

REDUNDANCY, DEDUCTION SCHEMES, AND MINIMUM-SIZE BASES FOR ASSOCIATION RULES

JOSÉ L. BALCÁZAR

Dep. de Matemáticas, Estadística y Computación, Universidad de Cantabria, Santander, Spain
e-mail address: joseluis.balcazar@unican.es

ABSTRACT. Association rules are among the most widely employed data analysis methods in the field of Data Mining. An association rule is a form of partial implication between two sets of binary variables. In the most common approach, association rules are parameterized by a lower bound on their confidence, which is the empirical conditional probability of their consequent given the antecedent, and/or by some other parameter bounds such as “support” or deviation from independence. We study here notions of redundancy among association rules from a fundamental perspective. We see each transaction in a dataset as an interpretation (or model) in the propositional logic sense, and consider existing notions of redundancy, that is, of logical entailment, among association rules, of the form “any dataset in which this first rule holds must obey also that second rule, therefore the second is redundant”. We discuss several existing alternative definitions of redundancy between association rules and provide new characterizations and relationships among them. We show that the main alternatives we discuss correspond actually to just two variants, which differ in the treatment of full-confidence implications. For each of these two notions of redundancy, we provide a sound and complete deduction calculus, and we show how to construct complete bases (that is, axiomatizations) of absolutely minimum size in terms of the number of rules. We explore finally an approach to redundancy with respect to several association rules, and fully characterize its simplest case of two partial premises.

1. INTRODUCTION

The relatively recent discipline of Data Mining involves a wide spectrum of techniques, inherited from different origins such as Statistics, Databases, or Machine Learning. Among them, Association Rule Mining is a prominent conceptual tool and, possibly, a cornerstone notion of the field, if there is one. Currently, the amount of available knowledge regarding association rules has grown to the extent that the tasks of creating complete surveys and websites that maintain pointers to related literature become daunting. A survey, with plenty

1998 ACM Subject Classification: I.2.3, H.2.8, I.2.4, G.2.3, F.4.1.

Key words and phrases: Data mining, association rules, implications, redundancy, deductive calculus, optimum bases.

This work is supported in part by project TIN2007-66523 (FORMALISM) of Programa Nacional de Investigación, Ministerio de Ciencia e Innovación (MICINN), Spain, and by the PASCAL2 Network of Excellence of the European Union.

of references, is [12], and additional materials are available in [25]; see also [2], [3], [18], [36], [44], [45], and the references and discussions in their introductory sections.

Given an agreed general set of “items”, association rules are defined with respect to a dataset that consists of “transactions”, each of which is, essentially, a set of items. Association rules are customarily written as $X \rightarrow Y$, for sets of items X and Y , and they hold in the given dataset with a specific “confidence” quantifying how often Y appears among the transactions in which X appears.

A close relative of the notion of association rule, namely, that of exact implication in the standard propositional logic framework, or, equivalently, association rule that holds in 100% of the cases, has been studied in several guises. Exact implications are equivalent to conjunctions of definite Horn clauses: the fact, well-known in logic and knowledge representation, that Horn theories are exactly those closed under bitwise intersection of propositional models leads to a strong connection with Closure Spaces, which are characterized by closure under intersection (see the discussions in [15] or [26]). Implications are also very closely related to functional dependencies in databases. Indeed, implications, as well as functional dependencies, enjoy analogous, clear, robust, hardly disputable notions of redundancy that can be defined equivalently both in semantic terms and through the same syntactic calculus. Specifically, for the semantic notion of entailment, an implication $X \rightarrow Y$ is entailed from a set of implications \mathcal{R} if every dataset in which all the implications of \mathcal{R} hold must also satisfy $X \rightarrow Y$; and, syntactically, it is known that this happens if and only if $X \rightarrow Y$ is derivable from \mathcal{R} via the Armstrong axiom schemes, namely, Reflexivity ($X \rightarrow Y$ for $Y \subseteq X$), Augmentation (if $X \rightarrow Y$ and $X' \rightarrow Y'$ then $XX' \rightarrow YY'$, where juxtaposition denotes union) and Transitivity (if $X \rightarrow Y$ and $Y \rightarrow Z$ then $X \rightarrow Z$).

Also, such studies have provided a number of ways to find implications (or functional dependencies) that hold in a given dataset, and to construct small subsets of a large set of implications, or of functional dependencies, from which the whole set can be derived; in Closure Spaces and in Data Mining these small sets are usually called “bases”, whereas in Dependency Theory they are called “covers”, and they are closely related to deep topics such as hypergraph theory. Associated natural notions of minimality (when no implication can be removed), minimum size, and canonicity of a cover or basis do exist; again it is inappropriate to try to give a complete set of references here, but see, for instance, [15], [16], [19], [23], [24], [26], [37], [43], [45], and the references therein.

However, the fact has been long acknowledged (e.g. already in [33]) that, often, it is inappropriate to search only for absolute implications in the analysis of real world datasets. Partial rules are defined in relation to their “confidence”: for a given rule $X \rightarrow Y$, 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 [20], [21]; we keep our focus on confidence because, besides being among the most common ones, it has a natural interpretation for educated users through its correspondence with the observed conditional probability.

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, called Apriori [3], and to an intense research activity. 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 and other parameters, huge amounts of association rules are often obtained. Therefore, besides the interesting progress in the topic of how to organize and query the rules discovered (see [31], [32], [42]),

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; and each proposed notion of redundancy opens up a major research problem, namely, to provide a general method for constructing bases of minimum size with respect to that notion of redundancy.

For partial rules, the Armstrong schemes are not valid anymore. Reflexivity does hold, but Transitivity takes a different form that affects the confidence of the rules: if the rule $A \rightarrow B$ (or $A \rightarrow AB$, which is equivalent) and the rule $B \rightarrow C$ both hold with confidence at least γ , we still know nothing about the confidence of $A \rightarrow C$; even the fact that both $A \rightarrow AB$ and $AB \rightarrow C$ hold with confidence at least γ only gives us a confidence lower bound of $\gamma^2 < \gamma$ for $A \rightarrow C$ (assuming $\gamma < 1$). Augmentation does not hold at all; indeed, enlarging the antecedent of a rule of confidence at least γ may give a rule with much smaller confidence, even zero: think of a case where 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 $X \rightarrow Z$ may be high whereas the confidence of $XY \rightarrow Z$ may be null. Similarly, if the confidence of $X \rightarrow YZ$ is high, it means that Y and Z appear together in most of the transactions having X , whence the confidences of $X \rightarrow Y$ and $X \rightarrow Z$ are also high; but, with respect to the converse, 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: only a ratio of $2\gamma - 1 < \gamma$ is guaranteed as a lower bound. In fact, if we look only for association rules with singletons as consequents (as in some of the analyses in [1], or in the “basic association rules” of [30], or even in the traditional approach to association rules [2] and the useful *apriori* implementation of Borgelt available on the web [8]) we are almost certain to lose information. As a consequence of these failures of the Armstrong schemes, the canonical and minimum-size cover construction methods available for implications or functional dependencies are not appropriate for partial association rules.

On the semantic side, a number of formalizations of the intuition of redundancy among association rules exist in the literature, often with proposals for defining irredundant bases (see [1], [13], [27], [33], [36], [38], [44], the survey [29], and section 6 of the survey [12]). All of these are weaker than the notion that we would consider natural by comparison with implications (of which we start the study in the last section of this paper). We observe here that one may wish to fulfill two different roles with a basis, and that both appear (somewhat mixed) in the literature: as a computer-supported data structure from which confidences and supports of rules are computed (a role for which we use the closures lattice instead) or, in our choice, as a means of providing the user with a smallish set of association rules for examination and, if convenient, posterior enumeration of the rules that follow from each rule in the basis. That is, we will not assume to have available, nor to wish to compute, exact values for the confidence, but only discern whether it stays above a certain user-defined threshold. We compute actual confidences out of the closure lattice only at the time of writing out rules for the user.

This paper focuses mainly on several such notions of redundancy, defined in a rather general way, by resorting to confidence and support inequalities: essentially, a rule is redundant with respect to another if it has at least the same confidence and support of the latter *for every dataset*. We also discuss variants of this proposal and other existing definitions given in set-theoretic terms. For the most basic notion of redundancy, we provide formal proofs of the so far unstated equivalence among several published proposals, including a

syntactic calculus and a formal proof of the fact, also previously unknown, that the existing basis known as the Essential Rules or the Representative Rules ([1], [27], [38]) is of absolutely minimum size.

It is natural to wish further progress in reducing the size of the basis. Our theorems indicate that, in order to reduce further the size without losing information, more powerful notions or redundancy must be deployed. We consider for this role the proposal of handling separately, to a given extent, full-confidence implications from lower-than-1-confidence rules, in order to profit from their very different combinatorics. This separation is present in many constructions of bases for association rules [33], [36], [44]. We discuss corresponding notions of redundancy and completeness, and prove new properties of these notions; we give a sound and complete deductive calculus for this redundancy; and we refine the existing basis constructions up to a point where we can prove again that we attain the limit of the redundancy notion.

Next, we discuss yet another potential for strengthening the notion of redundancy. So far, all the notions have just related one partial rule to another, possibly in the presence of full implications. Is it possible to combine two partial rules, of confidence at least γ , and still obtain a partial rule obeying that confidence level? Whereas the intuition is that these confidences will combine together to yield a confidence lower than γ , we prove that there is a specific case where a rule of confidence at least γ is nontrivially entailed by two of them. We fully characterize this case and obtain from the characterization yet another deduction scheme. We hope that further progress along the notion of a *set* of partial rules entailing a partial rule will be made along the coming years.

Preliminary versions of the results in sections 3.1, 4.2, 4.3, and 5 have been presented at Discovery Science 2008 [6]; preliminary versions of the remaining results (except those in section 4.5, which are newer and unpublished) have been presented at ECMLPKDD 2008 [5].

2. PRELIMINARIES

Our notation and terminology are quite standard in the Data Mining literature. All our developments take place in the presence of a “universe” set \mathcal{U} of atomic elements called *items*; their absence or presence in sets or items plays the same role as binary-valued attributes of a relational table. Subsets of \mathcal{U} are called *itemsets*. A dataset \mathcal{D} is assumed to be given; it consists of transactions, each of which is an itemset labeled by a unique transaction identifier. The identifiers allow us to distinguish among transactions even if they share 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, whereas $Z \subseteq X$ is used for the usual subset relationship with potential equality.

For a transaction t , we denote $t \models X$ the fact that X is a subset of the itemset corresponding to t , that is, the transaction satisfies the minterm corresponding to X in the propositional logic sense.

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 slightly, we also refer to that set of transactions itself as support. Whenever \mathcal{D} is clear, we drop the subindex: $s(X)$. Observe that $s(X) \geq s(Y)$ whenever $X \subseteq Y$; this is immediate from the definition. Note that many references resort to a normalized notion of support by dividing by the dataset size. We chose not to, but there is no essential issue

here. Often, research work in Data Mining assumes that a threshold on the support has been provided and that only sets whose support is above the threshold (then called “frequent”) are to be considered. We will require this additional constraint occasionally for the sake of discussing the applicability of our developments.

We immediately obtain by standard means (see, for instance, [19] or [44]) 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. Closed sets whose support is above the support threshold, if given, are usually termed closed frequent sets.

Association rules are pairs of itemsets, denoted as $X \rightarrow Y$ for itemsets X and Y . Intuitively, they suggest the fact 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 $\frac{s(XY)}{s(X)}$. As with support, often we drop the subindex \mathcal{D} . The support in \mathcal{D} of the association rule $X \rightarrow Y$ is $s_{\mathcal{D}}(X \rightarrow Y) = s_{\mathcal{D}}(XY)$.

We can switch rather freely between right-hand sides that include the left-hand side and right-hand sides that don't:

Definition 2.1. Rules $X_0 \rightarrow Y_0$ and $X_1 \rightarrow Y_1$ are *equivalent by reflexivity* if $X_0 = X_1$ and $X_0Y_0 = X_1Y_1$.

Clearly, $c_{\mathcal{D}}(X \rightarrow Y) = c_{\mathcal{D}}(X \rightarrow XY) = c_{\mathcal{D}}(X \rightarrow X'Y)$ and, likewise, $s_{\mathcal{D}}(X \rightarrow Y) = s_{\mathcal{D}}(X \rightarrow XY) = s_{\mathcal{D}}(X \rightarrow X'Y)$ for any $X' \subseteq X$; that is, the support and confidence of rules that are equivalent by reflexivity always coincide. A minor notational issue that we must point out is that, in some references, the left-hand side of a rule is required to be a subset of the right-hand side, as in [33] or [38], whereas many others require the left- and right-hand sides of an association rule to be disjoint, such as [29] or the original [2]. Both the rules whose left-hand side is a subset of the right-hand side, and the rules that have disjoint sides, may act as canonical representatives for the rules equivalent to them by reflexivity. We state explicitly one version of this immediate fact for later reference:

Proposition 2.2. *If rules $X_0 \rightarrow Y_0$ and $X_1 \rightarrow Y_1$ are equivalent by reflexivity, $X_0 \cap Y_0 = \emptyset$, and $X_1 \cap Y_1 = \emptyset$, then they are the same rule: $X_0 = X_1$ and $Y_0 = Y_1$.*

In general, we do allow, along our development, rules where the left-hand side, or a part of it, appears also at the right-hand side, because by doing so we will be able to simplify the mathematical arguments. We will assume here that, at the time of printing out the rules found, that is, for user-oriented output, the items in the left-hand side are removed from the right-hand side; accordingly, we write our rules sometimes as $X \rightarrow Y - X$ to recall this convention.

Also, many references require the right-hand side of an association rule to be nonempty, or even both sides. However, empty sets can be handled with no difficulty and do give meaningful, albeit uninteresting, rules. A partial rule $X \rightarrow \emptyset$ with an empty right-hand side is equivalent by reflexivity to $X \rightarrow X$, or to $X \rightarrow X'$ for any $X' \subseteq X$, and all of these rules have always confidence 1. A partial rule with empty left-hand side, as employed, for instance, in [29], actually gives the normalized support of the right-hand side as confidence value:

Fact 1. In a dataset \mathcal{D} of n transactions, $c(\emptyset \rightarrow Y) = s(Y)/n$.

Again, these sorts of rules could be omitted from user-oriented output, but considering them conceptually valid simplifies the mathematical development. We also resort to the convention that, if $s(X) = 0$ (which implies that $s(XY) = 0$ as well) we redefine the undefined confidence $c(X \rightarrow Y)$ as 1, since the intuitive expression “all transactions having X do have also Y ” becomes vacuously true. This convention is irrespective of whether $Y \neq \emptyset$.

Throughout the paper, “*implications*” are association rules of confidence 1, whereas “*partial rules*” are those having a confidence below 1. When the confidence could be 1 or could be less, we say simply “*rule*”.

3. REDUNDANCY NOTIONS

We start our analysis from one of the notions of redundancy defined formally in [1]. The notion is employed also, generally with no formal definition, in several papers on association rules, which subsequently formalize and study just some particular cases of redundancy (e.g. [27], [40]); thus, we have chosen to qualify this redundancy as “standard”. We propose also a small variation, seemingly less restrictive; we have not found that variant explicitly defined in the literature, but it is quite natural.

Definition 3.1.

- (1) [1] $X_0 \rightarrow Y_0$ has *standard redundancy* with respect to $X_1 \rightarrow Y_1$ if the confidence and support of $X_0 \rightarrow Y_0$ are larger than or equal to those of $X_1 \rightarrow Y_1$, in *all* datasets.
- (2) $X_0 \rightarrow Y_0$ has *plain redundancy* 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 $X_1 \rightarrow Y_1$, in *all* datasets.

Generally, we will be interested in applying these definitions only to rules $X_0 \rightarrow Y_0$ where $Y_0 \not\subseteq X_0$ since, otherwise, $c(X_0 \rightarrow Y_0) = 1$ for all datasets and the rule is trivially redundant. We state and prove separately, for later use, the following new technical claim:

Lemma 3.2. *Assume that rule $X_0 \rightarrow Y_0$ is plainly redundant with respect to rule $X_1 \rightarrow Y_1$, and that $Y_0 \not\subseteq X_0$. Then $X_0Y_0 \subseteq X_1Y_1$.*

Proof. Assume $X_0Y_0 \not\subseteq X_1Y_1$, to argue the contrapositive. Then, we can consider a dataset consisting of one transaction X_0 and, say, m transactions X_1Y_1 . No transaction includes X_0Y_0 , therefore $c(X_0 \rightarrow Y_0) = 0$; however, $c(X_1 \rightarrow Y_1)$ is either 1 or $m/(m+1)$, which can be pushed up as much as desired by simply increasing m . Then, plain redundancy does not hold, because it requires $c(X_0 \rightarrow Y_0) \geq c(X_1 \rightarrow Y_1)$ to hold for all datasets whereas, for this particular dataset, the inequality fails. \square

The first use of this lemma is to show that plain redundancy is not, actually, weaker than standard redundancy.

Theorem 3.3. *Consider any two rules $X_0 \rightarrow Y_0$ and $X_1 \rightarrow Y_1$ where $Y_0 \not\subseteq X_0$. Then $X_0 \rightarrow Y_0$ has standard redundancy with respect to $X_1 \rightarrow Y_1$ if and only if $X_0 \rightarrow Y_0$ has plain redundancy with respect to $X_1 \rightarrow Y_1$.*

Proof. Standard redundancy clearly implies plain redundancy by definition. Conversely, plain redundancy implies, first, $c(X_0 \rightarrow Y_0) \geq c(X_1 \rightarrow Y_1)$ by definition and, further, $X_0Y_0 \subseteq X_1Y_1$ by Lemma 3.2; this implies in turn $s(X_0 \rightarrow Y_0) = s(X_0Y_0) \geq s(X_1Y_1) = s(X_1 \rightarrow Y_1)$, for all datasets, and standard redundancy holds. \square

The reference [1] also provides two more direct definitions of redundancy:

Definition 3.4.

- (1) if $X_1 \subset X_0$ and $X_0Y_0 = X_1Y_1$, rule $X_0 \rightarrow Y_0$ is *simply redundant* with respect to $X_1 \rightarrow Y_1$.
- (2) if $X_1 \subseteq X_0$ and $X_0Y_0 \subset X_1Y_1$, rule $X_0 \rightarrow Y_0$ is *strictly redundant* with respect to $X_1 \rightarrow Y_1$.

Simple redundancy in [1] is explained as a potential connection between rules that come from the same frequent set, in our case $X_0Y_0 = X_1Y_1$. The formal definition is not identical to our rendering: in its original statement in [1], rule $XZ \rightarrow Y$ is *simply redundant* with respect to $X \rightarrow YZ$, provided that $Z \neq \emptyset$. The reason is that, in that reference, rules are always assumed to have disjoint sides, and then both formalizations are clearly equivalent. We do not impose disjointness, so that the natural formalization of their intuitive explanation is as we have just stated in Definition 3.4. The following is very easy to see (and is formally proved in [1]).

Fact 2. [1] Both simple and strict redundancies imply standard redundancy.

Note that, in principle, there could possibly be many other ways of being redundant beyond simple and strict redundancies: we show below, however, that, in essence, this is not the case. We can relate these notions also to the *cover operator* of [27]:

Definition 3.5. [27] Rule $X_1 \rightarrow Y_1$ covers rule $X_0 \rightarrow Y_0$ when $X_1 \subseteq X_0$ and $X_0Y_0 \subseteq X_1Y_1$.

Here, again, the original definition, according to which rule $X \rightarrow Y$ covers rule $XZ \rightarrow Y'$ if $Z \subseteq Y$ and $Y' \subseteq Y$ (plus some disjointness and nonemptiness conditions that we omit) is appropriate for the case of disjoint sides. The formalization we give is stated also in [27] as a property that characterizes covering. Both simple and strict redundancies become thus merged into a single definition. We observe as well that the same notion is also employed, without an explicit name, in [38].

Again, it should be clear that, in Definition 3.5, the covered rule is indeed plainly redundant: whatever the dataset, changing from $X_0 \rightarrow Y_0$ to $X_1 \rightarrow Y_1$ the confidence stays equal or increases since, in the quotient $\frac{s(XY)}{s(X)}$ that defines the confidence of a rule $X \rightarrow Y$, the numerator cannot decrease from $s(X_0Y_0)$ to $s(X_1Y_1)$, whereas the denominator cannot increase from $s(X_1)$ to $s(X_0)$. Also, the proposals in Definition 3.4 and 3.5 are clearly equivalent:

Fact 3. Rule $X_1 \rightarrow Y_1$ covers rule $X_0 \rightarrow Y_0$ if and only if rule $X_0 \rightarrow Y_0$ is either simply redundant or strictly redundant with respect to $X_1 \rightarrow Y_1$, or they are equivalent by reflexivity.

It turns out that all these notions are, in fact, fully equivalent to plain redundancy; indeed, the following converse statement is a main new contribution of this section:

Theorem 3.6. *Assume rule $X_0 \rightarrow Y_0$ is plainly redundant with respect to $X_1 \rightarrow Y_1$, where $Y_0 \not\subseteq X_0$. Then rule $X_1 \rightarrow Y_1$ covers rule $X_0 \rightarrow Y_0$.*

Proof. By Lemma 3.2, $X_0Y_0 \subseteq X_1Y_1$. To see the other inclusion, $X_1 \subseteq X_0$, assume to the contrary that $X_1 \not\subseteq X_0$. Then we can consider a dataset in which one transaction consists of X_1Y_1 and, say, m transactions consist of X_0 . Since $X_1 \not\subseteq X_0$, these m transactions do not count towards the supports of X_1 or X_1Y_1 , so that the confidence of $X_1 \rightarrow Y_1$ is 1;

also, X_0 is not adding to the support of X_0Y_0 since $Y_0 \not\subseteq X_0$. As $X_0Y_0 \subseteq X_1Y_1$, exactly one transaction includes X_0Y_0 , so that $c(X_0 \rightarrow Y_0) = 1/m$, which can be made as low as desired. This would contradict plain redundancy. Hence, plain redundancy implies the two inclusions in the definition of cover. \square

Combining the statements so far, we obtain the following characterization:

Corollary 3.7. *Consider any two rules $X_0 \rightarrow Y_0$ and $X_1 \rightarrow Y_1$ where $Y_0 \not\subseteq X_0$. The following are equivalent:*

- (1) $X_1 \subseteq X_0$ and $X_0Y_0 \subseteq X_1Y_1$ (that is, rule $X_1 \rightarrow Y_1$ covers rule $X_0 \rightarrow Y_0$);
- (2) rule $X_0 \rightarrow Y_0$ is either simply redundant or strictly redundant with respect to rule $X_1 \rightarrow Y_1$, or they are equivalent by reflexivity;
- (3) rule $X_0 \rightarrow Y_0$ is plainly redundant with respect to rule $X_1 \rightarrow Y_1$;
- (4) rule $X_0 \rightarrow Y_0$ is standard redundant with respect to rule $X_1 \rightarrow Y_1$.

Marginally, we note here an additional strength of the proofs given. One could consider attempts at weakening the notion of plain redundancy by allowing for a “margin” or “slack”, appropriately bounded, but whose value is independent of the dataset, upon comparing confidences. The slack could be additive or multiplicative: conditions such as $c_{\mathcal{D}}(X_0 \rightarrow Y_0) \geq c_{\mathcal{D}}(X_1 \rightarrow Y_1) - \delta$ or $c_{\mathcal{D}}(X_0 \rightarrow Y_0) \geq \delta c_{\mathcal{D}}(X_1 \rightarrow Y_1)$, for all \mathcal{D} and for δ independent of \mathcal{D} , could be considered. However, such approaches do not define different redundancy notions: they result in formulations actually equivalent to plain redundancy. This is due to the fact that the proofs in Lemma 3.2 and Theorem 3.6 show that the gap between the confidences of rules that do not exhibit redundancy can be made as large as desired within $(0, 1)$. Likewise, if we fix a confidence threshold $\gamma \in (0, 1)$ beforehand and use it to define redundancy as $c_{\mathcal{D}}(X_0 \rightarrow Y_0) \geq \gamma \Rightarrow c_{\mathcal{D}}(X_1 \rightarrow Y_1) \geq \gamma$ for all \mathcal{D} , again an equivalent notion is obtained, independently of the concrete value of γ ; whereas, for $\gamma = 1$, this is, instead, a characterization of Armstrong derivability.

3.1. Deduction Schemes for Plain Redundancy. From the characterization just given, we extract now a sound and complete deductive calculus. It consists of three inference schemes: right-hand Reduction (rR), where the consequent is diminished; right-hand Augmentation (rA), where the consequent is enlarged; and left-hand Augmentation (ℓA), where the antecedent is enlarged. As customary in logic calculi, our rendering of each rule means that, if the facts above the line are already derived, we can immediately derive the fact below the line.

$$\begin{aligned} (rR) \quad & \frac{X \rightarrow Y, \quad Z \subseteq Y}{X \rightarrow Z} \\ (rA) \quad & \frac{X \rightarrow Y}{X \rightarrow XY} \\ (\ell A) \quad & \frac{X \rightarrow YZ}{XY \rightarrow Z} \end{aligned}$$

We also allow always to state trivial rules:

$$(r\emptyset) \quad \frac{}{X \rightarrow \emptyset}$$

Clearly, scheme (ℓA) could be stated equivalently with $XY \rightarrow YZ$ below the line, by (rA) :

$$(\ell A') \quad \frac{X \rightarrow YZ}{XY \rightarrow YZ}$$

In fact, (ℓA) is exactly the simple redundancy from Definition 3.4 and, in the cases where $Y \subseteq X$, it provides a way of dealing with one direction of equivalence by reflexivity; the other direction is a simple combination of the other two schemes. The Reduction

Scheme (rR) allows us to “lose” information from the right-hand side; it corresponds to strict redundancy.

As further alternative options, it is easy to see that we could also join (rR) and (rA) into a single scheme:

$$(rA') \quad \frac{X \rightarrow Y, \quad Z \subseteq XY}{X \rightarrow Z}$$

but we consider that this option does not really simplify, rather obscures a bit, the proof of our Corollary 3.8 below. Also, we could allow as trivial rules $X \rightarrow Y$ whenever $Y \subseteq X$, which includes the case of $Y = \emptyset$; such rules also follow from the calculus given by combining $(r\emptyset)$ with (rA) and (rR) .

The following can be derived now from Corollary 3.7:

Corollary 3.8. *The calculus given is sound and complete for plain redundancy; that is, rule $X_0 \rightarrow Y_0$ is plainly redundant with respect to rule $X_1 \rightarrow Y_1$ if and only if $X_0 \rightarrow Y_0$ can be derived from $X_1 \rightarrow Y_1$ using the inference schemes (rR) , (rA) , and (ℓA) .*

Proof. Soundness, that is, all rules derived are plainly redundant, is simple to argue by checking that, in each of the inference schemes, the confidence of the rule below the line is greater than or equal to the confidence of the rule above the line: these facts are actually the known statements that each of equivalence by reflexivity, simple redundancy, and strict redundancy imply plain redundancy. Also, trivial rules with empty right-hand side always hold. To show completeness, assume that rule $X_0 \rightarrow Y_0$ is plainly redundant with respect to rule $X_1 \rightarrow Y_1$. If $Y_0 \subseteq X_0$, apply $(r\emptyset)$ and use (rA) to copy X_0 and, if necessary, (rR) to leave just Y_0 in the right-hand side. If $Y_0 \not\subseteq X_0$, by Corollary 3.7, we know that this implies that $X_1 \subseteq X_0$ and $X_0 Y_0 \subseteq X_1 Y_1$. Now, to infer $X_0 \rightarrow Y_0$ from $X_1 \rightarrow Y_1$, we chain up applications of our schemes as follows:

$$X_1 \rightarrow Y_1 \vdash_{(rA)} X_1 \rightarrow X_1 Y_1 \vdash_{(rR)} X_1 \rightarrow X_0 Y_0 \vdash_{(\ell A)} X_0 \rightarrow Y_0$$

where the second step makes use of the inclusion $X_0 Y_0 \subseteq X_1 Y_1$, and the last step makes use of the inclusion $X_1 \subseteq X_0$. Here, the standard derivation symbol \vdash denotes derivability by application of the scheme indicated as a subscript. \square

We note here that [38] proposes a simpler calculus that consists, essentially, of (ℓA) (called there “weak left augmentation”) and (rR) (called there “decomposition”). The point is that these two schemes are sufficient to prove completeness of the “representative basis” as given in that reference, due to the fact that, in that version, the rules of the representative basis include the left-hand side as part of the right-hand side; but such a calculus is incomplete with respect to plain redundancy because it offers no rule to move items from left to right.

3.2. Optimum-Size Basis for Plain Redundancy. A basis is a way of providing a shorter list of rules for a given dataset, with no loss of information, in the following sense:

Definition 3.9. Given a set of rules \mathcal{R} , $\mathcal{B} \subseteq \mathcal{R}$ is a *complete basis* if every rule of \mathcal{R} is plainly redundant with respect to some rule of \mathcal{B} .

Bases are analogous to covers in functional dependencies, and we aim at constructing bases with properties that correspond to minimum size and canonical covers. The solutions for functional dependencies, however, are not valid for partial rules due to the failure of the Armstrong schemes.

In all practical applications, \mathcal{R} is the set of all the rules “mined from” a given dataset \mathcal{D} at a confidence threshold $\gamma \in (0, 1]$. That is, the basis is a set of rules that hold with confidence at least γ in \mathcal{D} , and such that each rule holds with confidence at least γ in \mathcal{D} if and only if it is plainly redundant with respect to some rule of \mathcal{B} ; equivalently, the rules in \mathcal{R} can be inferred from \mathcal{B} through the corresponding deductive calculus. All along this paper, such a confidence threshold is denoted γ , and always $\gamma > 0$. We will employ two simple but useful definitions.

Definition 3.10. Fix a dataset \mathcal{D} . Given itemsets Y and $X \subseteq Y$, X is a γ -antecedent for Y if $c(X \rightarrow Y) \geq \gamma$, that is, $s(Y) \geq \gamma s(X)$.

Note that we allow $X = Y$, that is, the set itself as its own γ -antecedent; this is just to simplify the statement of the following rather immediate lemma:

Lemma 3.11. *If X is a γ -antecedent for Y and $X \subseteq Z \subseteq Y$, then X is a γ -antecedent for Z and Z is a γ -antecedent for Y .*

Proof. From $X \subseteq Z \subseteq Y$ we have $s(X) \geq s(Z) \geq s(Y)$, so that $s(Z) \geq s(Y) \geq \gamma s(X) \geq \gamma s(Z)$. The lemma follows. \square

We make up for proper antecedents as part of the next notion:

Definition 3.12. Fix a dataset \mathcal{D} . Given itemsets Y and $X \subset Y$ (*proper subset*), X is a *valid γ -antecedent* for Y if the following holds:

- (1) X is a γ -antecedent of Y ,
- (2) no proper subset of X is a γ -antecedent of Y , and
- (3) no proper superset of Y has X as a γ -antecedent.

The basis we will focus on now is constructed from each Y and each valid antecedent of Y ; we consider that this is the most clear way to define and study it, and we explain below why it is essentially identical to two existing, independent proposals.

Definition 3.13. Fix a dataset \mathcal{D} and a confidence threshold γ . The *representative rules* for \mathcal{D} at confidence γ are all the rules $X \rightarrow Y - X$ for all itemsets Y and for all valid γ -antecedents X of Y .

In the following, we will say “let $X \rightarrow Y - X$ be a representative rule” to mean “let Y be a set having valid γ -antecedents, and let X be one of them”; the parameter $\gamma > 0$ will always be clear from the context. Note that some sets Y may not have valid antecedents, and then they do not generate any representative rules.

By the conditions on valid antecedents in representative rules, the following relatively simple but crucial property holds; beyond the use of our Corollary 3.7, the argument follows closely that of related facts in [29]:

Proposition 3.14. *Let rule $X \rightarrow Y - X$ be among the representative rules for \mathcal{D} at confidence γ . Assume that it is plainly redundant with respect to rule $X' \rightarrow Y'$, also of confidence at least γ ; then, they are equivalent by reflexivity and, in case $X' \cap Y' = \emptyset$, they are the same rule.*

Proof. Let $X \rightarrow Y - X$ be a representative rule, so that $X \subset Y$ and X is a valid γ -antecedent of Y . By Corollary 3.7, $X' \rightarrow Y'$ must cover $X \rightarrow Y - X$: $X' \subseteq X \subseteq X(Y - X) = Y \subseteq X'Y'$. As $c(X' \rightarrow Y') \geq \gamma$, X' is a γ -antecedent of $X'Y'$. We first show that $Y = X'Y'$; assume $Y \subset X'Y'$, and apply Lemma 3.11 to $X' \subseteq X \subseteq Y \subset X'Y'$: X' is also a γ -antecedent of Y ,

and the minimality of valid γ -antecedent X gives us $X = X'$. X is, thus, a γ -antecedent of $X'Y'$ which properly includes Y , contradicting the third property of valid antecedents.

Hence, $Y = X'Y'$, so that X' is a γ -antecedent of $X'Y' = Y$; but again X is a minimal γ -antecedent of $X'Y' = Y$, so that necessarily $X = X'$, which, together with $X'Y' = Y = XY$, proves equivalence by reflexivity. Under the additional condition $X' \cap Y' = \emptyset$, both rules coincide as per Proposition 2.2. \square

It easily follows that our definition is equivalent to the definition given in [27], except for a support bound that we will explain later; indeed, we will show in Section 4.5 that all our results carry over when a support bound is additionally enforced.

Corollary 3.15. *Fix a dataset \mathcal{D} and a confidence threshold γ . Let $X \subset Y$. The following are equivalent:*

- (1) *Rule $X \rightarrow Y - X$ is among the representative rules for \mathcal{D} at confidence γ ;*
- (2) *[27] $c(X \rightarrow Y - X) \geq \gamma$ and there does not exist any other rule $X' \rightarrow Y'$ with $X' \cap Y' = \emptyset$, of confidence at least γ in \mathcal{D} , that covers $X \rightarrow Y - X$.*

Proof. Let rule $X \rightarrow Y - X$ be among the representative rules for \mathcal{D} at confidence γ , and let rule $X' \rightarrow Y'$ cover it, while being also of confidence at least γ and with $X' \cap Y' = \emptyset$. Then, by Corollary 3.7 $X' \rightarrow Y'$ makes $X \rightarrow Y - X$ plainly redundant, and by Proposition 3.14 they must coincide. To show the converse, we must see that $X \rightarrow Y - X$ is a representative rule under the conditions given. The fact that $c(X \rightarrow Y - X) \geq \gamma$ gives that X is a γ -antecedent of Y , and we must see its validity. Assume that a proper subset $X' \subset X$ is also a γ -antecedent of Y : then the rule $X' \rightarrow Y - X'$ would be a different rule of confidence at least γ covering $X \rightarrow Y - X$, which cannot be. Similarly, assume that X is a γ -antecedent of Y' where $Y \subset Y'$: then the rule $X \rightarrow Y' - X$ would be a different rule of confidence at least γ covering $X \rightarrow Y - X$, which cannot be either. \square

Similarly, and with the same proviso regarding support, our definition is equivalent to the “essential rules” of [1]. There, the set of minimal γ -antecedents of a given itemset is termed its “boundary”. The following statement is also easy to prove:

Corollary 3.16. *Fix a dataset \mathcal{D} and a confidence threshold γ . Let $X \subseteq Y$. The following are equivalent:*

- (1) *Rule $X \rightarrow Y - X$ is among the representative rules for \mathcal{D} at confidence γ ;*
- (2) *[1] X is in the boundary of Y but is not in the boundary of any proper superset of Y ; that is, X is a minimal γ -antecedent of Y but is not a minimal γ -antecedent of any itemset strictly containing Y .*

Proof. If $X \rightarrow Y - X$ is among the representative rules, X must be a minimal γ -antecedent of Y by the conditions of valid antecedents; also, X is not a γ -antecedent at all (and, thus, not a minimal γ -antecedent) of any Y' properly including Y . Conversely, assume that X is in the boundary of Y but is not in the boundary of any proper superset of Y ; first, X must be a minimal γ -antecedent of Y so that the first two conditions of valid γ -antecedents hold. Assume that $X \rightarrow Y - X$ is not among the representative rules; the third property must fail, and X must be a γ -antecedent of some Y' with $Y \subset Y'$. Our hypotheses tell us that X is not a *minimal* γ -antecedent of Y' . That is, there is a proper subset $X' \subset X$ that is also a γ -antecedent of Y' . It suffices to apply Lemma 3.11 to $X' \subset X \subseteq Y \subset Y'$ to reach a contradiction, since it implies that X' is a γ -antecedent of Y and therefore X would not be a minimal γ -antecedent of Y . \square

The representative rules are indeed a basis:

Fact 4. ([1], [27]) Fix a dataset \mathcal{D} and a confidence threshold γ , and consider the set of representative rules constructed from \mathcal{D} ; it is a complete basis:

- (1) all the representative rules hold with confidence at least γ ;
- (2) all the rules of confidence at least γ in \mathcal{D} are plainly redundant with respect to the representative rules.

The first part follows directly from the use of γ -antecedents as left-hand sides of representative rules. For the second part, also almost immediate, suppose $c(X \rightarrow Y) \geq \gamma$, and let $Z = XY$; since X is now a γ -antecedent of Z , it must contain a minimal γ -antecedent of Z , say $X' \subseteq X$. Let Z' be the largest superset of Z such that X' is still a γ -antecedent of Z' . Thus, $X' \rightarrow Z' - X'$ is among the representative rules and covers $X \rightarrow Y$. Small examples of the construction of representative rules can be found in the same references; we also provide one below.

An analogous fact is proved in [38] through an incomplete deductive calculus consisting of the schemes that we have called (lA) and (rR) , and states that every rule of confidence at least γ can be inferred from the representative rules by application of these two inference schemes. Since representative rules in the formulation of [38] have a right-hand side that includes the left-hand side, this inference process does not need to employ (rA) .

Now we can state and prove the most interesting novel property of this basis, which again follows from our main result in this section, Corollary 3.7. As indicated, representative rules were known to be irredundant with respect to simple and strict redundancy or, equivalently, with respect to covering. But, for standard redundancy, in principle there was actually the possibility that some other basis, constructed in an altogether different form, could have less rules. We can 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 bases, a notion of redundancy more powerful than plain (or standard) redundancy is unavoidably necessary.

Theorem 3.17. *Fix a dataset \mathcal{D} , and let \mathcal{R} be the set of rules that hold with confidence γ in \mathcal{D} . Let $\mathcal{B}' \subseteq \mathcal{R}$ be an arbitrary basis, complete so that all the rules in \mathcal{R} are plainly redundant with respect to \mathcal{B}' . Then, \mathcal{B}' must have at least as many rules as the representative rules. Moreover, if the rules in \mathcal{B}' are such that antecedents and consequents are disjoint, then all the representative rules belong to \mathcal{B}' .*

Proof. By the assumed completeness of \mathcal{B}' , each representative rule $X \rightarrow Y - X$ must be redundant with respect to some rule $X' \rightarrow Y' \in \mathcal{B}' \subseteq \mathcal{R}$. By Corollary 3.7, $X' \rightarrow Y'$ covers $X \rightarrow Y - X$. Then Proposition 3.14 applies: they are equivalent by reflexivity. This means $X = X'$ and $Y = X'Y'$, hence $X' \rightarrow Y'$ uniquely identifies which representative rule it covers, if any; hence, \mathcal{B}' needs, at least, as many rules as the number of representative rules. Moreover, as stated also in Proposition 3.14, if the disjointness condition $X' \cap Y' = \emptyset$ holds, then both rules coincide. \square

Example 3.18. We consider a small example consisting of 12 transactions, where there are actually only 7 itemsets, but some of them are repeated across several transactions. We can simplify our study as follows: if X is not a closed set for the dataset, that is, if it has some superset $X' \supset X$ with the same support, then clearly it has no valid γ -antecedents (see also Fact 5 below); thus we concentrate on closed sets. Figure 1 shows the example dataset and

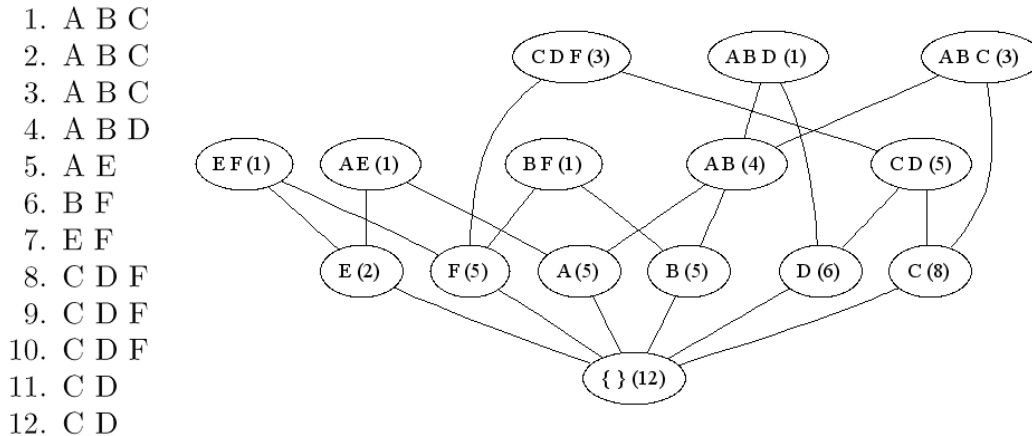


Figure 1: Closed itemsets for a small example

the corresponding (semi-)lattice of closures, depicted as a Hasse diagram (that is, transitive edges have been removed to clarify the drawing); edges stand for the inclusion relationship.

For this example, the implications can be summarized by six rules, namely, $AC \Rightarrow B$, $BC \Rightarrow A$, $AD \Rightarrow B$, $BD \Rightarrow A$, $CF \Rightarrow D$, and $DF \Rightarrow C$, which are also the representative rules at confidence 1. At confidence $\gamma = 0.75$, we find that, first, the left-hand sides of the six implications are still valid γ -antecedents even at this lower confidence, so that the implications still belong to the representative basis. Then, we see that two of the closures, ABC and CD , have additionally one valid γ -antecedent each, whereas AB has two. The following four rules hold: $A \rightarrow B$, $B \rightarrow A$, $AB \rightarrow C$, and $D \rightarrow C$. These four rules, jointly with the six implications indicated, constitute exactly the ten representative rules at confidence 0.75.

4. CLOSURE-BASED REDUNDANCY

Theorem 3.17 in the previous section tells us that, for plain redundancy, the absolute limit of a basis at any given confidence threshold is reached by the set of representative rules. Several studies, prominently [44], have put forward a different notion of redundancy; namely, they give a separate role to the full-confidence implications, often through their associated closure operator. Along this way, one gets a stronger notion of redundancy and, therefore, a possibility that 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 certain forms of transitivity with partial rules: as a simple example, if $c(X \rightarrow Y) \geq \gamma$ and $c(Y \rightarrow Z) = 1$, that is, if a fraction γ or more of the support of X has Y and *all* the transactions containing Y do have Z as well, clearly this implies that $c(X \rightarrow Z) \geq \gamma$. Observe, however, that the directionality is relevant: from $c(X \rightarrow Y) = 1$ and $c(Y \rightarrow Z) \geq \gamma$ we infer nothing about $c(X \rightarrow Z)$, since the high confidence of $Y \rightarrow Z$ might be due to a large number of transactions that do not include X .

We will need some notation about closures. Given a dataset \mathcal{D} , the closure operator associated to \mathcal{D} maps each itemset X to the largest itemset \bar{X} that contains X and has the

same support as X in \mathcal{D} : $s(X) = s(\overline{X})$, and \overline{X} is as large as possible under this condition. It is known and easy to prove that \overline{X} exists and is unique. Implications that hold in the dataset correspond to the closure operator ([19], [23], [36], [43], [44]): $c(X \rightarrow \overline{X}) = 1$, and \overline{X} is as large as possible under this condition. Equivalently, the closure of itemset X is the intersection of all the transactions that contain X ; this is because $X \subseteq \overline{X}$ implies that all transactions counted for the support of \overline{X} are counted as well for the support of X , hence, if the support counts coincide they must count exactly the same transactions.

Along this section, as in [36], we denote full-confidence implications using the standard logic notation $X_0 \Rightarrow Y_0$; thus, $X_0 \Rightarrow Y_0$ if and only if $Y_0 \subseteq \overline{X_0}$.

A basic fact from the theory of Closure Spaces is that closure operators are characterized by three properties: extensivity ($X \subseteq \overline{X}$), idempotency ($\overline{\overline{X}} = \overline{X}$), and monotonicity (if $X \subseteq Y$ then $\overline{X} \subseteq \overline{Y}$). As an example of the use of these properties, we note the following simple consequence for later use:

Lemma 4.1. $XY \subseteq \overline{XY} \subseteq \overline{X}\overline{Y} \subseteq \overline{XY}$, and $\overline{\overline{XY}} = \overline{\overline{X}\overline{Y}} = \overline{\overline{X}\overline{Y}} = \overline{XY} = \overline{XY}$.

We omit the immediate proof. A set is closed if it coincides with its closure. Usually we speak of the *lattice* of closed sets (technically it is just a semilattice but it allows for a standard transformation into a lattice [14]). When $\overline{X} = Y$ we also say that X is a generator of Y ; if the closures of all proper subsets of X are different from Y , we say that X is a minimal generator. Note that some references use the term “generator” to mean our “minimal generator”; we prefer to make explicit the minimality condition in the name. In some works, often database-inspired, minimal generators are termed sometimes “keys”. In other works, often matroid-inspired, they are termed also “free sets”. Our definition says explicitly that $s(X) = s(\overline{X})$. We will make liberal use of this fact, which is easy to check also with other existing alternative definitions of the closure operator, as stated in [36], [44], and others. Several quite good algorithms exist to find the closed sets and their supports (see section 4 of [12]).

Redundancy based on closures is a natural generalization of equivalence by reflexivity; it works as follows ([44], see also [29] and section 4 in [36]):

Lemma 4.2. *Given a dataset and the corresponding closure operator, two partial rules $X_0 \rightarrow Y_0$ and $X_1 \rightarrow Y_1$ such that $\overline{X_0} = \overline{X_1}$ and $\overline{X_0Y_0} = \overline{X_1Y_1}$ have the same support and the same confidence.*

The rather immediate reason is that $s(X_0) = s(\overline{X_0}) = s(\overline{X_1}) = s(X_1)$, and $s(X_0Y_0) = s(\overline{X_0Y_0}) = s(\overline{X_1Y_1}) = s(X_1Y_1)$. 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. On account of these properties, there are some proposals of basis constructions from closed sets in the literature, reviewed below. But the first fact that we must mention to relate the closure operator with our explanations so far is the following:

Fact 5. [28] Let $X \rightarrow Y - X$ be a representative rule as per Definition 3.13. Then Y is a closed set and X is a minimal generator.

The proof is direct from Definitions 3.12 and 3.13, and can be found in [28], [29], [38]. These references employ this property to improve on the earlier algorithms to compute the representative rules, which considered all the frequent sets, by restricting the exploration to closures and minimal generators. Also the authors of [40] do the same, seemingly unaware that the algorithm in [28] already works just with closed itemsets. Fact 5 may shed doubts

on whether closure-based redundancy actually can lead to smaller bases. We prove that this is sometimes the case, due to the fact that the redundancy notion itself changes, and allows for a form of Transitivity, which we show can take again the form of a deductive calculus. Then, we will be able to refine the notion of valid antecedent of the previous section and provide a basis for which we can prove that it has the smallest possible size among the bases for partial rules, with respect to closure-based completeness. That is, we will reach the limit of closure-based redundancy in the same manner as we did for standard redundancy in the previous section.

4.1. Characterizing Closure-Based Redundancy. Let \mathcal{B} be the set of implications in the dataset \mathcal{D} ; alternatively, \mathcal{B} can be any of the bases already known for implications in a dataset. In our empirical validations below we have used as \mathcal{B} the Guigues-Duquenne basis, or GD-basis, that has been proved to be of minimum size [23], [43]. An apparently popular and interesting alternative, that has been rediscovered over and over in different guises, is the so-called *iteration-free basis* of [43], which coincides with the proposal in [37] and with the *exact min-max basis* of [36] (also called sometimes *generic basis* [29]); because of Fact 5, it coincides exactly also with the representative rules of confidence 1, that is: implications that are not plainly redundant with any other implication according to Definition 3.1. Also, it coincides with the “closed-key basis” for frequent sets in [39], which in principle is not intended as a basis for rules, and has a different syntactic sugar, but differs in essence from the iteration-free basis only in the fact that the support of each rule is explicitly recorded together with it.

Closure-based redundancy takes into account \mathcal{B} as follows:

Definition 4.3. Let \mathcal{B} be a set of implications. Partial rule $X_0 \rightarrow Y_0$ has *closure-based redundancy relative to \mathcal{B}* with respect to rule $X_1 \rightarrow Y_1$, denoted $\mathcal{B}, \{X_1 \rightarrow Y_1\} \models X_0 \rightarrow Y_0$, if any dataset \mathcal{D} in which all the rules in \mathcal{B} hold with confidence 1 gives $c_{\mathcal{D}}(X_0 \rightarrow Y_0) \geq c_{\mathcal{D}}(X_1 \rightarrow Y_1)$.

In some cases, it might happen that the dataset at hand does not satisfy any nontrivial rule with confidence 1; then, this notion will not be able to go beyond plain redundancy. However, it is usual that some full-confidence rules do hold, and, in these cases, as we shall see, closure-based redundancy may give more economical bases. More generally, all our results only depend on the implications reaching indeed full confidence in the dataset; but they are not required to capture all of these: the implications in \mathcal{B} (with their consequences according to the Armstrong schemes) could constitute just a part of the full-confidence rules in the dataset. In particular, plain redundancy reappears by choosing $\mathcal{B} = \emptyset$, whether the dataset satisfies or not any full-confidence implication.

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 4.4. *Let \mathcal{B} be a set of exact rules, with associated closure operator mapping each itemset Z to its closure \overline{Z} . Let $X_0 \rightarrow Y_0$ be a rule not implied by \mathcal{B} , that is, where $Y_0 \not\subseteq \overline{X_0}$. Then, the following are equivalent:*

- (1) $X_1 \subseteq \overline{X_0}$ and $X_0 Y_0 \subseteq \overline{X_1 Y_1}$;
- (2) $\mathcal{B}, \{X_1 \rightarrow Y_1\} \models X_0 \rightarrow Y_0$.

Proof. The direct proof is simple: the inclusions given imply that $s(X_1) \geq s(\overline{X_0}) = s(X_0)$ and $s(X_0 Y_0) \geq s(\overline{X_1 Y_1}) = s(X_1 Y_1)$; then $c(X_0 \rightarrow Y_0) = \frac{s(X_0 Y_0)}{s(X_0)} \geq \frac{s(X_1 Y_1)}{s(X_1)} = c(X_1 \rightarrow Y_1)$.

Conversely, for $Y_0 \not\subseteq \overline{X_0}$, we argue that, if either of $X_1 \subseteq \overline{X_0}$ and $X_0Y_0 \subseteq \overline{X_1Y_1}$ fails, then there is a dataset where \mathcal{B} holds with confidence 1 and $X_1 \rightarrow Y_1$ holds with high confidence but the confidence of $X_0 \rightarrow Y_0$ is low.

We observe first that, in order to satisfy \mathcal{B} , it suffices to make sure that all the transactions in the dataset we are to construct are closed sets according to the closure operator corresponding to \mathcal{B} .

Assume now that $X_1 \not\subseteq \overline{X_0}$: then a dataset consisting only of one or more transactions with itemset $\overline{X_0}$ satisfies (vacuously) $X_1 \rightarrow Y_1$ with confidence 1 but, given that $Y_0 \not\subseteq \overline{X_0}$, leads to confidence zero for $X_0 \rightarrow Y_0$. It is also possible to argue without resorting to vacuous satisfaction: simply take one transaction consisting of $\overline{X_1Y_1}$ and, in case this transaction satisfies $X_0 \rightarrow Y_0$, obtain as low a confidence as desired for $X_0 \rightarrow Y_0$ by adding as many transactions $\overline{X_0}$ as necessary; these will not change the confidence of $X_1 \rightarrow Y_1$ since $X_1 \not\subseteq \overline{X_0}$.

Then consider the case where $X_1 \subseteq \overline{X_0}$, whence the other inclusion fails: $X_0Y_0 \not\subseteq \overline{X_1Y_1}$. Consider a dataset of, say, n transactions, where one transaction consists of the itemset $\overline{X_0}$ and $n - 1$ transactions consist of the itemset $\overline{X_1Y_1}$. The confidence of $X_1 \rightarrow Y_1$ is at least $\frac{n-1}{n}$, which can be made as close to 1 as desired by increasing n , whereas the presence of at least one X_0 and no transaction at all containing X_0Y_0 gives confidence zero to $X_0 \rightarrow Y_0$. Thus, in either case, we see that redundancy does not hold. \square

4.2. Deduction Schemes for Closure-Based Redundancy. We provide now a stronger calculus that is sound and complete for this more general case of closure-based redundancy. For clarity, we chose to avoid the closure operator in our deduction schemes, writing instead explicitly each implication.

Our calculus for closure-based redundancy consists of four inference schemes, each of which reaches a partial rule from premises including a partial rule. Two of the schemes correspond to variants of Augmentation, one for enlarging the antecedent, the other for enlarging the consequent. The other two correspond to composition with an implication, one in the antecedent and one in the consequent: a form of controlled transitivity. Their names (rA) , (ℓA) , (rI) , and (ℓI) indicate whether they operate at the right or left-hand side and whether their effect is Augmentation or composition with an Implication.

$$\begin{aligned}
 (rA) \quad & \frac{X \rightarrow Y, \quad X \Rightarrow Z}{X \rightarrow YZ} \\
 (rI) \quad & \frac{X \rightarrow Y, \quad Y \Rightarrow Z}{X \rightarrow Z} \\
 (\ell A) \quad & \frac{X \rightarrow YZ}{XY \rightarrow Z} \\
 (\ell I) \quad & \frac{X \rightarrow Y, \quad Z \subseteq X, \quad Z \Rightarrow X}{Z \rightarrow Y}
 \end{aligned}$$

Again we allow to state rules with empty right-hand side directly:

$$(r\emptyset) \quad \overline{X \rightarrow \emptyset}$$

Alternatively, we could state trivial rules with a subset of the left-hand side at the right-hand side. Note that this opens the door to using (rA) with an empty Y , and this allows us to “downgrade” an implication into the corresponding partial rule. Again, (ℓA) could be stated equivalently as $(\ell A')$ like in Section 3.1. In fact, the whole connection with the simpler calculus in Section 3.1 should be easy to understand: first, observe that the (ℓA) rules are identical. Now, if implications are not considered separately, the closure operator

trivializes to identity, $\overline{Z} = Z$ for every Z , and the only cases where we know that $X_1 \Rightarrow Y_1$ are those where $Y_1 \subseteq X_1$; we see that (rI) corresponds, in that case, to (rR) , whereas the (rA) schemes only differ on cases of equivalence by reflexivity. Finally, in that case (ℓI) becomes fully trivial since $Z \Rightarrow X$ becomes $X \subseteq Z$ and, together with $Z \subseteq X$, leads to $X = Z$: then, the partial rules above and below the line would coincide.

Similarly to the plain case, there exists an alternative deduction system, more compact, whose equivalence with our four schemes is rather easy to see. It consists of just two forms of combining a partial rule with an implication:

$$(rI') \quad \frac{X \rightarrow Y, \quad XY \Rightarrow Z}{X \rightarrow Z}$$

$$(\ell I') \quad \frac{X \rightarrow Y, \quad Z \subseteq XY, \quad Z \Rightarrow X}{Z \rightarrow Y}$$

However, in our opinion, the use of these schemes in our further developments is less intuitive, so we keep working with the four schemes above.

In the remainder of this section, we denote as $\mathcal{B}, \{X \rightarrow Y\} \vdash X' \rightarrow Y'$ the fact that, in the presence of the implications in the set \mathcal{B} , rule $X' \rightarrow Y'$ can be derived from rule $X \rightarrow Y$ using zero or more applications of the four deduction schemes; along such a derivation, any rule of \mathcal{B} (or derived from \mathcal{B} by the Armstrong schemes) can be used whenever an implication of the form $X \Rightarrow Y$ is required.

4.3. Soundness and Completeness. We can characterize the deductive power of this calculus as follows: it is sound and complete with respect to the notion of closure-based redundancy; that is, all the rules it can prove are redundant, and all the redundant rules can be proved:

Theorem 4.5. *Let \mathcal{B} consist of implications. Then, $\mathcal{B}, \{X_1 \rightarrow Y_1\} \vdash X_0 \rightarrow Y_0$ if and only if rule $X_0 \rightarrow Y_0$ has closure-based redundancy relative to \mathcal{B} with respect to rule $X_1 \rightarrow Y_1$: $\mathcal{B}, \{X_1 \rightarrow Y_1\} \models X_0 \rightarrow Y_0$.*

Proof. Soundness corresponds to the fact that every rule derived is redundant: it suffices to prove it individually for each scheme; the essentials of some of these arguments are also found in the literature. For (rA) , the inclusions $XY \subseteq XYZ \subseteq \overline{XY}$ prove that the partial rules above and below the line have the same confidence. For (rI) , one has $XZ \subseteq X\overline{Y} \subseteq \overline{XY}$, thus $s(XZ) \geq s(XY)$ and the confidence of the rule below the line is at least that of the one above, or possibly greater. Scheme (ℓA) is unchanged from the previous section. Finally, for (ℓI) , we have $Z \subseteq X \subseteq \overline{Z}$ so that $s(Z) = s(X)$, and $ZY \subseteq XY$ so that $s(ZY) \geq s(XY)$, and again the confidence of the rule below the line is at least the same as the confidence of the one above.

To prove completeness, we must see that all redundant rules can be derived. We assume $\mathcal{B}, \{X_1 \rightarrow Y_1\} \models X_0 \rightarrow Y_0$ and resort to Theorem 4.4: we know that the inclusions $X_1 \subseteq \overline{X_0}$ and $X_0 Y_0 \subseteq \overline{X_1 Y_1}$ must hold. From Lemma 4.1, we have that $\overline{X_0 Y_0} \subseteq \overline{X_1 Y_1}$.

Now we can write a derivation in our calculus, taking into account these inclusions, as follows:

$$X_1 \rightarrow Y_1 \vdash_{(rA)} X_1 \rightarrow X_1 Y_1 \vdash_{(rI)} X_1 \rightarrow \overline{X_0 Y_0} \vdash_{(\ell A)} \overline{X_0} \rightarrow Y_0 \vdash_{(\ell I)} X_0 \rightarrow Y_0$$

Thus, indeed the redundant rule is derivable, which proves completeness. \square

4.4. Optimum-Size Basis for Closure-Based Redundancy. In a similar way as we did for plain redundancy, we study here bases corresponding to closure-based redundancy.

Since the implications become “factored out” thanks to the stronger notion of redundancy, we can focus on the partial rules. A formal definition of completeness for a basis is, therefore, as follows:

Definition 4.6. Given a set of partial rules \mathcal{R} and a set of implications \mathcal{B} , *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}' .

Again \mathcal{R} is intended to be the set of all the partial rules “mined from” a given dataset \mathcal{D} at a confidence threshold $\gamma < 1$ (recall that always $\gamma > 0$), whereas \mathcal{B} is intended to be the subset of rules in \mathcal{R} that hold with confidence 1 in \mathcal{D} or, rather, a basis for these implications. There exist several proposals for constructing bases while taking into account the implications and their closure operator. We use the same intuitions and *modus operandi* to add a new proposal which, conceptually, departs only slightly from existing ones. Its main merit is not the conceptual novelty of the basis itself but the mathematical proof that it achieves the minimum possible size for a basis with respect to closure-based redundancy, and is therefore at most as large as any alternative basis and, in many cases, smaller than existing ones.

Our new basis is constructed as follows. For each closed set Y , we will consider a number of closed sets X properly included in Y as candidates to act as antecedents:

Definition 4.7. Fix a dataset \mathcal{D} , and consider the closure operator corresponding to the implications that hold in \mathcal{D} with confidence 1. For each closed set Y , a closed proper subset $X \subset Y$ is a *basic γ -antecedent* if the following holds:

- (1) X is a γ -antecedent of Y : $s(Y) \geq \gamma s(X)$;
- (2) no proper closed subset of X is a γ -antecedent of Y , and
- (3) no proper closed superset of Y has X as a γ -antecedent.

Basic antecedents follow essentially the same pattern as the valid antecedents (Definition 3.12), but restricted to closed sets only, that is, instead of minimal antecedents, we pick just minimal *closed* antecedents. Then we can use them as before:

Definition 4.8. Fix a dataset \mathcal{D} and a confidence threshold γ .

- (1) The basis \mathcal{B}_γ^* consists of all the rules $X \rightarrow Y - X$ for all closed sets Y and all basic γ -antecedents X of Y .
- (2) A *minmax variant* of the basis \mathcal{B}_γ^* is obtained by replacing each left-hand side in \mathcal{B}_γ^* by a minimal generator: that is, for a closed set Y , each rule $X \rightarrow Y - X$ becomes $X' \rightarrow Y - X$ for one minimal generator X' of the (closed) basic γ -antecedent X .
- (3) A *minmin variant* of the basis \mathcal{B}_γ^* is obtained by replacing by a minimal generator both the left-hand and the right-hand sides in \mathcal{B}_γ^* : for each closed set Y and each basic γ -antecedent X of Y , the rule $X \rightarrow Y - X$ becomes $X' \rightarrow Y' - X$ where Y' is chosen a minimal generator of Y and X' is chosen a minimal generator of X .

The variants are defined only for the purpose of discussing the relationship to previous works along the next few paragraphs; generally, we will use only the first version of \mathcal{B}_γ^* . Note the following: in a minmax variant, at the time of substituting a generator for the left-hand side closure, in case we consider a rule from \mathcal{B}_γ^* that has a left-hand side with

several minimal generators, only one of them is to be used. Also, all of X (and not only X') can be removed from the right-hand side: (rA) can be used to recover it.

The basis \mathcal{B}_γ^* is uniquely determined by the dataset and the confidence threshold, but the variants can be constructed, in general, in several ways, because each closed set in the rule may have several minimal generators, and even several different generators of minimum size. We can see the variants as applications of our deduction schemes. The result of substituting a generator for the left-hand side of a rule is equivalent to the rule itself: in one direction it is exactly scheme (ℓI) , and in the other is a chained application of (rA) to add the closure to the right-hand side and (ℓA) to put it back in the left-hand side. Substituting a generator for the right-hand side corresponds to scheme (rI) in both directions.

The use of generators instead of closed sets in the rules is discussed in several references, such as [36] or [44]. In the style of [36], we would consider a minmax variant, which allows one to show to the user minimal sets of antecedents together with all their nontrivial consequents. In the style of [44], we would consider a minmin variant, thus reducing the total number of symbols if minimum-size generators are used, since we can pick any generator. Each of these known bases incurs a risk of picking more than one minimum generator for the same closure as left-hand sides of rules with the same closure of the right-hand side: this is where they may be (and, in actual cases, have been empirically found to be) larger than \mathcal{B}_γ^* , because, in a sense, they would keep in the basis all the variants. Facts analogous to Corollaries 3.15 and 3.16 hold as well if the closure condition is added throughout, and provide further alternative definitions of the same basis. We use one of them in our experimental setting, described in Section 4.6. We now see that this set of rules entails exactly the rules that reach the corresponding confidence threshold in the dataset:

Theorem 4.9. *Fix a dataset \mathcal{D} and a confidence threshold γ . Let \mathcal{B} be any basis for implications that hold with confidence 1 in \mathcal{D} .*

- (1) *All the rules in \mathcal{B}_γ^* hold with confidence at least γ .*
- (2) *\mathcal{B}_γ^* is a complete basis for the partial rules under closure-based redundancy.*

Proof. All the rules in \mathcal{B}_γ^* must hold indeed because all the left-hand sides are actually γ -antecedents. To prove that all the partial rules that hold are entailed by rules in \mathcal{B}_γ^* , assume that indeed $X \rightarrow Y$ holds with confidence γ , that is, $s(\overline{XY}) = s(XY) \geq \gamma s(X)$; thus X is a γ -antecedent of \overline{XY} . If $Y \subseteq \overline{X}$, then $c(X \rightarrow Y) = 1$ and the implication will follow from \mathcal{B} ; we have to discuss only the case where $Y \not\subseteq \overline{X}$, which implies that $\overline{X} \subset \overline{XY}$. Consider the family of closed sets that include XY and have X as γ -antecedent; it is a nonempty family, since \overline{XY} fulfills these conditions. Pick Z maximal in that family. Then $\overline{X} \subset \overline{Z} = Z$ since $X \subseteq Z$ and $\overline{X} \subset \overline{XY} \subseteq Z$. Now, \overline{X} is a γ -antecedent of Z , but not of any strictly larger closed itemset. Also, any subset of \overline{X} is a proper subset of Z .

Let $X' \subseteq \overline{X}$ be closed, a γ -antecedent of Z , and minimal with respect to these properties; assume that X' is a γ -antecedent of a closed set Z' strictly larger than Z . From $X' \subseteq \overline{X} \subseteq Z \subset Z'$ and Lemma 3.11, \overline{X} would be also a γ -antecedent of Z' , which would contradict the maximality of Z . Therefore, X' cannot be a γ -antecedent of a closed set strictly larger than Z and, together with the facts that define X' , we have that X' is a basic γ -antecedent of Z whence $X' \rightarrow Z - X' \in \mathcal{B}_\gamma^*$.

We gather the following inequalities: $X' \subseteq \overline{X}$ and $XY \subseteq Z = \overline{Z} = \overline{X'(Z - X')}$; this is exactly what we need to infer that $\mathcal{B}, \{X' \rightarrow Z - X'\} \models X \rightarrow Y$ from Theorem 4.4. \square

Now we can move to the main result of this section: this basis has a minimum number of rules among all bases that are complete for the partial rules, according to closure-based redundancy with respect to \mathcal{B} .

Theorem 4.10. *Fix a dataset \mathcal{D} , and let \mathcal{R} be the set of rules that hold with confidence γ in \mathcal{D} . Let \mathcal{B} be a basis for the set of implications in \mathcal{R} . Let $\mathcal{B}' \subseteq \mathcal{R}$ be an arbitrary basis, having closure-based completeness for \mathcal{R} with respect to \mathcal{B} . Then, \mathcal{B}' must have at least as many rules as \mathcal{B}_γ^* .*

Proof. First, we will prove the following intermediate claim: for each partial rule in \mathcal{B}_γ^* , say $X \rightarrow Y - X$, there is in \mathcal{B}' a corresponding partial rule of the form $X' \rightarrow Y'$ with $\overline{X'Y'} = Y$ and $\overline{X'} = X$. We pick any rule $X \rightarrow Y - X \in \mathcal{B}_\gamma^*$, that is, where X is a basic γ -antecedent of Y ; this rule must be redundant, relative to the implications in \mathcal{B} , with respect to the new basis \mathcal{B}' under consideration: for some rule $X' \rightarrow Y' \in \mathcal{B}'$, we have that $\mathcal{B}, \{X' \rightarrow Y'\} \models X \rightarrow Y - X$ which, by Theorem 4.4, is the same as $X' \subseteq \overline{X} = X$ and $Y \subseteq \overline{X'Y'}$, together with $c(X' \rightarrow Y') \geq \gamma$. We consider some support ratios: $\frac{s(\overline{X'Y'})}{s(X)} = \frac{s(X'Y')}{s(\overline{X})} \geq \frac{s(X'Y')}{s(X')} \geq \gamma$, which means that X is a γ -antecedent of $\overline{X'Y'}$, a closed set including Y ; by the second condition in the definition of basic γ -antecedent, this cannot be the case unless $\overline{X'Y'} = Y$.

Then, again, $c(X' \rightarrow Y) = c(X' \rightarrow X'Y') = c(X' \rightarrow Y') \geq \gamma$, that is, X' is a γ -antecedent of Y , and $\overline{X'} \subseteq \overline{Y} = Y$ is as well; but $\overline{X'} \subseteq \overline{X} = X$ and, by minimality of X as a basic γ -antecedent of Y , it must be that $\overline{X'} = X$.

Now, to complete the proof of the theorem, we observe that each such rule $X' \rightarrow Y'$ in \mathcal{B}' determines univocally both closed sets X and Y , so that the same rule in \mathcal{B}' cannot correspond to more than one of the rules in \mathcal{B}_γ^* . This requires \mathcal{B}' , therefore, to have at least as many rules as \mathcal{B}_γ^* . \square

In applications of \mathcal{B}_γ^* , one needs, in general, as a basis both \mathcal{B}_γ^* and a basis for the implications, such as the GD-basis. On the other hand, in many practical cases, implications provide little new knowledge, most often just showing existing (and known) properties of the attributes. If a user is satisfied with the \mathcal{B}_γ^* basis, and does not ask for a basis for the implications nor the representative rules, then (s)he may get results faster, since in this case the algorithms would not need to compute minimal generators, and just mining closures and their supports (and organizing them via the subset relation) would suffice.

Note that the joint consideration of the GD-basis and \mathcal{B}_γ^* incurs the risk of being a larger set of rules than the representative rules, due to the fact that some rules in the GD-basis could be, in fact, plainly redundant (ignoring the closure-related issues) with a representative rule. We have observed empirically that, at high confidence thresholds, the representative rules tend to be a large basis due to the lack of specific minimization of implications, whereas the union of the GD-basis and \mathcal{B}_γ^* tends to be quite smaller; conversely, at lower confidence levels, the availability of many partial rules increases the chances of covering a large part of the GD-basis, so that the representative rules are a smaller basis than the union of \mathcal{B}_γ^* plus GD, even if they are more in number than \mathcal{B}_γ^* . That is: closure-based redundancy may be either stronger or weaker, in terms of the optimum basis sizes, than plain redundancy. Sometimes, \mathcal{B}_γ^* even fully coincides with the partial representative rules. This is, in fact, illustrated in the following example.

Example 4.11. We revisit the example in Figure 1. As indicated at the end of Section 3.2, the basis for implications consists of six rules: $AC \Rightarrow B$, $AD \Rightarrow B$, $BC \Rightarrow A$, $BD \Rightarrow A$,

$CF \Rightarrow D$, and $DF \Rightarrow C$; the iteration-free basis [43] and the Guigues-Duquenne basis [23] coincide here, and these implications are also the representative rules at confidence 1. At confidence $\gamma = 0.75$, these are kept and four *representative rules* are added: $A \rightarrow B$, $B \rightarrow A$, $AB \rightarrow C$, and $D \rightarrow C$. Since the four left-hand sides are, actually, closed sets, which is not guaranteed in general, the basis \mathcal{B}_γ^* at this confidence includes exactly these four rules: no other closure is a basic γ -antecedent.

However, if the confidence threshold is lowered to $\gamma = 0.6$, we find seven rules in the $\mathcal{B}_{0.6}^*$ basis: $A \rightarrow BC$, $B \rightarrow AC$, $C \rightarrow D$, $D \rightarrow C$, $CD \rightarrow F$, and $F \rightarrow CD$, plus the somewhat peculiar $\emptyset \rightarrow C$, since indeed the support of C is above the same threshold; the rules $A \rightarrow B$, $B \rightarrow A$, and $AB \rightarrow C$ also hold, but they are redundant with respect to $A \rightarrow BC$ or $B \rightarrow AC$: A and B are γ -antecedents of AB but are not basic (by way of being also γ -antecedents of ABC), whereas AB is a γ -antecedent of ABC but is not basic either since it is not minimal.

Additionally, the sizes of the rules can be reduced somewhat: $A \rightarrow C$ suffices to give $A \rightarrow BC$ or indeed $A \rightarrow ABC$ since $A \rightarrow C$ is equivalent by reflexivity to $A \rightarrow AC$ and there is a full-confidence implication $AC \Rightarrow B$ in the GD-basis that gives us $A \rightarrow ABC$. This form of reasoning is due to [44], and a similar argument can be made for several of the other rules. Alternatively, there exists the option of omitting those implications that, seen as partial rules, are already covered by a partial rule: in this example, these are $AC \Rightarrow B$ and $BC \Rightarrow A$, covered by $A \rightarrow BC$ (but *not* by $A \rightarrow C$, which needs $AC \Rightarrow B$ to infer $A \rightarrow BC$); similarly, $CF \Rightarrow D$ and $CD \Rightarrow F$ are plainly redundant with $C \rightarrow DF$. In fact, it can be readily checked that the seven partial rules in $\mathcal{B}_{0.6}^*$ plus the two remaining implications in the GD-basis, $AD \Rightarrow B$ and $BD \Rightarrow A$, form exactly the representative rules at this confidence threshold.

4.5. Double-Support Mining. For many real-life datasets, including all the standard benchmarks in the field, the closure space is huge, and reaches easily hundreds of thousands of nodes, or indeed even millions. A standard practice, as explained in the introduction, is to impose a support constraint, that is, to ignore (closed) sets that do not appear often enough. It has been observed also that the rules removed by this constraint are often appropriately so, in that they are less robust and prone to represent statistical artifacts rather than true information [34]. Hence, we discuss briefly what happens to our basis proposal if we work under such a support constraint.

For a dataset \mathcal{D} and confidence and support thresholds γ and τ , respectively, denote by $\mathcal{R}_{\gamma,\tau}$ the set of rules that hold in \mathcal{D} with confidence at least γ and support at least τ . We may want to construct either of two similar but different sets of rules: we can ask just how to compute the set of rules in \mathcal{B}_γ^* that reach that support or, more likely, we may wish a minimum-size basis for $\mathcal{R}_{\gamma,\tau}$. We solve both problems.

We first discuss a minimum-size basis for $\mathcal{R}_{\gamma,\tau}$. Of course, the natural approach is to compute the rule basis exactly as before, but only using closed sets above the support threshold. Indeed this works:

Proposition 4.12. *Fix a dataset \mathcal{D} . For any fixed confidence threshold γ and support threshold τ , the construction of basic γ -antecedents, applied only to closed sets of support at least τ , provides a minimum-size basis for $\mathcal{R}_{\gamma,\tau}$.*

Proof. Consider any rule $X \rightarrow Y$ of support at least τ and confidence at least γ . Then \overline{X} is a γ -antecedent of \overline{XY} ; also, $s(\overline{X}) = s(X) \geq s(XY) = s(\overline{XY}) \geq \tau$.

Arguing as in the proof of Theorem 4.9 but restricted to the closures with support at least τ , we can find a rule $X' \rightarrow Y' - X'$ where both $\overline{X'}$ and $\overline{X'Y'}$ have support at least τ , $\overline{X'}$ is a basic γ -antecedent of $\overline{X'Y'}$, and such that $X' \subseteq \overline{X}$ and $XY \subseteq \overline{X'Y'}$ so that it covers $X \rightarrow Y$. Minimum size is argued exactly as in the proof of Theorem 4.10: following the same steps, one proves that any complete basis consisting of rules in $\mathcal{R}_{\gamma,\tau}$ must have separate rules to cover each of the rules formed by basic γ -antecedents of closures of support τ . \square

We are therefore safe if we apply the basis construction for \mathcal{B}_γ^* to a lattice of frequent closed sets above support τ , instead of the whole lattice of closed sets. However, this fact does not ensure that the basis obtained coincides with the set of rules in the whole basis \mathcal{B}_γ^* having support above τ . There may be rules that are not in \mathcal{B}_γ^* because a large closure, of low support, prevents some X from being a basic antecedent. If the large closure is pruned by the support constraint, then X may become a basic antecedent. The following result explains with more precision the relationship between the basis \mathcal{B}_γ^* and the rules of support τ .

Proposition 4.13. *Fix a dataset \mathcal{D} , a confidence threshold γ , and a support threshold τ . Assume that $X \subseteq Y$ and that $s(Y) \geq \tau$; then $X \rightarrow Y - X \in \mathcal{B}_\gamma^*$ if and only if X is a basic γ -antecedent of Y in the set of all closures of support at least $\gamma \times \tau$.*

This proposition says that, in order to find $\mathcal{B}_\gamma^* \cap \mathcal{R}_{\gamma,\tau}$, that is, the set of rules in \mathcal{B}_γ^* that have support at least τ , we do not need to compute all the closures and construct the whole of \mathcal{B}_γ^* ; it suffices to perform the \mathcal{B}_γ^* construction on the set of closures of support $\gamma \times \tau$. Of course, in both cases we must then discard the rules of support less than τ . We call this sort of process *double-support mining*: given user-defined γ and τ , use the product to find all closures of support $\gamma \times \tau$, compute \mathcal{B}_γ^* on these closures, and finally prune out the rules with support less than τ to obtain $\mathcal{B}_\gamma^* \cap \mathcal{R}_{\gamma,\tau}$, if that is what is desired.

Proof. Consider a pair of closed sets $X \subset Y$ with $s(X) > s(Y) \geq \tau$; we must discuss whether X is a basic γ -antecedent of Y in two different closure lattices: the one of all the closed sets and the one of frequent closures at support threshold $\gamma \times \tau$.

The properties of being a γ -antecedent and of being minimally so refer to X and Y themselves or to even smaller sets, and are therefore unaffected by the support constraint. We must discuss just the existence of some proper superset of Y having X as a γ -antecedent. In case X is a basic γ -antecedent of Y , no proper superset Z of Y has X as γ -antecedent, whatever the support of Z ; therefore, X will be found to be a basic γ -antecedent of Y also in the smaller lattice of frequent closures.

To show the converse, it suffices to argue that, for any proper superset Z of Y , if X is a γ -antecedent of Z , then $s(Z) \geq \gamma \times \tau$. Indeed, $s(Z) \geq \gamma s(X) \geq \gamma \times \tau$; hence, if no such Z is found in the frequent closures lattice at support threshold $\gamma \times \tau$, no such Z exists at all. \square

4.6. Empirical Evaluation. Whereas our interests in this paper are rather foundational, we wish to describe briefly the direct applicability of our results so far. We have chosen an approach that conveniently uses as a black-box a separate closed itemsets miner due to Borgelt [8]. We have implemented a construction of the GD basis using a hypergraph transversal method to construct representative rules of confidence 1 following the guidelines of [37] and subsequently simplifying them to obtain the GD basis as per [4]; and we have

Dataset	S/C	Traditional	Closure-based	RR Imp	GD	\mathcal{B}_γ^*	Sum
Chess	80	552564	27711	2228	5	226	231
Chess	70	8171198	152074	13193	10	891	901
Connect	97	8092	1116	161	4	41	45
Connect	90	3640704	18848	3359	14	222	236
Mushroom	40	7020	475	170	24	41	65
Mushroom	20	19191656	5741	1739	177	158	328
Pumsb	95	1170	267	62	2	32	34
Pumsb	85	1408950	44483	9559	9	1080	1089
Pumsb_star	60	2358	192	43	5	6	11
Pumsb_star	40	5659536	13479	2939	48	82	129
T10I4D100K	0.5	2216	1231	0	0	585	585
T10I4D100K	0.1	431838	86902	582	214	4054	4268

Table 1: Number of rules in various bases for benchmark datasets.

implemented a simple algorithm that scans repeatedly the closed sets mined by the separate program and constructs all basic γ -antecedents. A first scan picks up γ -antecedents from the proper closed subsets and filters them for minimality; once all minimal antecedents are there for all closures, a subsequent scan filters out those that are not basic by way of being antecedents of larger sets. Effectively the algorithm does not implement the definition but the immediate extension of the characterization in Corollary 3.16 to the closure-based case.

A natural alternative consists in preprocessing 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, we found that the computation slows down unacceptably, probably due to a worse fit to virtual memory caching. Our implementation gives us 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.

On the basis of this implementation, we have undertaken some empirical evaluations of the sizes of the basis. We consider that the key point of our contribution is the mathematical proof of absolute size minimality, but, as a mere illustration, we show the figures of some of the cases explored in [44] in Table 1. The datasets and thresholds are set exactly as per that reference; column “S/C” is the confidence and support parameters. Columns “Traditional” (for the number of rules under the standard traditional definition [2]) and “Closure-based” (for the number of rules obtained by the closure-based method proposed in [44]) are taken verbatim from the same reference. We have added the number of rules in the representative basis for implications at 100% confidence “RRImp”, that coincides with the iteration-free basis [43] and other proposals as discussed at the beginning of Subsection 4.1; the size of the GD basis for the same implications (often yielding huge savings); and the number of rules in the \mathcal{B}_γ^* basis of partial rules, which, in the totality of these cases, did coincide with the representative rules at the corresponding thresholds. As discussed in the end of Section 4.4, representative rules encompass implications but \mathcal{B}_γ^* must be taken jointly with the GD basis, so we give also the corresponding sum.

The confidence chosen in [44] for this comparison, namely, coincident with the support threshold, is, in our opinion, too low to provide a good perspective; at these thresholds, representative rules essentially correspond to support bounds (rules with empty left-hand side).

To complement the intuition, we provide the evolution of the sizes of the representative rules and the \mathcal{B}_γ^* basis for the dataset `pumsb-star`, downloaded from [17], at the same support thresholds of 40% and 60% used in Table 1, with confidence ranging from 99% to 51%, at 1% granularity. The Guigues-Duquenne bases at these support thresholds consist of 48 and 5 rules respectively. These have been added to the size of \mathcal{B}_γ^* in Figures 2 and 3. At these confidence thresholds, the traditional notion of association rules gives from 105086 up to 179684 rules at support 40%, and between 268 and 570 rules at support 60%. Note that, in that notion, association rules are restricted, by definition, to singleton consequents; larger numbers would be found if this condition is lifted for a fairer comparison with the bases we study. These figures show the advantage of the closure-based basis over representative rules up to the point where the implications become subsumed by partial representative rules.

We want to point out as well one interesting aspect of the figures obtained. The standard settings for association rules lead to a monotonicity property, by which lower confidence thresholds allow for more rules, so that the size of the output grows (sometimes enormously) as the confidence threshold decreases. However, in the case of the \mathcal{B}_γ^* basis and the representative rules, some datasets exhibit a nonmonotonic evolution: at lesser confidence thresholds, sometimes less rules are obtained. 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.

5. TOWARDS GENERAL ENTAILMENT

We move on towards a further contribution of this paper: we propose a stronger notion of redundancy, as progress towards a complete logical approach, where redundancy would play the role of entailment and a sound and complete deductive calculus is sought. Considering the redundancy notions described so far, the following question naturally arises: beyond all these notions of redundancy that relate one partial rule to another partial rule, possibly in presence of implications, is it indeed possible that a partial rule is entailed jointly by two partial rules, but not by a single one of them? and, if so, when does this happen? We will fully answer this question below.

The failures of Transitivity and Augmentation may suggest the intuition of a negative answer: it looks like any combination of two partial rules of confidence at least γ , but with $\gamma < 1$, will require us to multiply confidences, reaching as low as γ^2 or lower; but this intuition is wrong. We will characterize precisely the case where, at a fixed confidence threshold, a partial rule follows from exactly two partial rules, a case where our previous calculus becomes incomplete; and we will identify one extra deduction scheme that allows us to conclude as consequent a partial rule from two premise partial rules in a sound form. The calculus obtained is complete with respect to entailment from two premise rules. We present the whole setting in terms of closure-based redundancy, but the development carries over for plain redundancy, simply by taking the identity as closure operator.

A first consideration is that we no longer have a single value of the confidence to compare; therefore, we take a position like the one in most cases of applications of association rule mining in practice, namely: we fix a confidence threshold, and consider only rules whose confidence is above it. An alternative view, further removed from practice, would be

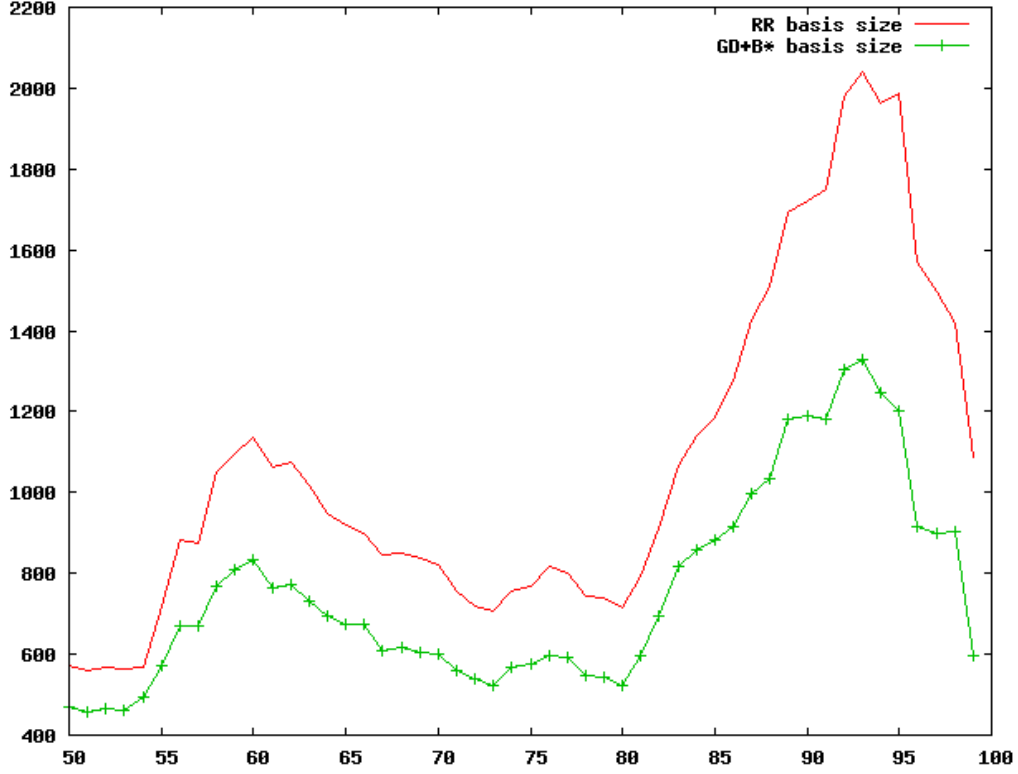


Figure 2: Basis sizes per confidence in pumsb-star at 40% support

to require just that the confidence of all our conclusions should be at least the same as the minimum of the confidences of the premises.

As an example, consider the following fact (the analogous statement for $\gamma < 1/2$ does not hold, as discussed below):

Proposition 5.1. *Let $\gamma \geq 1/2$. Assume that items A, B, C, D are present in \mathcal{U} and that the confidence of the rules $A \rightarrow BC$ and $A \rightarrow BD$ is above γ in dataset \mathcal{D} . Then, the confidence of the rule $ACD \rightarrow B$ in \mathcal{D} is also above γ .*

We do not provide a formal proof of this claim since it is just the simplest particular case of Theorem 5.3 below. We consider the following definition:

Definition 5.2. Given a set \mathcal{B} of implications, and a set \mathcal{R} of partial rules, rule $X_0 \rightarrow Y_0$ is γ -redundant with respect to them (or also γ -entailed by them), denoted $\mathcal{B}, \mathcal{R} \models_{\gamma} X_0 \rightarrow Y_0$, if every dataset in which the rules of \mathcal{B} have confidence 1 and the confidence of all the rules in \mathcal{R} is at least γ must satisfy as well $X_0 \rightarrow Y_0$ with confidence at least γ . The entailment is called “proper” if it does not hold for proper subsets of \mathcal{R} ; otherwise it is “improper”.

Note that, in this case, the parameter γ is necessary to qualify the entailment relation itself. In previous sections we had a mere confidence inequality that did not depend on γ . The main result of this section is now:

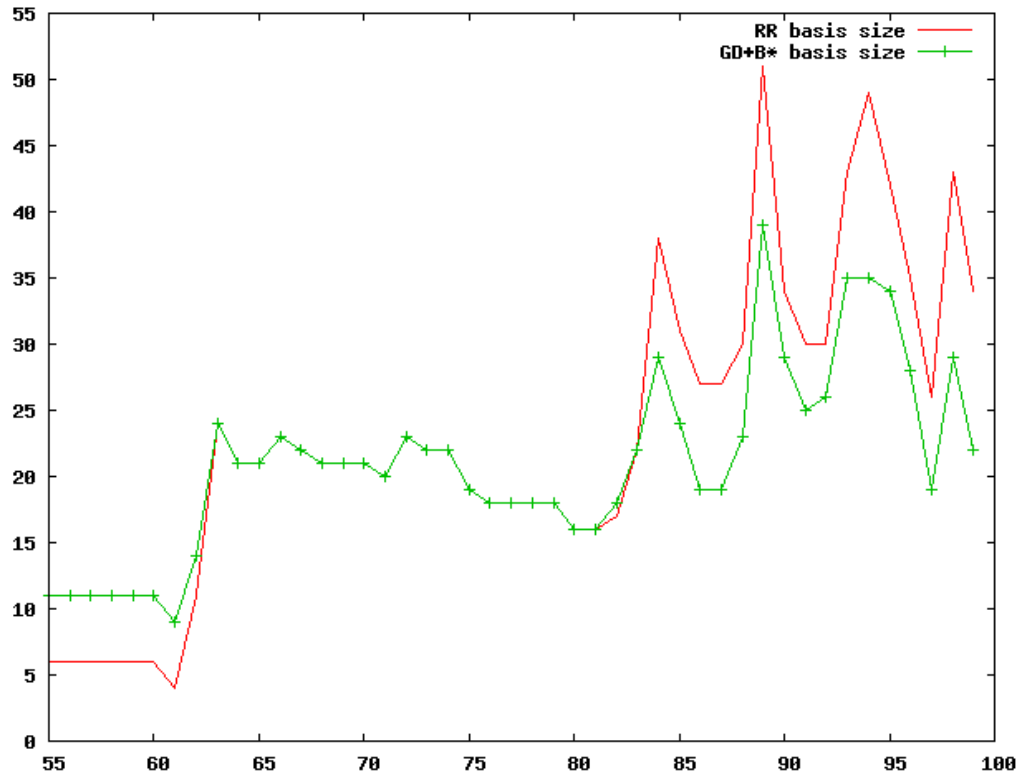


Figure 3: Basis sizes per confidence in pumsb-star at 60% support

Theorem 5.3. *Let \mathcal{B} be a set of implications, and let $1/2 \leq \gamma < 1$. Consider three partial rules, $X_0 \rightarrow Y_0$, $X_1 \rightarrow Y_1$, and $X_2 \rightarrow Y_2$. Then, $\mathcal{B}, \{X_1 \rightarrow Y_1, X_2 \rightarrow Y_2\} \models_{\gamma} X_0 \rightarrow Y_0$ if and only if either:*

- (1) $Y_0 \subseteq \overline{X_0}$, or
- (2) $\mathcal{B}, \{X_1 \rightarrow Y_1\} \models X_0 \rightarrow Y_0$, or
- (3) $\mathcal{B}, \{X_2 \rightarrow Y_2\} \models X_0 \rightarrow Y_0$, or
- (4) all the following conditions simultaneously hold:
 - (i) $X_1 \subseteq \overline{X_0}$
 - (ii) $X_2 \subseteq \overline{X_0}$
 - (iii) $X_1 \subseteq \overline{X_2 Y_2}$
 - (iv) $X_2 \subseteq \overline{X_1 Y_1}$
 - (v) $X_0 \subseteq \overline{X_1 Y_1 X_2 Y_2}$
 - (vi) $Y_0 \subseteq \overline{X_0 Y_1}$
 - (vii) $Y_0 \subseteq \overline{X_0 Y_2}$

Proof. Let us discuss first the leftwards implication. In case (1), rule $X_0 \rightarrow Y_0$ holds trivially. Clearly cases (2) and (3) also give (improper) entailment. For case (4), we must argue that, if all the seven conditions hold, then the entailment relationship also holds. Thus, fix any dataset \mathcal{D} where the confidences of the premise rules are at least γ : these assumptions can

be written, respectively, $s(X_1Y_1) \geq \gamma s(X_1)$ and $s(X_2Y_2) \geq \gamma s(X_2)$, or equivalently for the corresponding closures.

We have to show that the confidence of $X_0 \rightarrow Y_0$ in \mathcal{D} is also at least γ . Consider the following four sets of transactions from \mathcal{D} :

$$\begin{aligned} A &= \{t \in \mathcal{D} \mid t \models X_0Y_0\} \\ B &= \{t \in \mathcal{D} \mid t \models X_0, t \not\models X_0Y_0\} \\ C &= \{t \in \mathcal{D} \mid t \models X_1Y_1, t \not\models X_0\} \\ D &= \{t \in \mathcal{D} \mid t \models X_2Y_2, t \not\models X_0\} \end{aligned}$$

and let a, b, c , and d be the respective cardinalities.

We first argue that all four sets are mutually disjoint.

This is easy for most pairs: clearly A and B have incompatible behavior with respect to Y_0 ; and a tuple in either A or B has to satisfy X_0 , which makes it impossible that that tuple is accounted for in either C or D . The only place where we have to argue a bit more carefully is to see that C and D are disjoint as well: but a tuple t that satisfies both X_1Y_1 and X_2Y_2 , that is, satisfies their union $X_1Y_1X_2Y_2$, must satisfy every subset of the corresponding closure as well, such as X_0 , due to condition (v). Hence, C and D are disjoint.

Now we bound the supports of the involved itemsets as follows: clearly, by definition of A , $s(X_0Y_0) = a$. All tuples that satisfy X_0 are accounted for either as satisfying Y_0 as well, in A , or in B in case they don't; disjointness then guarantees that $s(X_0) = a + b$.

We see also that $s(X_1) \geq a + b + c + d$, because X_1 is satisfied by the tuples in C , by definition; by the tuples in A or B , by condition (i); and by the tuples in D , by condition (iii); again disjointness allows us to sum all four cardinalities. Similarly, using instead (ii) and (iv), we obtain $s(X_2) \geq a + b + c + d$.

The next delicate point is to show an upper bound on $s(X_1Y_1)$ (and on $s(X_2Y_2)$ symmetrically). We split all the tuples that satisfy X_1Y_1 into two sets, those that additionally satisfy X_0 , and those that don't. Tuples that satisfy X_1Y_1 and not X_0 are exactly those in C , and there are exactly c many of them. Satisfying X_1Y_1 and X_0 is the same as satisfying X_0Y_1 by condition (i), and tuples that do it must also satisfy Y_0 by condition (vi). Therefore, they satisfy both X_0 and Y_0 , must belong to A , and there can be at most a many of them. That is, $s(X_1Y_1) \leq a + c$ and, symmetrically, resorting to (ii) and (vii), $s(X_2Y_2) \leq a + d$.

Thus we can write the following inequations:

$$\begin{aligned} a + c &\geq s(X_1Y_1) \geq \gamma s(X_1) \geq \gamma(a + b + c + d) \\ a + d &\geq s(X_2Y_2) \geq \gamma s(X_2) \geq \gamma(a + b + c + d) \end{aligned}$$

Adding them up, using $\gamma \geq \frac{1}{2}$, we get

$$2a + c + d \geq 2\gamma(a + b + c + d) = 2\gamma(a + b) + 2\gamma(c + d) \geq 2\gamma(a + b) + c + d$$

that is, $a \geq \gamma(a + b)$, so that

$$c(X_0 \rightarrow Y_0) = \frac{s(X_0Y_0)}{s(X_0)} = \frac{a}{a + b} \geq \gamma$$

as was to be shown.

Now we prove the rightwards direction; the bound $\gamma \geq \frac{1}{2}$ is not necessary for this part. Since all our supports are integers, we can assume that the threshold is a rational number, $\gamma = \frac{m}{n}$, so that we can count on $n - m > 0$ and $1 \leq m \leq n - 1$. We will argue the

contrapositive, assuming that we are in neither of the four cases, and showing that the entailment does not happen, that is, it is possible to construct a counterexample dataset for which all the implications in \mathcal{B} hold, and the two premise partial rules have confidence at least γ , whereas the rule in the conclusion has confidence strictly below γ . This requires us to construct a number of counterexamples through a somewhat long case analysis. In all of them, all the tuples will be closed sets with respect to \mathcal{B} ; this ensures that these implications are satisfied in all the transactions. We therefore assume that case (1) does not happen, that is, $Y_0 \not\subseteq \overline{X_0}$; and that cases (2) and (3) do not happen either. Now, Theorem 4.4 tells us that $X_1 \subseteq \overline{X_0}$ implies $X_0Y_0 \not\subseteq \overline{X_1Y_1}$, and that $X_2 \subseteq \overline{X_0}$ implies $X_0Y_0 \not\subseteq \overline{X_2Y_2}$. Along the rest of the proof, we will refer to the properties explained in this paragraph as the “known facts”.

Then, assuming that case (4) does not hold either, we have to consider multiple ways for the conditions (i) to (vii) to fail. Failures of (i) and (ii), however, cannot be argued separately, and we discuss them together.

Case A. Exactly one of (i) and (ii) fails. By symmetry, renaming $X_1 \rightarrow Y_1$ into $X_2 \rightarrow Y_2$ if necessary, we can assume that (i) fails and (ii) holds. Thus, $X_1 \not\subseteq \overline{X_0}$ but $X_2 \subseteq \overline{X_0}$. Then, by the known facts, $X_0Y_0 \not\subseteq \overline{X_2Y_2}$. We consider a dataset consisting of one transaction with the itemset $\overline{X_2Y_2}$, $mn - 1$ transactions with the set $\overline{X_0X_1Y_1X_2Y_2}$, and $n(n - m)$ transactions with the set $\overline{X_0}$, for a total of n^2 transactions. Then, the support of X_0 is either $n^2 - 1$ or n^2 , and the support of X_0Y_0 is at most $mn - 1$, for a confidence bounded by $\frac{mn-1}{n^2-1} < \frac{mn}{n^2} = \gamma$ for the rule $X_0 \rightarrow Y_0$. However, the premise rules hold: since (i) fails, the support of X_1 is at most mn , and the support of X_1Y_1 is at least $mn - 1$, for a confidence at least $\frac{mn-1}{mn} \geq \frac{m}{n} = \gamma$ for $X_1 \rightarrow Y_1$; whereas the support of X_2 is n^2 , that of X_2Y_2 is at least nm , and therefore the confidence is at least $m/n = \gamma$.

Case B. This corresponds to both of (i) and (ii) failing. Then, for a dataset consisting only of $\overline{X_0}$'s, the premise rules hold vacuously whereas $X_0 \rightarrow Y_0$ fails. We can also avoid arguing through rules holding vacuously by means of a dataset consisting of one transaction $\overline{X_0X_1Y_1X_2Y_2}$ and n transactions $\overline{X_0}$.

Remark. For the rest of the cases, we will assume that both of (i) and (ii) hold, since the other situations are already covered. Then, by the known facts, we can freely use the properties $X_0Y_0 \not\subseteq \overline{X_1Y_1}$ and $X_0Y_0 \not\subseteq \overline{X_2Y_2}$.

Case C. Assume (iii) fails, $X_1 \not\subseteq \overline{X_2Y_2}$, and consider a dataset consisting of one transaction $\overline{X_0}$, n transactions $\overline{X_1Y_1}$, and n^2 transactions $\overline{X_2Y_2}$. Here, by the known facts, the support of X_0Y_0 is zero. It suffices to check that the antecedent rules hold. Since (iii) fails, and (i) holds, the support of X_1 is exactly $n + 1$ and the support of X_1Y_1 is at least n , for a confidence of at least $\frac{n}{n+1} > \frac{n-1}{n} \geq \frac{m}{n} = \gamma$; whereas the support of X_2 is at most $n^2 + n + 1$ (depending on whether (iv) holds) for a confidence of rule $X_2 \rightarrow Y_2$ of at least $\frac{n^2}{n^2+n+1}$ which is easily seen to be above $\frac{n-1}{n} \geq \frac{m}{n} = \gamma$.

The case where (iv) fails is fully symmetrical and can be argued just interchanging the roles of $X_1 \rightarrow Y_1$ and $X_2 \rightarrow Y_2$.

Case D. Assume (v) fails. It suffices to consider a dataset with one transaction $\overline{X_0}$ and $n - 1$ transactions $\overline{X_1Y_1X_2Y_2}$. Using (i) and (ii), for both premises the confidence is $\frac{n-1}{n} \geq \gamma$, the support of X_0 is 1, and the support of X_0Y_0 is zero by the known fact $Y_0 \not\subseteq \overline{X_0}$ and the failure of (v).

Case E. We assume that (vi) fails, but a symmetric argument takes care of the case where (vii) fails. Thus, we have $Y_0 \not\subseteq \overline{X_0Y_1}$. By treating this case last, we can assume (i), (ii), and (v) hold, and also the known facts that $X_0Y_0 \not\subseteq \overline{X_1Y_1}$ and $X_0Y_0 \not\subseteq \overline{X_2Y_2}$. We consider a dataset with one transaction $\overline{X_0Y_1}$, one transaction $\overline{X_2Y_2}$, $m-1$ transactions $\overline{X_1Y_1X_2Y_2}$, and $n-m-1$ transactions $\overline{X_0}$ (note that this last part may be empty, but $n-m-1 \geq 0$; the total is n transactions). By (v), the support of X_0 is at least $n-1$, whereas the support of X_0Y_0 is at most $m-1$, given the available facts. Since $\frac{m-1}{n-1} < \gamma$, rule $X_0 \rightarrow Y_0$ does not hold. However, the premises hold: all supports are at most n , the total size, and the supports of X_1Y_1 (using (i)) and X_2Y_2 are both m .

This completes the proof. \square

A small point that remains to be clarified is the role of the condition $\gamma \geq 1/2$. As indicated in the proof of the theorem, that condition is only necessary in one of the two directions. If there is entailment, the conditions enumerated must hold irrespective of the value of γ . In fact, for $0 < \gamma < 1/2$, proper entailment from a set of two (or more) premises never holds, and γ -entailment in general is characterized as (closure-based) redundancy as per Theorem 4.4 and the corresponding calculus. Indeed:

Theorem 5.4. *Let $0 < \gamma < 1/2$. Then, $\mathcal{B}, \{X_1 \rightarrow Y_1, X_2 \rightarrow Y_2\} \models_\gamma X_0 \rightarrow Y_0$ if and only if either:*

- (1) $Y_0 \subseteq \overline{X_0}$, or
- (2) $\mathcal{B}, \{X_1 \rightarrow Y_1\} \models X_0 \rightarrow Y_0$, or
- (3) $\mathcal{B}, \{X_2 \rightarrow Y_2\} \models X_0 \rightarrow Y_0$.

Proof. The leftwards proof is already part of Theorem 5.3. For the converse, assume that the three conditions fail: similarly to the previous proof, we have as known facts the following: $Y_0 \not\subseteq \overline{X_0}$, $X_1 \subseteq \overline{X_0}$ implies $X_0Y_0 \not\subseteq \overline{X_1Y_1}$ and $X_2 \subseteq \overline{X_0}$ implies $X_0Y_0 \not\subseteq \overline{X_2Y_2}$. We prove that there are datasets giving low confidence to $X_0 \rightarrow Y_0$ and high confidence to both premise rules. If both $X_1 \not\subseteq \overline{X_0}$ and $X_2 \not\subseteq \overline{X_0}$ then we consider one transaction $\overline{X_1Y_1}$, one transaction $\overline{X_2Y_2}$, and a large number m of transactions $\overline{X_0}$ which do not change the confidences of the premises but lead to a confidence of at most $2/m$ for $X_0 \rightarrow Y_0$. Also, if $X_1 \not\subseteq \overline{X_0}$ but $X_2 \subseteq \overline{X_0}$, where the symmetric case is handled analogously, we are exactly as in Case A in the proof of Theorem 5.3 and argue in exactly the same way.

The interesting case is when both $X_1 \subseteq \overline{X_0}$ and $X_2 \subseteq \overline{X_0}$; then both $X_0Y_0 \not\subseteq \overline{X_1Y_1}$ and $X_0Y_0 \not\subseteq \overline{X_2Y_2}$. We fix any integer $k \geq \frac{\gamma}{1-2\gamma}$ and use the fact that $\gamma < 1/2$ to ensure that the fraction is positive and that the inequality can be transformed, by solving for γ , into $\frac{k}{2k+1} \geq \gamma$ (following these steps for $\gamma \geq 1/2$ either makes the denominator null or reverses the inequality due to a negative sign). We consider a dataset with one transaction for $\overline{X_0}$ and k transactions for each of $\overline{X_1Y_1}$ and $\overline{X_2Y_2}$. Even in the worst case that either or both of X_1 and X_2 show up in all transactions, the confidences of $X_1 \rightarrow Y_1$ and $X_2 \rightarrow Y_2$ are at least $\frac{k}{2k+1} \geq \gamma$, whereas the confidence of $X_0 \rightarrow Y_0$ is zero. \square

5.1. Extending the calculus. We work now towards a rule form, in order to enlarge our calculus with entailment from larger sets of premises. We propose the following additional rule:

$$(2A) \quad \frac{X_1 \rightarrow Y_1, \quad X_2 \rightarrow Y_2, \quad X_1Y_1 \Rightarrow X_2, \quad X_2Y_2 \Rightarrow X_1, \quad X_1Y_1X_2Y_2 \Rightarrow Z_1, \quad X_1Y_1Z_1 \Rightarrow Z_2, \quad X_2Y_2Z_1 \Rightarrow Z_2}{X_1X_2Z_1 \rightarrow Z_2}$$

and state the following properties:

Theorem 5.5. *Given a threshold $\gamma \geq 1/2$ and a set \mathcal{B} of implications,*

- (1) *this deduction scheme is sound, and*
- (2) *together with the deduction schemes in Section 4.2, it gives a calculus complete with respect to all entailments with two partial rules in the antecedent.*

Proof. This follows easily from Theorem 5.3, in that it implements the conditions of case (4); soundness is seen by directly checking that the conditions (i) to (vii) in case 4 of Theorem 5.3 hold: let $X_0 = X_1X_2Z_1$ and $Y_0 = Z_2$; then, conditions (i) and (ii) hold trivially, and the rest are explicitly required in the form of implications in the premises (notice that $X_1Y_1 \Rightarrow X_2$ implies that $X_1Y_1Z_1 \Rightarrow Z_2$ and $X_1X_2Y_1Z_1 \Rightarrow Z_2$ are equivalent). Completeness is argued by considering any rule $X_0 \rightarrow Y_0$ entailed by $X_1 \rightarrow Y_1$ and $X_2 \rightarrow Y_2$ jointly with respect to confidence threshold γ ; if the entailment is improper, apply Theorem 4.5, otherwise just apply this new deduction scheme with $Z_1 = \overline{X_0}$ and $Z_2 = Y_0$ to get $\overline{X_0} \rightarrow Y_0$ and apply (ℓI) to obtain $X_0 \rightarrow Y_0$. It is easy to see that the scheme is indeed applicable: proper entailment implies that all seven conditions in case (4) hold and, for $Z_1 = \overline{X_0}$, we get from (i) and (ii) that $X_1X_2Z_1 = Z_1$; under this equality, the remaining five conditions provide exactly the premises of the new deduction scheme. \square

6. DISCUSSION

Our main contribution, at a glance, is a study of confidence-bounded association rules in terms of a family of notions of redundancy. We have provided characterizations of several existing redundancy notions; we have described how these previous proposals, once the relationship to the most robust definitions has been clarified, provide a sound and complete deductive calculus for each of them; and we have been able to prove global optimality of an existing basis proposal, for the plain notion of redundancy, and also to improve the constructions of bases for closure-based redundancy, up to global optimality as well.

Many existing notions of redundancy discuss redundancy of a partial rule only with respect to another single partial rule; in our Section 5, we have moved beyond into the use of two partial rules. For this approach to redundancy, we believe that this last step has been undertaken for the first time here; the only other reference we are aware of, where a consideration is made of several partial rules entailing a partial rule, is the early [33], which used a much more demanding notion of redundancy in which the exact values of the confidence of the rules were both available on the premises and required in the conclusion. In our simpler context, we have shown that the following holds: for $0 < \gamma < 1/2$, there is no case of proper γ -entailment from two premises; beyond $1/2$, there are such cases, and they are fully captured in terms of set inclusion relationships between the itemsets involved. We conjecture that a more general pattern holds.

More precisely, we conjecture the following: for values of the confidence parameter $\gamma \neq 0$, such that $\frac{n-1}{n} \leq \gamma < \frac{n}{n+1}$ (where $n \geq 1$), there are partial rules that are properly entailed from n premises, partial rules themselves, but there are no proper entailments from $n+1$ or more premises. That is, intuitively, higher values of the confidence threshold correspond, successively, to the ability of using more and more partial premises. However, the combinatorics to fully characterize the case of two premises are already difficult enough for the current state of the art, and progress towards proving this conjecture requires to build intuition to much further a degree.

This may be, in fact, a way towards stronger redundancy notions and always smaller bases of association rules. We wish to be able to establish such more general methods to reach absolutely minimum-size bases with respect to general entailment, possibly depending on the value of the confidence threshold γ as per our conjecture as just stated.

We observe the following: after constructing a basis, be it either the representative rules or the \mathcal{B}_γ^* family, it is a simple matter to scan it and check for the existence of pairs of rules that generate a third rule in the basis according to Theorem 5.3: then, removing such third rules gives a smaller basis with respect to this more general entailment. However, we must say that some preliminary empirical tests suggest that this sort of entailments from two premises seems to appear in practice very infrequently, so that the check is computationally somewhat expensive compared to the scarce savings it provides for the basis size.

Now that all our contributions are in place, let us review briefly a point that we made in the Introduction regarding what is expected to be the role of the basis. The statement that association rule mining produces huge outputs, and that this is indeed a problem, not only is acknowledged in many papers but also becomes self-evident to anyone who has looked at the output of any of the association miner implementations freely accessible on the web (say [8] for one). However, we do not agree that it is *one* problem: to us, it is, in fact, *two* slightly different problems, and confusing them may lead to controversies that are easier to settle if we understand that different persons may be interested in different problems, even if they are stated similarly. Specifically, let us ask whether a huge output of an association miner is a problem for the user, who needs to receive the output of the mining process in a form that a human can afford to read and understand, or for the software that is to store all these rules, with their supports and confidences. Of course, the answer is “both”, but the solutions may not coincide.

Indeed, sophisticated conceptual advances have provided data structures to be computed from the given dataset in such a way that, within reasonable computational resource limits, they are able to give us the support and confidence of any given rule in the given dataset; maybe a good approximation is satisfactory enough, and this may allow us to obtain some efficiency advantages. The set of frequent sets, the set of frequent closures, and many other methods have been proposed for this task; see [3], [9], [10], [13], [33], [35], [36], [44], and the surveys [11] and [29].

Our approach is, rather, logical in nature, and aimed at the other variant of the problem: what rules are irredundant, in a general sense. From these, redundant rules reaching the thresholds can be found, “just as rules”. So, we formalize a situation closer to the practitioner’s process, where a confidence threshold γ is enforced beforehand and the rules with confidence at least γ are to be discussed; but we do *not* need to infer from the basis the value of the confidence of each of these other rules, because we can recompute it immediately as a quotient of two supports, found in an additional data structure that we assume kept, such as the closures lattice with the supports of each closed set.

Therefore, our bases, namely, the already-known representative rules and our new closure-based proposal \mathcal{B}_γ^* , are rather “user-oriented”: we know that all rules above the threshold can be obtained from the basis, and we know how to infer them when necessary; thus, we could, conceivably, guide (or be guided by) the user if (s)he wishes to see all the rules that can be derived from one of the rules in the basis; this user-guided exploration of the rules resulting from the mining process is alike to the “direction-setting rules” of [31], with the difference that their proposal is based on statistical considerations rather than the logic-based approach we have followed.

The advantage is that our basis is not required to provide as much information as the bases we have mentioned so far, because the notion of redundancy does not require us to be able to compute the confidence of the redundant rules. This is why we can reach an optimum size, and indeed, compared to [36] or [44], \mathcal{B}_γ^* differs because these proposals, essentially, pick all minimal generators of each antecedent, which we avoid. The difference is marginal in the conceptual sense; however the figures in practical cases may differ considerably, and the main advantage of our construction is that we can actually prove that there is no better alternative as a basis for the partial rules with respect to closure-based redundancy.

Further research may proceed along several questions. We believe that a major breakthrough in intuition is necessary to fully understand entailment among partial rules in its full generality, either as per our conjecture above or against it; variations of our definition may be worth study as well, such as removing the separate confidence parameter and requiring that the conclusion holds with a confidence at least equal to the minimum of the confidences of the premises.

Other questions are how to extend this approach to the mining of more complex dependencies [41] or of dependencies among structured objects; however, extending the development to sequences, partial orders, and trees, is not fully trivial, because, as demonstrated in [7], there are settings where the combinatorial structures may make redundant certain 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 ([20], [21]) are used?

ACKNOWLEDGEMENTS

The author is grateful to his research group at UPC and to the regular seminar attendees; also, for many observations, suggestions, comments, references, and improvements, the author gratefully acknowledges Cristina Tîrnăuică, Verónica Dahl, Tijn de Bie, Jean-François Boulicaut, the participants in seminars where the author has presented this work, the reviewers of the conference papers where most of the results of this paper were announced, and the reviewers of the present paper.

REFERENCES

- [1] C C Aggarwal, P S Yu: A New Approach to Online Generation of Association Rules. *IEEE Transactions on Knowledge and Data Engineering*, 13 (2001), 527–540. See also ICDE'98.
- [2] R Agrawal, T Imielinski, A Swami: Mining Association Rules between Sets of Items in Very Large Databases. *ACM SIGMOD 1993*, 207–216.
- [3] R Agrawal, H Mannila, R Srikant, H Toivonen, A I Verkamo: Fast Discovery of Association Rules. In: *Advances in Knowledge Discovery and Data Mining*, U Fayyad et al. (eds.), AAAI Press 1996, 307–328.
- [4] M Arias, J L Balcázar: Construction and Learnability of Canonical Horn Formulas. Submitted. Preliminary version in ALT'2009.
- [5] J L Balcázar: Minimum-Size Bases of Association Rules. *ECML-PKDD'08*, Antwerp, 86–101.
- [6] J L Balcázar: Deduction Schemes for Association Rules. *Discovery Science 2008*, 124–135.
- [7] J L Balcázar, A Bifet, A Lozano: Mining Implications from Lattices of Closed Trees. *Extraction et Gestion des Connaissances 2008*.
- [8] C Borgelt: Efficient Implementations of Apriori and Eclat. *Workshop on Frequent Itemset Mining Implementations (2003)*. See borgelt.net
- [9] J-F Boulicaut, A Bykowski, C Rigotti: Free-Sets: A Condensed Representation of Boolean Data for the Approximation of Frequency Queries. *Data Min. Knowl. Discov.* 7, 1 (2003), 5–22.
- [10] T Calders, B Goethals: Mining All Non-Derivable Frequent Itemsets. *PKDD 2002*, LNCS 2431, 74–85.

- [11] T Calders, C Rigotti, J-F Boulicaut: A Survey on Condensed Representations for Frequent Sets. *Constraint-Based Mining and Inductive Databases* 2004, 64–80.
- [12] A Ceglar, J F Roddick: Association Mining. *ACM Computing Surveys* 38 (2006).
- [13] L Cristofor, D Simovici: Generating an Informative Cover for Association Rules. ICDM 2002, 597–613.
- [14] B A Davey, H A Priestley: *Introduction to Lattices and Order*. Cambridge University Press, 1990.
- [15] R Dechter, J Pearl: Structure Identification in Relational Data. *Artificial Intelligence* 58 (1992), 237–270.
- [16] T Eiter, G Gottlob: Identifying the Minimal Transversals of a Hypergraph and Related Problems. *SIAM J. Comput.* 24 (1995), 1278–1304.
- [17] Frequent Itemset Mining Implementations Repository (<http://fimi.cs.helsinki.fi>).
- [18] A Freitas: Understanding the Crucial Differences between Classification and Discovery of Association Rules. *SIGKDD Explorations*, 2 (2000), 65–69.
- [19] B Ganter, R Wille: *Formal Concept Analysis*. Springer 1999.
- [20] G C Garriga: Statistical Strategies for Pruning All the Uninteresting Association Rules. ECAI 2004, 430–434.
- [21] L Geng, H J Hamilton: Interestingness Measures for Data Mining: A Survey. *ACM Computing Surveys* 38 (2006).
- [22] B Goethals, J Muhonen, H Toivonen: Mining Non-Derivable Association Rules. SDM 2005.
- [23] J-L Guigues, V Duquenne: Famille minimale d’implications informatives résultant d’un tableau de données binaires. *Mathématiques et Sciences Humaines* 24 (1986), 5–18.
- [24] D Gunopulos, R Khardon, H Mannila, S Saluja, H Toivonen, RS Sharma: Discovering All Most Specific Sentences. *ACM Trans. Database Syst.*, 28 (2003), 140–174.
- [25] M Hahsler: Annotated Bibliography on Association Rule Mining (http://michael.hahsler.net/research/association_rules).
- [26] R Khardon, D Roth: Reasoning with Models. *Artificial Intelligence* 87 (1996), 187–213.
- [27] M Kryszkiewicz: Representative Association Rules. Pacific-Asia KDD Conference, PAKDD’98, LNCS 1394, 198–209.
- [28] M Kryszkiewicz: Fast Discovery of Representative Association Rules. RSCTC, 1998, 214–221.
- [29] M Kryszkiewicz: Concise Representations of Association Rules. Pattern Detection and Discovery 2002 (LNCS 2447), 187–203.
- [30] G Li, H Hamilton: Basic Association Rules. SDM 2004.
- [31] B Liu, W Hsu, Y Ma: Pruning and Summarizing the Discovered Associations. KDD 1999, 125–134.
- [32] B Liu, M Hu, W Hsu: Multi-Level Organization and Summarization of the Discovered Rules. KDD 2000, 208–217.
- [33] M Luxenburger: Implications Partielles dans un Contexte. *Mathématiques et Sciences Humaines* 29 (1991), 35–55.
- [34] N Megiddo, R Srikant: Discovering Predictive Association Rules. KDD 1998, 274–278
- [35] J Muhonen, H Toivonen: Closed Non-Derivable Itemsets. PKDD 2006, 601–608.
- [36] N Pasquier, R Taouil, Y Bastide, G Stumme, L Lakhal: Generating a Condensed Representation for Association Rules. *Journal of Intelligent Information Systems* 24 (2005), 29–60.
- [37] J L Pfaltz, C M Taylor: Scientific Discovery through Iterative Transformations of Concept Lattices. Workshop on Discrete Mathematics and Data Mining at SDM 2002, 65–74.
- [38] V Phan-Luong: The Representative Basis for Association Rules. ICDM 2001, 639–640.
- [39] V Phan-Luong: The Closed Keys Base of Frequent Itemsets. DaWaK 2002, 181–190.
- [40] J Saquer, J S Deogun: Using Closed Itemsets for Discovering Representative Association Rules. IS-MIS 2000, LNAI 1932, 495–504.
- [41] D A Simovici, D Cristofor, L Cristofor: Mining Purity Dependencies in Databases. Extraction et Gestion des Connaissances EGC 2002, 257–268.
- [42] A Tuzhilin, B Liu: Querying Multiple Sets of Discovered Rules. KDD 2002, 52–60.
- [43] M Wild: A Theory of Finite Closure Spaces Based on Implications. *Advances in Mathematics* 108 (1994), 118–139.
- [44] M Zaki: Mining Non-Redundant Association Rules. *Data Mining and Knowledge Discovery* 9 (2004), 223–248.
- [45] M Zaki, M Ogihara: Theoretical Foundations of Association Rules. Workshop on research issues in DMKD (1998).