that maintain pointers to literature on association rules become daunting (a recent survey is [8] but additional materials appear in `http://wwwai.wu-wien.ac.at/~hahsler/research/association_rules/`, for instance, at the time of writing); see also [3], [26], [32], [33], and the references and discussions in their introductory sections.

A close relative of the notion of association rule, namely, that of exact implication, that is, an association rule that holds in 100% of the cases, had been studied before in the research area of closure spaces, where a number of methods have been found to construct, for every binary dataset, sets of implications (often called "bases") that are complete in the sense that all other implications can be derived from them; some of these bases enjoy minimality properties depending on the notion of derivation at hand [10], [12], [27], [31], in fact, such implications can

be seen also as conjunctions of definite Horn clauses, and the closure under intersection that characterizes closures spaces corresponds to the fact, well-known in logic and knowledge representation, that Horn theories are exactly those closed under bitwise intersection of propositional models (see e.g. [17]). Thus, as a form of knowledge gathered from a dataset, implications have several advantages: explicit or implicit correspondence with Horn logic, therefore a tight parallel with functional dependencies and a simple and well-known calculus through the Armstrong axioms (explained below), whence a clear notion of redundancy.

However, the fact has been long acknowledged (e.g. already in [23]) that, often, it is inappropriate to search only for absolute implications in the analysis of real world datasets. There may be many reasons to consider interesting a co-occurrence pattern, even if the perceived implication does not hold in all the cases. Already in [23], partial rules are defined in relation to their so-called-there "precision", that is, the notion of intensity of implication now widely called "confidence": for a given rule $X \to Y$, the ratio of how often X and Y are seen together to how often X is seen. Many other alternative measures of intensity of implication exist, and several sources describe them (see [13], [14]); we keep our focus on confidence, which is among the most common ones, certainly the first one proposed, and has a natural interpretation for educated users because it corresponds to a lower bound to the observed conditional probability.

The first attempts at mining partial rules were also proposed in [23]; yet, the process of searching for implications or for partial rules was not used on really large datasets until the introduction of the support bound: a threshold on how often the itemsets under analysis appear in the dataset. The idea of restricting the exploration for association rules to frequent itemsets, with respect to a support threshold, gave rise to the most widely discussed and applied algorithm, Apriori [3], and to an intense research activity. Unfortunately, if the combinatorial properties of implications are already nontrivial to handle, those of partial rules are even harder. Already with full-confidence implications, the output of an association mining process often consists of large sets of rules, and a well-known difficulty in applied association rule mining lies in that, on large datasets, and for sensible settings of the confidence and support thresholds, huge amounts of association rules are often obtained, much beyond what any user of the data mining process may be expected to look at; and the difficulty of studying the formal properties of partial rules makes it very difficult to select in a principled, provably optimal way, a subset of the rules without losing information.

Therefore, besides the interesting progress in the topic of how to organize and query the rules discovered (see [21], [22], [28]), one research topic that has been worthy of attention is the identification of patterns that indicate redundancy of rules, and ways to avoid that redundancy [1], [9], [18], [19], [23], [26], [32] (see also section 6 of [8] and the references therein). A major problem, open since [23], would be to give a general method for constructing bases of minimum size: a basis for a given dataset would be a subset of the rules that hold in the dataset, that is complete, in the sense that it makes all the remaining rules redundant. Therefore, restricting ourselves to the basis does not incur loss of information.

But the very notion of completeness of a basis depends on the concrete ways specified to construct "redundant" rules out of the basis. Therefore, we discuss this point briefly now and propose one specific standpoint. Imagine that a standard association rule miner has been run on a given dataset, with user-specified thresholds of support and confidence — a situation that fully matches most cases of application; its output $\mathcal{R}$ is now available, in the form of a (probably large) set of rules, each labeled with its confidence, all of these above the threshold. We want to select a basis $\mathcal{B} \subseteq \mathcal{R}$, aiming at choosing it as small as possible and, simultaneously, making sure that we do not lose information in doing so. (In fact, we are after better algorithmics that obtain directly $\mathcal{B}$ instead of mining for the whole of $\mathcal{R}$ and postprocessing it, but, for the sake of the properties of the basis, the discussion is clearer if we assume $\mathcal{R}$ known.) Thus, we are to find a subset of rules $\mathcal{B} \subseteq \mathcal{R}$ such that all the rules in $\mathcal{R}$ become redundant; and, of course, the crux now is how to define formally "redundant".

For the case of exact implications, "redundancy" has several equivalent natural, robust logical formalizations, such as entailment among definite Horn clauses. Alternatively, it also corresponds to derivability under the so-called Armstrong axiom schemes [30]: Reflexivity ($X \to X$), Augmentation (if $X \to Z$ and $Y \to W$ then $XY \to ZW$) and Transitivity (if $X \to Y$ and $Y \to Z$ then $X \to Z$).

But, in our context of partial rules with a hard confidence threshold in place, Augmentation and Transitivity, and also other natural inference schemes, are not valid anymore: for instance, if most of the times X appears it comes with Z, but it only comes with Y when Z is not present, then the confidence of $X \to Z$ may be high whereas the confidence of $XY \to Z$ may be null; that is, Augmentation (with $W = \emptyset$ here) is not valid. Neither is Transitivity: knowing that $A \to B$ and $B \to C$ (or even $AB \to C$) hold with confidence γ does not inform us about whether $A \to C$ holds with confidence γ. Additionally, a rule with several items in the consequent is *not* equivalent to the conjunction of the Horn-style rules with the same antecedent and each item of the consequent separately, and, if we look only into rules with singletons as consequents, we are almost certain to lose information. Indeed, if the confidence of $X \to YZ$ is high, it means that Y and Z appear together in most of the transactions having X; whereas the fact that both Y and Z appear in fractions at least γ of the transactions having X does not inform us that they show up *together* at a similar ratio of these transactions. This is also a failing form of Augmentation, with $X = Y$ this time.

Thus, we lack characterizations of derivability, and are left with the task of identifying, little by little, specific cases of redundancy, working them out, and seeing whether they give us bases and with which properties. This task indeed has been performed, and with great results already, but there is some progress to achieve yet. Most notably for our work here, we find that [1], [19] [26], and [32] have all proposed interesting notions of redundancy and methods to construct nonredundant bases; some of them work (as does [23]) in a setting where all the partial rules, plus their confidences, are to be inferred from the basis; and still the size of their basis can be suboptimal. The basis of [1] and the representative rules of [19] are of

minimum size in a well-defined sense, but this fact is a contribution of the present paper. The "basic association rules" of [20] suffer a limitation that consequents must be singletons, which loses information as indicated above (the same limitation applies to the useful `apriori` implementation of Borgelt available on the web [6]). Along a different avenue, some works are set up in a context of "using all the information available": namely, combining the supports of some sets in various ways, one can determine, through short computations, the supports of many other sets and the confidence of many rules. The "nonderivable" itemsets and rules [7], [15], [25] and the "covering" scheme of [9] all refer to the possibility of deriving rules of confidence above the threshold from information about the supports of specific sets. This seems a very effective approach, employing information about actual supports, that depend on the dataset at hand. These works were inspiring, and crucial to our research; however, we wish to study here a non-comparable approach: as discussed above, in our setting we are after a notion of redundancy based only on little information, essentially as in [1] and [19]. Instead of considering a rule redundant when its confidence can be somehow inferred from others, we take the slightly different view that it is redundant when "the fact that its confidence is above γ can be inferred from others", where these other rules are known to have confidence above γ but the inference process does not use their actual confidence values. We believe that this approach will be a good complement to the existing works, and expect that it would be particularly useful in cases where the user of the data mining system is not familiar with inclusion-exclusion principles and similar facts used in the "nonderivable itemsets" approach: only the rules (or almost only the rules, as explained next) are brought to bear in the derivation of other rules. We will see that this is possible, and indeed can be achieved through definitions that are already in the literature [1], [18], of which we establish new, important properties; and we will push this approach provably to the limit.

Then, we follow the proposals of [26], [32], and several other works, pushing through beyond that limit by assuming that we are allowed one single additional bit per rule: we will handle separately, to a given extent, full-confidence implications from lower-than-1-confidence rules, in order to profit from their very different combinatorics. We discuss adequate notions of redundancy and completeness, prove new properties, and refine the existing basis constructions up to a point where we can prove again that we attain the limit of the redundancy notion. We close the paper with some empirical data regarding our proposals (the author is grateful to Bart Goethals and Christian Borgelt for making their respective implementations of `fpgrowth` and `apriori` so easily accessible) and a short section of conclusions. Due to the space limits, here we defer the proofs of the theorems and additional examples and discussions to an extended version available from the author.

2   Preliminaries

A dataset $\mathcal{D}$ is given; it consists of transactions, each of which is an itemset labeled by a unique transaction identifier. The identifiers allow for many

transactions sharing the same itemset. Upper-case, often subscripted letters from the end of the alphabet, like X_1 or Y_0, denote itemsets. Juxtaposition denotes union of itemsets, as in XY; and $Z \subset X$ denotes proper subsets. For a transaction t, we denote $t \models X$ the fact that X is a subset of the itemset corresponding to t.

From the given dataset we obtain a notion of support of an itemset: $s_{\mathcal{D}}(X)$ is the cardinality of the set of transactions that include it, $\{t \in \mathcal{D} \mid t \models X\}$, sometimes, abusing language, we also refer to that set of transactions itself as support. Whenever $\mathcal{D}$ is clear, we drop the subindex: $s(X)$.

We immediately obtain by standard means (see, for instance, [12] or [32]) a notion of closed itemsets, namely, those that cannot be enlarged while maintaining the same support. The function that maps each itemset to the smallest closed set that contains it is known to be monotonic, extensive, and idempotent, that is, it is a closure operator. This notion will be reviewed in more detail later on.

Association rules are pairs of itemsets, denoted as $X \rightarrow Y$ for itemsets X and Y. Intuitively, they express that Y occurs particularly often among the transactions in which X occurs. More precisely, each such rule has a confidence associated: the confidence $c_{\mathcal{D}}(X \rightarrow Y)$ of an association rule $X \rightarrow Y$ in a dataset $\mathcal{D}$ is $\frac{s(XY)}{s(X)}$, that is, the ratio by which transactions having X have also Y, or, again, the observed empirical approximation to a conditional probability of Y given X. As with support, often we drop the subindex $\mathcal{D}$. This view suggests a form of correlation that, in many applications, is interpreted implicitly as a form of causality (which, however, is not guaranteed in any formal way; see the interesting discussion in [11]).

We denote as $\mathcal{R}$ the set of all the rules of confidence at least γ for the given dataset $\mathcal{D}$, again we should label $\mathcal{R}$ with $\mathcal{D}$ and γ as subscripts but these will be always clear from the context. Always $\gamma > 0$. We resort to the convention that, if $s(X) = 0$ (which implies $s(XY) = 0$) we redefine the undefined confidence as 1, since the intuitive expression "all transactions having X do have also Y" becomes vacuously true. Additionally, note that $c_{\mathcal{D}}(X \rightarrow Y) = c_{\mathcal{D}}(X \rightarrow XY)$, and we will switch rather freely between right hand sides that include the left hand side and right hand sides that don't: when two rules have the same left hand side, and the union of left and right hand sides also coincide, we say that they are *equivalent by reflexivity*. Clearly their confidences will always coincide.

3   Redundancy Notions

We start our analysis from one of the notions of redundancy proposed in [1], and from a variation thereof, seemingly less restrictive.

Definition 1. *1. $X_0 \rightarrow Y_0$ is AY-redundant with respect to $X_1 \rightarrow Y_1$ if the confidence and support of the former are always larger than or equal to those of the latter, whatever the dataset [1].*
  2. $X_0 \rightarrow Y_0$ is plainly redundant with respect to $X_1 \rightarrow Y_1$ if the confidence of $X_0 \rightarrow Y_0$ is larger than or equal to the confidence of the latter, whatever the dataset.

Thus, plain redundancy is like AY-redundancy, but forgetting the condition regarding support. It turns out that the condition about confidence is already rather strong, due to the "whatever the dataset" clause, to the point that our first new result is that the simplified version is as powerful as the original one:

Theorem 1. *Consider any two rules $X_0 \to Y_0$ and $X_1 \to Y_1$ where $Y_0 \not\subseteq X_0$. Then $X_0 \to Y_0$ is AY-redundant with respect to $X_1 \to Y_1$ if and only if $X_0 \to Y_0$ is plainly redundant with respect to $X_1 \to Y_1$.*

This will allow us to concentrate on confidence bounds at the time of discussing complete bases, since support bounds will follow essentially from that result. The reference indicated [1] also provides two simpler definitions of redundancy:

Definition 2. *(From [1].)*

1. *If $Z_0 \neq \emptyset$, rule $X_0 Z_0 \to Y_0$ is simply redundant with respect to $X_0 \to Y_0 Z_0$;*
2. *if $X_1 \subseteq X_0$ and $X_0 Y_0 \subset X_1 Y_1$, rule $X_0 \to Y_0$ is strictly redundant with respect to $X_1 \to Y_1$.*

It is rather easy to check that moving attributes from the right hand side into the left hand side can only increase the confidence and leave equal the support: this fact corresponds to simple redundancy, and relates rules obtained from the same frequent set $X_0 Y_0 Z_0$. Strict redundancy focuses, instead, on rules extracted from two different (frequent) itemsets, say $X_0 Y_0$ where X_0 will be considered as antecedent, versus $X_1 Y_1$, where X_1 will be antecedent, and under the conditions that $X_1 \subseteq X_0$ and $X_0 Y_0 \subset X_1 Y_1$ (the case $X_0 Y_0 = X_1 Y_1$ is already covered by simple redundancy). Both simple and strict redundancies imply AY-redundancy; this is argued in [1], which discusses most of the results just in terms of these two simplified notions. Note that, in principle, there could possibly be many other ways of being redundant beyond simple and strict redundancies: we show next, however, that, in essence, this is not the case, as we can state the following new, far from obvious characterization:

Theorem 2. *Consider any two rules $X_0 \to Y_0$ and $X_1 \to Y_1$ where $Y_0 \not\subseteq X_0$. The following are equivalent:*

1. *$X_1 \subseteq X_0$ and $X_0 Y_0 \subseteq X_1 Y_1$;*
2. *rule $X_0 \to Y_0$ is either simply redundant or strictly redundant with respect to $X_1 \to Y_1$, or they are equivalent by reflexivity;*
3. *rule $X_0 \to Y_0$ is plainly redundant with respect to $X_1 \to Y_1$.*

The two inclusions in the first statement form a condition that characterizes exactly the cover operator of [18] (Property 3.3 there): hence that operator is also fully equivalent to plain redundancy.

3.1   Optimum-Size Basis for Plain Redundancy

The main property of a basis, namely completeness, or not losing information upon deletion of the remaining rules, corresponds now to the following formalization:

Definition 3. *Given the set of rules $\mathcal{R}$ that hold in a given dataset $\mathcal{D}$ at confidence at least γ, $\mathcal{B} \subseteq \mathcal{R}$ is a complete basis if every rule of $\mathcal{R}$ is plainly redundant with respect to some rule of $\mathcal{B}$.*

We describe now, briefly, the construction of a basis as proposed, almost simultaneously, in [1] and in [19] (called there "representative rules").

Definition 4. *Given itemsets Y and $X \subseteq Y$, X is a γ-antecedent for Y if $c(X \to Y) \geq \gamma$, that is, $s(Y) \geq \gamma s(X)$.*

This is the same as organizing all the rules of $\mathcal{R}$ according to the itemset resulting from union of antecedent and consequent. For each itemset Z, we will keep some rules $X \to Y$ with $XY = Z$, or equivalently, we will keep some antecedents X for Z. Keeping all γ-antecedents of all sets yields, essentially, the whole of $\mathcal{R}$. We will keep only a part of them, as few as possible, but losing no information.

Definition 5. *Given itemsets Y and $X \subseteq Y$, X is a valid γ-antecedent for Y if the following holds:*

1. *X is a γ-antecedent of Y,*
2. *X fulfills this preceding property* minimally*: no proper subset of X is a γ-antecedent of Y, and*
3. *X is not a minimal γ-antecedent of an itemset strictly larger than Y.*

Now, the basis we are studying consists of all rules $X \to Z - X$ for all itemsets Z and for all valid antecedents X of Z. We refer to this basis as $\mathcal{B}_0$. It is immediate to see that all the rules in the basis $\mathcal{B}_0$ actually hold with confidence at least γ since, for $X \subseteq Z$, $c(X \to Z - X) = c(X \to Z) \geq \gamma$ which is explicitly required for X to be an antecedent of Z. It is proved in [1] that this basis is irredundant with respect to simple and strict redundancies. Completeness can be stated as follows (see also [19]):

Proposition 1. *$\mathcal{B}_0$ is a complete basis: all the rules in $\mathcal{R}$ are plainly redundant with respect to $\mathcal{B}_0$.*

Hence $\mathcal{B}_0$ contains rules that hold and which imply all the rules that hold, that is, it is indeed a basis. Efficient algorithms to construct $\mathcal{B}_0$ are provided in [1] and [19], where small examples can be found as well; we have developed alternative algorithmics that we will describe elsewhere. Our contribution here, rather than algorithmic, is foundational: now we can state and prove the most interesting novel property of this basis. It was known [1] that it is irredundant with respect to simple or strict redundancy, that is, none of the rules in it is simply redundant, nor strictly redundant. Our characterization in Theorem 2 then applies, so that none of the rules in $\mathcal{B}_0$ is, in fact, plainly redundant. But this irredundancy does not rule out the possibility that some other basis, constructed in an altogether different form, could have less many rules. We state and prove now that this is not so: there is absolutely no other way of constructing a basis smaller than this one, while preserving completeness with respect to plain redundancy, because it

has absolutely minimum size among all complete bases. Therefore, in order to find smaller ways of listing association rules from $\mathcal{R}$, and not losing information, a notion of redundancy more powerful than plain redundancy is unavoidably necessary.

Theorem 3. *Let $\mathcal{B}' \subseteq \mathcal{R}$ be any set of rules such that all the rules of $\mathcal{R}$ are plainly redundant with respect to $\mathcal{B}'$. Then, $\mathcal{B}'$ has at least as many rules as $\mathcal{B}_0$.*

Implementation and test of this approach reveals interesting reductions of the sizes of the rules, but these are still somewhat large. We concentrate efforts henceforth on the study of alternative existing notions of redundancy, as an attempt at getting bases smaller than $\mathcal{B}_0$.

4   Closure-Based Redundancy

The theorems in the previous section tell us that, for plain redundancy, the absolute limit of a basis, thus without losing information, is reached by the construction of [1]. Several studies, prominently [32], have put forward a different notion of redundancy; namely, they give a separate role to the full-confidence implications. Along this way, one gets a stronger notion of redundancy and, therefore, smaller bases can be constructed.

Indeed, implications can be summarized better, because they allow for transitivity and augmentation to apply in order to find redundancies; moreover, they can be combined in a certain form of transitivity with a partial rule of confidence at least γ to give rules that also have confidence at least γ. The best way to handle them is through a closure operator ([10], [12], [32], [26], [31]).

Given a dataset $\mathcal{D}$, the closure operator associated to $\mathcal{D}$ maps each itemset X to the largest itemset $\overline{X}$ that has the same support as X in $\mathcal{D}$; it can be defined in several alternative ways. A set is closed if it coincides with its closure. When $\overline{X} = Y$ we also say that X is a generator of Y. Our definition gives directly that always $s(X) = s(\overline{X})$. We will make liberal use of this fact, which is easy to check also with other definitions of the closure operator, as stated in [26], [32], and others.

Implications, or association rules of confidence 1, are closely related to this closure operator: $c(X \rightarrow Y) = 1$ if and only if $Y \subseteq \overline{X}$. Several quite good algorithms exist to find the closed sets and their supports. There are proposals of basis constructions from closed sets in the literature. In the min-max basis of [26], antecedents are minimal generators, that is, as small as possible, whereas consequents are closures, that is, as large as possible. We will follow the approach of [32] where both antecedents and consequents are chosen as small as possible; but, if we consider only number of rules, and not their sizes, both approaches share many similarities.

Redundancy based on closures is a natural generalization of equivalence by reflexivity; it works as follows ([32], see also section 4 in [26]): given a dataset and a closure operator corresponding to implications that have confidence 1 in the dataset, two partial rules $X_0 \rightarrow Y_0$ and $X_1 \rightarrow Y_1$ such that $\overline{X_0} = \overline{X_1}$ and

$\overline{X_0 Y_0} = \overline{X_1 Y_1}$ turn out to be equivalent in terms of support and confidence; the reason is that $s(X_0) = s(\overline{X_0}) = s(\overline{X_1}) = s(X_1)$, and $s(X_0 Y_0) = s(\overline{X_0 Y_0}) = s(\overline{X_1 Y_1}) = s(X_1 Y_1)$. Therefore, groups of rules sharing the same closure of the antecedent, and the same closure of the union of antecedent and consequent, give cases of redundancy. The notion of redundancy in [32] leads to selecting as irredundant rules from each such group that have inclusion-minimal antecedents and consequents. The size of the basis obtained in this way is analyzed empirically in [32], where it is also combined with the strategy from [23] of using only neighbor closures. This basis was found to be smaller than the set of all the rules in all cases, many times exhibiting a huge reduction factor. We will provide additional improvements by refining the closure analysis and by combining the idea of closure-based redundancy with the notion of valid antecedents of the previous section. Most interestingly, we provide again a proof that, with our variant, we reach the limit of closure-based redundancy: our basis will be shown again to have the smallest possible size with respect to closure-based completeness.

4.1   Characterizing Closure-Based Redundancy

Let $\mathcal{B}$ be the set of implications, of confidence 1, in the dataset $\mathcal{D}$; alternatively, $\mathcal{B}$ can be any of the bases already known for implications in a dataset. In our experiments later on we will use as $\mathcal{B}$ the Guigues-Duquenne basis, that has been proved to be of minimum size [10], [31]. From here on, we require $0 < \gamma < 1$, leaving the rules of confidence 1 to be handled from $\mathcal{B}$.

Definition 6. *Let $\mathcal{B}$ be a set of implications. Rule $X_2 \to Y_2$ has closure-based redundancy relative to $\mathcal{B}$ with respect to rule $X_1 \to Y_1$, which we denote by $\mathcal{B} \cup \{X_1 \to Y_1\} \models X_2 \to Y_2$, if any dataset $\mathcal{D}$ in which all the rules in $\mathcal{B}$ hold with confidence 1 gives $c_{\mathcal{D}}(X_2 \to Y_2) \geq c_{\mathcal{D}}(X_1 \to Y_1)$.*

We continue our study by showing a necessary and sufficient condition for closure-based redundancy, along the same lines as the one in the previous section.

Theorem 4. *Let $\mathcal{B}$ be a set of exact rules, with associated closure operator mapping each itemset Z to its closure $\overline{Z}$. Let $X_2 \to Y_2$ be a rule not implied by $\mathcal{B}$, that is, where $Y_2 \not\subseteq \overline{X_2}$. Then, the following are equivalent:*

1. *$X_1 \subseteq \overline{X_2}$ and $X_2 Y_2 \subseteq \overline{X_1 Y_1}$*
2. *$\mathcal{B} \cup \{X_1 \to Y_1\} \models X_2 \to Y_2$*

4.2   Optimum-Size Basis for Closure-Based Redundancy

In a similar way as in the previous section, we give here a basis, similar to the one proposed in [32] but smaller, by combining closure-based redundancy with the conditions of Definition 5. As in [32], the rules of confidence 1 are handled separately: we focus on the partial rules. We show first that our construction indeed gives a basis, that is, consists of rules that hold and make redundant all other rules that hold, in the following sense:

Definition 7. *Given the set of rules $\mathcal{R}$ that hold in a given dataset $\mathcal{D}$ at confidence at least γ, and given in it the rules $\mathcal{B}$ that hold with confidence 1, closure-based completeness of a set of partial rules $\mathcal{B}' \subseteq \mathcal{R}$ holds if every partial rule of $\mathcal{R}$ has closure-based redundancy relative to $\mathcal{B}$ with respect to some rule of $\mathcal{B}'$.*

Conceptually, our new basis departs only slightly from the bases of [26] and [32], but is nonetheless different in most cases (and therefore smaller, in some cases by an important factor, as shown below). It is constructed as follows. For each closed set X, we will consider a number of closed sets Y properly included in X to act as antecedents. They follow a similar pattern to the one of valid antecedents; but, instead of minimal antecedents, we will pick just minimal closed antecedents. That is:

Definition 8. *For each closed set X, a closed set $Y \subset X$ (proper inclusion) is a* basic γ-antecedent *if the following holds:*

1. *Y is a γ-antecedent of X: $s(X) \geq \gamma s(Y)$;*
2. *Y is* not *a γ-antecedent of X' for any larger closed set $X' \supset X$: $s(X') < \gamma s(Y)$;*
3. *Y is minimal among the closed proper subsets of X for which the previous two conditions hold.*

Then we use these antecedents for our basis:

Definition 9. *The basis $B_\gamma^\star$ consists of all the rules $Y \to X - Y$ for all closed sets X and all basic γ-antecedents Y of X.*

This set of rules entails exactly the rules that hold:

Theorem 5. *Let $\mathcal{B}$ be any basis for implications that hold with confidence 1.*

1. *All the rules in $B_\gamma^\star$ hold with confidence at least γ.*
2. *$B_\gamma^\star$ is a complete basis for closure-based redundancy: if the rule $Y \to Z$ holds with confidence at least γ, then, taken together with the full-confidence implications, $\mathcal{B} \cup B_\gamma^\star \models Y \to Z$.*

Now we can move to the main result of this section, and in fact of the whole paper: this basis is of absolutely minimum size.

Theorem 6. *Let $\mathcal{B}' \subseteq \mathcal{R}$ be an arbitrary basis having closure-based completeness for $\mathcal{R}$. Then, for each implication $Y \to Z \in B_\gamma^\star$, there is in $\mathcal{B}'$ an implication of the form $Y' \to Z'$ with $\overline{Y'Z'} = \overline{YZ}$ and $\overline{Y'} = Y$.*

That is, for each $Y \to X - Y \in B_\gamma^\star$, there is a corresponding rule in $Y' \to Z' \in \mathcal{B}'$; this rule in $\mathcal{B}'$ provides us with $X = \overline{Y'Z'}$ and $Y = \overline{Y'}$. Thus, both X and Y are univocally determined by $Y' \to Z'$ and, hence, the same rule in $\mathcal{B}'$ cannot yield but one of the rules in $B_\gamma^\star$, so that $\mathcal{B}'$ must have at least as many rules as $B_\gamma^\star$. Therefore, $B_\gamma^\star$ has a minimum number of rules, in an absolute sense, among all bases that are complete according to closure-based redundancy.

4.3 Support Bounds

Now we discuss, briefly, what happens if we work under a support threshold. In fact, for most datasets, if we do not impose a support bound then even the lattice including just closed sets reaches easily dozens or hundreds of thousands of nodes, or indeed even millions.

Assume that we do not mine the rule basis from a lattice including all closed sets but only those above a support threshold. Is there any risk of obtaining a wrong basis? The answer is negative:

Proposition 2. *For any fixed confidence threshold γ, mining the $B_\gamma^\star$ basis only on closed sets of support at least τ, for $\tau \leq \gamma$, provides a basis of the whole set of rules that hold with confidence at least γ and support at least τ.*

This proposition is proved by combining Theorem 1 with an easy observation: if the rule $X \rightarrow Y$ has support at least τ, both $\overline{X}$ and $\overline{XY}$ have also support at least τ, so that we can argue as in the proof of completeness above. We are therefore safe if we apply the basis construction for $B_\gamma^\star$ to a lattice of frequent closed sets above support τ, instead of the whole lattice of closed sets. However, this proposition does not ensure us that the basis obtained under a support bound is minimal anymore. There is a strategy (to be described in a forthcoming paper) that provides us with a correct and provably minimum basis also under a support bound, at the price of somewhat longer computations. For our development here, we just consider basis $B_\gamma^\star$ as computed from the lattice of closed sets above the support bound.

5 Empirical Validation

As indicated, this paper focuses rather on the foundational properties of redundancy and bases, and its algorithmic content is not the main contribution. However, we present some algorithmics and empirical results for the sake of completeness. We have implemented a known construction of a minimal basis for the full-confidence implications [10] to compute the closure operator, and an algorithm that constructs our proposed basis. For this computation, we consider an algorithm that conveniently uses as a black-box a separate closed itemsets miner. It is explained in Table 1: it scans the lattice of closed sets repeatedly to construct the basic γ-antecedents. That implementation has provided us with all the figures in Table 3. The initialization of the lists scan the whole lattice to pick up closed proper predecessors: a natural alternative would preprocess the lattice as a graph in order to find the predecessors of a node directly; however, in practice, with this alternative, whenever the graph requires too much space the computation slows down unacceptably, probably due to a worse fit to virtual memory caching. The search of the optimal algorithmic compromise between avoiding repeated computations while efficiently handling virtual memory will be the topic of further work; the current implementation gives answers in just seconds in most cases, on a mid-range Windows XP laptop, taking a few minutes when the closure space reaches a couple dozen thousand itemsets.

Table 1. Algorithmic approach to $B_\gamma^\star$

Algorithm $B_\gamma^\star$-1(closed sets,γ):
    for each of the closed sets:
        construct a list of closed proper subsets
        filter it to leave only γ-antecedents
        filter again to leave only minimal γ-antecedents
    for each of the closed sets:
        filter out from the list minimal γ-antecedents of larger closed sets
    for each of the closed sets:
        for each antecedent in its list:
            output as rule:
                left hand side: a minimum-size generator of the antecedent
                right hand side: a minimum-size generator of the closed set,
                    removing from it items appearing in the left hand side

Table 2. Parameters of the datasets

| Dataset	Num. Items	Size in Transactions
Chess	76	3196
Connect	130	67557
Mushroom	120	8124
Pumsb	7117	49046
Pumsb_star	7117	49046
T10I4D100K	1000	100000

An important property of our approach, shared with all the closure-based works, is that the key parameter is neither the size nor the dimensionality of the dataset, rather the size of the closures lattice. If that structure is of moderate size, our proposal works very well; the average degree of the corresponding Hasse graph is the next crucial value. If this degree is sublinear, which tends to be the case, then the computation of the rules is quadratic on the number of closures.

We have run our implementation exactly on the same real datasets (downloaded from `http://fimi.cs.helsinki.fi`) as the main table in [32], and with the same values of the parameters. Of course further comparisons are desirable, but in this way it is clear that we did not pick our experiments specifically to favor our approach. We have also included one of the synthetic datasets treated there. Some parameters of the datasets are indicated in Table 2.

Numbers of rules appear in Table 3. The values of "Supp/Conf" is the value of the confidence and support parameters; "Traditional" is the number of rules obtained via the original definition; "Closure-based" is the number of rules of the closure-based basis of [32] (which is already an impressive improvement); these columns are taken literally from [32] (there the value of the confidence parameter is made to coincide with the support parameter, so we do the same). These numbers are to be compared with the number of rules for our approach, that is, the sum of the columns "GDbasis" (number of rules in the Guigues-Duquenne minimum-size basis for full-confidence implications) and $B_\gamma^\star$. That

Table 3. Sizes of sets of rules for some datasets

Dataset	Supp/Conf	Traditional	Closure-based	GD basis	$B_\gamma^\star$ basis	Total
Chess	80	552564	27711	5	226	231
Chess	70	8171198	152074	10	891	901
Connect	97	8092	1116	4	41	45
Connect	90	3640704	18848	14	222	236
Mushroom	40	7020	475	24	41	65
Mushroom	20	19191656	5741	170	158	328
Pumsb	95	1170	267	2	32	34
Pumsb	85	1408950	44483	9	1080	1089
Pumsb_star	60	2358	192	5	6	11
Pumsb_star	40	5659536	13479	47	82	129
T10I4D100K	0.5	2216	1231	0	585	585
T10I4D100K	0.1	431838	86902	214	4054	4268
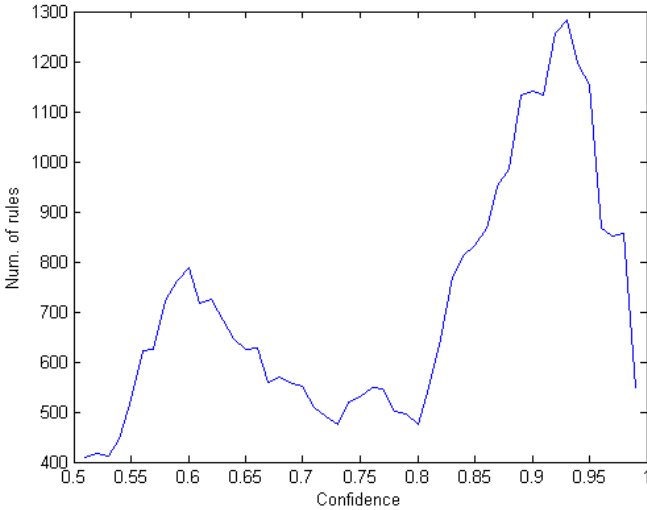
Fig. 1. Number of rules in the basis $B_\gamma^\star$ for `pumsb-star` at 20% support

total is provided in the last column. The improvements are apparent: however, they are just consequences of our main theorems.

One very interesting outcome of the experiments was the following. Some of us are used to a monotonicity intuition, by which lower confidence thresholds allow for more rules, so that the size of the output grows (sometimes enormously) as the confidence drops. However, in the case of the basis $B_\gamma^\star$, some datasets exhibit a nonmonotonic evolution: at lesser confidence thresholds, sometimes less many rules are mined. Inspecting the actual rules, we can find the reason: sometimes there are several rules at, say, 90% confidence that become simultaneously redundant due to a single rule of smaller confidence, say 85%, which does not appear at 90% confidence. This may reduce the set of rules upon lowering the

confidence threshold. An example illustrating this point is given by the dataset `pumsb-star`, mined for our basis $B_\gamma^\star$ at 20% support threshold with confidence ranging from 99% to 51%. The number of full-confidence implications in the Guigues-Duquenne basis at this support threshold is 47. The number of partial rules varies between 476 (at 80% confidence) and 1282 (at 93%), except near 50% confidence where the number of rules drops a bit more. The graphic in Figure 1 indicates the number of rules for each confidence level between 0.51 and 0.99, computed at a granularity of 0.01.

6   Conclusions

Our main contribution, at a glance, is a way of thinking about confidence-bounded association rules in terms of notions of redundancy. We have provided characterizations of existing redundancy notions, from which we have been able to exactly pinpoint the limitation of an existing basis proposal, for a plain notion of redundancy, and also to improve the constructions of bases for closure-based redundancy. As a consequence, analysis of specific datasets is now more feasible, from the perspective of the human analyst who must read through the output of the rule miner.

Several questions are worth further study. Mainly, it is not difficult to propose stronger notions of redundancy, and, in fact, we could see these contributions as progress towards a complete logical approach, where redundancy would play the role of entailment: in a forthcoming paper [4], will describe logical calculi that exactly correspond to plain redundancy and to closure-based redundancy. Then, the following natural question arises: our notions of redundancy only relate one partial rule to another partial rule, possibly in presence of full-confidence implications, and always with respect to a fixed confidence threshold: is it indeed possible that a partial rule is entailed by two partial rules, but not by a single one? The failures of Transitivity and Augmentation strongly suggest the intuition of a negative answer; in a forthcoming paper [4] we will see that this is so for certain confidence thresholds, but that there are confidence thresholds where this intuition is wrong: for instance, the reader may enjoy analyzing the case of rules $A \rightarrow BC$ and $A \rightarrow BD$, assuming that they hold with 65% confidence, and trying to make rule $ACD \rightarrow B$ fall below the same confidence threshold in the same dataset. This is, in fact, impossible, and our next paper [4] will characterize exactly when two partial rules entail a third one, either in presence or in absence of a closure operator for the exact rules. This could be a way towards stronger redundancy notions and even smaller bases.

We are studying as well ways of computing bases of provably minimum size under a support threshold; as discussed above, our basis is correct, and small, but, for this particular case, full optimality is not guaranteed yet. We will discuss an alternative in future works. Additional comparatives with other approaches to redundancy, based on additional information beyond the list of rules mined, is also necessary, including, for instance, the nonderivable itemsets of [7] or the cover operator of [9].

We plan also to extend this approach to the mining of more complex dependencies [29] or of dependencies among structured objects; however, extending the development to sequences, partial orders, and trees, is not fully trivial, because, as demonstrated in [5], the combinatorial structures may make redundant rules that would not be redundant in a propositional (item-based) framework; additionally, an intriguing question is: what part of all this discussion remains true if implication intensity measures different from confidence are used?

References

1. Aggarwal, C.C., Yu, P.S.: A new approach to online generation of association rules. IEEE Transactions on Knowledge and Data Engineering 13, 527–540 (2001)
2. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in very large databases. In: ACM SIGMOD, pp. 207–216 (1993)
3. Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, A.I.: Fast discovery of association rules. In: Fayyad, U., et al. (eds.) Advances in Knowledge Discovery and Data Mining, pp. 307–328. AAAI Press, Menlo Park
4. Balcázar, J.L.: Deduction Schemes for Association Rules. Discovery Science (to appear, 2008)
5. Balcázar, J.L., Bifet, A., Lozano, A.: Mining implications from lattices of closed trees. Extraction et Gestion des Connaissances (2008)
6. Borgelt, C.: Efficient Implementations of Apriori and Eclat Workshop on Frequent Itemset Mining Implementations (2003)
7. Calders, T., Goethals, B.: Mining all non-derivable frequent itemsets. In: Elomaa, T., Mannila, H., Toivonen, H. (eds.) PKDD 2002. LNCS (LNAI), vol. 2431, pp. 74–85. Springer, Heidelberg (2002)
8. Ceglar, A., Roddick, J.F.: Association mining. ACM Computing Surveys 38 (2006)
9. Cristofor, L., Simovici, D.: Generating an informative cover for association rules. In: ICDM 2002, pp. 597–613 (2002)
10. Guigues, J.-L., Duquenne, V.: Famille minimale d'implications informatives résultant d'un tableau de données binaires. Mathématiques et Sciences Humaines 24, 5–18 (1986)
11. Freitas, A.: Understanding the crucial differences between classification and discovery of association rules. SIGKDD Explorations 2, 65–69 (2000)
12. Ganter, B., Wille, R.: Formal Concept Analysis. Springer, Heidelberg (1999)
13. Garriga, G.C.: Statistical Strategies for Pruning All the Uninteresting Association Rules. In: ECAI 2004, pp. 430–434 (2004)
14. Geng, L., Hamilton, H.J.: Interestingness measures for data mining: a survey. ACM Computing Surveys 38 (2006)
15. Goethals, B., Muhonen, J., Toivonen, H.: Mining non-derivable association rules. In: Jonker, W., Petković, M. (eds.) SDM 2005. LNCS, vol. 3674. Springer, Heidelberg (2005)
16. Gunopulos, D., Khardon, R., Mannila, H., Saluja, S., Toivonen, H., Sharma, R.S.: Discovering all most specific sentences. ACM Trans. Database Syst. 28, 140–174 (2003)
17. Khardon, R., Roth, D.: Reasoning with models. Artificial Intelligence 87, 187–213 (1996)
18. Kryszkiewicz, M.: Representative Association Rules. In: Wu, X., Kotagiri, R., Korb, K.B. (eds.) PAKDD 1998. LNCS, vol. 1394, pp. 198–209. Springer, Heidelberg (1998)

19. Kryszkiewicz, M.: Representative Association Rules and Minimum Condition Maximum Consequence Association Rules. In: Żytkow, J.M. (ed.) PKDD 1998. LNCS, vol. 1510, pp. 361–369. Springer, Heidelberg (1998)
20. Li, G., Hamilton, H.: In: Jonker, W., Petković, M. (eds.) SDM 2004. LNCS, vol. 3178. Springer, Heidelberg (2004)
21. Liu, B., Hsu, W., Ma, Y.: Pruning and summarizing the discovered associations. In: KDD 1999, pp. 125–134 (1999)
22. Liu, B., Hu, M., Hsu, W.: Multi-level organization and summarization of the discovered rules. In: KDD 2000 (2000)
23. Luxenburger, M.: Implications partielles dans un contexte. Mathématiques et Sciences Humaines 29, 35–55 (1991)
24. Mannila, H., Toivonen, H.: Levelwise search and borders of theories in knowledge discovery. Data Mining and Knowledge Discovery 1, 241–258 (1997)
25. Muhonen, J., Toivonen, H.: Closed non-derivable itemsets. In: Fürnkranz, J., Scheffer, T., Spiliopoulou, M. (eds.) PKDD 2006. LNCS (LNAI), vol. 4213, pp. 601–608. Springer, Heidelberg (2006)
26. Pasquier, N., Taouil, R., Bastide, Y., Stumme, G., Lakhal, L.: Generating a condensed representation for association rules. Journal of Intelligent Information Systems 24, 29–60 (2005)
27. Pfaltz, J.L., Taylor, C.M.: Closed set mining of biological data. In: BIOKDD 2002, pp. 43–48 (2002)
28. Tuzhilin, A., Liu, B.: Querying multiple sets of discovered rules. In: KDD 2002, pp. 52–60 (2002)
29. Simovici, D.A., Cristofor, D., Cristofor, L.: Mining purity dependencies in databases. In: Extraction et Gestion des Connaissances EGC 2002, pp. 257–268 (2002)
30. Ullman, J., Widom, J.: A First Course in Database Systems (1997)
31. Wild, M.: A theory of finite closure spaces based on implications. Advances in Mathematics 108, 118–139 (1994)
32. Zaki, M.: Mining non-redundant association rules. Data Mining and Knowledge Discovery 9, 223–248 (2004)
33. Zaki, M., Ogihara, M.: Theoretical foundations of association rules. In: DMKD Workshop on research issues in DMKD (1998)