Deduction Schemes for Association Rules

José L. Balcázar[1]

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

Abstract. Several notions of redundancy exist for Association Rules. Often, these notions take the form "any dataset in which this first rule holds must obey also that second rule, therefore the second is redundant"; if we see datasets as interpretations (or models) in the logical sense, this is a form of logical entailment. In many logics, entailment has a syntactic counterpart in the form of a deduction calculus. We provide such a deduction calculus for existing notions of redundancy; then, we consider a very general notion of entailment, where a confidence threshold is fixed and several rules can act as simultaneous premises, and identify exactly the cases where a partial rule follows from two partial rules; we also give a deduction calculus for this setting.

Keywords: association rules, redundancy, deductive calculus

1 Motivation and Related Work

Data mining involves a wide spectrum of techniques; among them, Association Rule Mining is a prominent conceptual tool and, possibly, a cornerstone notion of the field, if there is one. Indeed, association rules are both among the most widely studied topic in data mining research and among the most widely employed data mining techniques in actual applications. Practitioners have reported impressive success stories (failures being less prone to receive publicity); researchers have provided a wealth of algorithms to compute diverse variants of association rules on datasets of diverse characteristics, and there are many extensions into similar notions for complex data. The volume of literature about the topic is daunting. A recent survey is [5] but additional materials appear in `http://wwwai.wu-wien.ac.at/~hahsler/research/association_rules/`, for instance, at the time of writing.

Implications, that is, association rules that hold in 100% of the cases, had been studied before in the research area of closure spaces (see, for instance, [9] and [19]). Implications can be seen also as conjunctions of definite Horn clauses, and the closure under intersection property that characterizes closures spaces corresponds to the fact, well-known in logic and knowledge representation, that Horn theories are exactly those closed under bitwise intersection of propositional models (see e.g. [11]). Thus, as a form of knowledge gathered from a dataset, implications have several advantages: explicit or implicit correspondence with Horn logic, therefore a tight parallel with functional dependencies and a clear, hardly disputable notion of redundancy that can be defined equivalently both in semantic terms and through a syntactic calculus. Specifically, in semantic terms, an implication $X \to Y$ is entailed from a set of implications $\mathcal{R}$ if every dataset in which all the implications of $\mathcal{R}$ hold must also satisfy $X \to Y$; and, syntactically, it is known that this happens if and only if $X \to Y$ is derivable from $\mathcal{R}$ via the Armstrong axiom schemes [17], namely, Reflexivity ($X \to X$), Augmentation (if $X \to Y$ and $X' \to Y'$ then $XX' \to YY'$, where juxtaposition denotes union) and Transitivity (if $X \to Y$ and $Y \to Z$ then $X \to Z$).

However, absolute implication analysis may become too limited for many application tasks. Already in [15] we find a proposal to consider partial rules, defined in relation to their so-called-there "precision", that is, the notion of intensity of implication now widely called "confidence": for a given rule $X \to Y$, the ratio of how often X and Y are seen together to how often X is seen.

Such search for implications or for partial rules was not used on really large datasets until the introduction of the notion of support bound, that is, an absolute threshold on how often the itemsets under analysis appear in the dataset. The idea of restricting the exploration for association rules to frequent itemsets, with respect to a support threshold, gave rise to the most widely discussed and applied algorithm, Apriori [2], and to an intense research activity.

A well-known difficulty in applied association rule mining lies in that, on large datasets, and for sensible settings of the confidence and support thresholds, huge amounts of association rules are often obtained, much beyond what any user of the data mining process may be expected to look at. Therefore, one research topic that has been worthy of attention is the identification of patterns that indicate redundancy of rules, and ways to avoid that redundancy [1], [6], [12], [13], [15], [16], [18]; see also section 6 of [5] and the references therein. All these definitions of redundancy are given either set-theoretically, in the sense of inclusions among sets of attributes, or in a more general way, by resorting to a confidence inequality: a rule is redundant with respect to another if it has at least the same confidence of the latter *for every dataset*.

By analogy to the case of implications, it is natural to raise the question of whether a deductive calculus for these different notions of redundancy among partial rules can be designed. (We will keep this terminology throughout the paper: implications are association rules of confidence 1, whereas partial rules are those having a confidence below 1.) The Armstrong axiom schemes that play such a role for implications are, in fact, no longer adequate: Reflexivity does hold for partial association rules, but Augmentation does not hold at all, whereas Transitivity takes a different form that affects the confidence of the rules: if the rule $A \to B$ (or $A \to AB$, which is equivalent) and the rule $B \to C$ both hold with confidence γ, we still know nothing about the confidence of $A \to C$; even the fact that both $A \to AB$ and $AB \to C$ hold with confidence γ only gives us a confidence of γ^2 for $A \to C$. Regarding Augmentation, enlarging the antecedent of a rule of confidence 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 Y, but it only comes with Z when Y is not present; then the confidence of $X \to Y$ may be high whereas the confidence of $XZ \to Y$ may be null. Similarly, a rule with several items in the consequent is *not* equivalent to the conjunction of the Horn-style rules with the same antecedent and each item of the consequent separately: if we look only into rules with singletons as consequents (as in the "basic association rules" of [14] or in the useful `apriori` implementation of Borgelt available on the web [4]) we are almost certain to lose information. Indeed, if the confidence of $X \to YZ$ is high, it means that Y and Z appear together in most of the transactions having X; 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 to hold.

We have provided in a recent contribution [3] an analysis of several particular notions of redundancy in the literature; here, we progress further along that study. First, we provide a deductive calculus that characterizes these notions of redundancy. In a way similar to the Armstrong axiom schemes, we give deduction schemes such that exactly the rules that are redundant can be deduced through these schemes. Then, we depart from most proposed notions of redundancy by considering the possibility that a rule is redundant with respect to a set of rules, instead of a single one; for that case we have our main contribution here. 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: fix a confidence threshold, and consider only rules whose confidence is above it; alternatively, an equivalent view would be that the confidence of all our conclusions should be at least the same as the minimum of the confidences of the premises. For instance, with items A, B, C, and D, assume that the confidence of the rules $A \to BC$ and $A \to BD$ is above γ in a dataset $\mathcal{D}$. What can be said, then, about the confidence of the rule $ACD \to B$ in $\mathcal{D}$? For instance, could one construct a dataset where the rules $A \to BC$ and $A \to BD$ hold with 65% confidence and, simultaneously, rule $ACD \to B$ falls below the same confidence threshold? The answer is counterintuitive: it is, in fact, impossible, and we will provide a full answer, characterizing exactly inference from two partial rules, as main result of this paper, and a corresponding deduction scheme extending our calculus. In fact, the existing notions of redundancy correspond exactly to entailment among association rules just for a specific confidence interval, and our results suggest the possibility of a pattern, where further values of the threshold would correspond, successively, to the ability of using three partial premises, four, and so on. However, to attain such a result, further efforts are still necessary.

2 Preliminaries

A dataset $\mathcal{D}$ is given; it consists of transactions, each of which is an itemset labeled by a unique transaction identifier. The identifiers allow for many transactions sharing the same itemset. Upper-case, often subscripted letters from the end of the alphabet, like X_1 or Y_0, denote itemsets. Juxtaposition denotes union of itemsets, as in XY; and $Z \subset X$ denotes proper subsets. For a transaction t, we denote $t \models X$ the fact that X is a subset of the itemset corresponding to t.

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

We immediately obtain by standard means (see, for instance, [9] or [18]) 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, a closure operator. This notion will be reviewed in more detail later on.

Association rules are pairs of itemsets, denoted as $X \to Y$ for itemsets X and Y. Intuitively, they express that Y occurs particularly often among the transactions in which X occurs. More precisely, the confidence $c_{\mathcal{D}}(X \to Y)$ of an association rule $X \to Y$ in a dataset $\mathcal{D}$ is $\frac{s(XY)}{s(X)}$, that is, the ratio by which transactions having X have also Y; or, again, the observed empirical approximation to a conditional probability of Y given X. As with support, often we drop the subindex $\mathcal{D}$. This view suggests a form of correlation that, in many applications, is interpreted implicitly as a form of causality (which, however, is not guaranteed in any formal way; see the interesting discussion in [8]).

We resort to the convention that, if $s(X) = 0$ (which implies $s(XY) = 0$) we redefine the undefined confidence as 1, since the intuitive expression "all transactions having X do have also Y" becomes vacuously true. Also, it is immediate to check that $c_{\mathcal{D}}(X \to Y) = c_{\mathcal{D}}(X \to XY) = c_{\mathcal{D}}(X \to X'Y)$ for any subset $X' \subseteq X$. When two rules have the same left hand side, and the same union of left and right hand sides, we say that they are *equivalent by reflexivity*. Clearly their supports and confidences will always coincide.

We discuss briefly now the following natural notion of redundancy:

Definition 1. *$X_0 \to Y_0$ is plainly redundant with respect to $X_1 \to Y_1$ if the confidence of $X_0 \to Y_0$ is larger than or equal to the confidence of $X_1 \to Y_1$, whatever the dataset.*

That is: in that case, if a data mining process with confidence threshold γ provides both rules as output, $X_0 \to Y_0$ and $X_1 \to Y_1$, then rule $X_0 \to Y_0$ is uniformative, and can be ignored, because the fact that its confidence is at least γ is already guaranteed without computing it. Of course, such a redundancy will happen only if the various itemsets involved, such as X_0 and X_1, have some correlation. Several cases of redundancy have been identified already in the literature, and compared in [3]. We briefly review some related results.

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

3. [12] *Rule* $X_1 \to Y_1$ *covers rule* $X_0 \to Y_0$ *if* $X_1 \subseteq X_0$ *and* $X_0Y_0 \subseteq X_1Y_1$*.*

The original definition of cover in [12] is different but the same reference proves the equivalence with the formulation we are using here.

In these three cases, it is not difficult to see that the confidence and support of $X_0 \to Y_0$ is at least that of $X_1 \to Y_1$ ([1], [12]). Note that, in principle, there could possibly be many other ways of a rule being redundant with respect to another beyond covering, simple, and strict redundancies. Simple redundancy relates rules obtained from the same (frequent) set $X_0Y_0Z_0$. Strict redundancy focuses, instead, on rules extracted from two different (frequent) itemsets, say X_0Y_0 where X_0 will be considered as antecedent, and X_1Y_1, where X_1 will be antecedent, and under the conditions that $X_1 \subseteq X_0$ and $X_0Y_0 \subset X_1Y_1$ (the case $X_0Y_0 = X_1Y_1$ is already covered by simple redundancy). Our previous work has contributed the following result:

Theorem 1. [3] *Consider any two rules* $X_0 \to Y_0$ *and* $X_1 \to Y_1$ *where* $Y_0 \not\subseteq X_0$*. The following are equivalent:*

1. $X_1 \to Y_1$ *covers* $X_0 \to Y_0$*;*
2. *rule* $X_0 \to Y_0$ *is either simply redundant or strictly redundant with respect to* $X_1 \to Y_1$*, or they are equivalent by reflexivity;*
3. *both the confidence* **and** *the support of* $X_0 \to Y_0$ *are larger than or equal to those of* $X_1 \to Y_1$*, whatever the dataset;*
4. *rule* $X_0 \to Y_0$ *is plainly redundant with respect to* $X_1 \to Y_1$*.*

That is, the additional consideration of the support bound in formulation 3 (due to [1] as well) does not make the notion more restrictive, whereas the notion of covering catches all possible nontrivial situations of plain redundancy. The equivalence of the first two statements is immediate. Note that rules with $Y_0 \subseteq X_0$ have confidence 1: they state just reflexivity, and are uninformative.

A major application of the notion of redundancy is the construction of "bases": sets of rules that make redundant all the remaining rules mined. Our previous paper [3] has shown that the existing techniques for constructing bases with respect to these equivalent notions of redundancy do attain the minimum possible size of a basis. However, further reduction of the bases is still desirable, and the only way to obtain it is through some stronger notion of redundancy.

One such stronger notion existing in the literature, which indeed allows for smaller bases (most of the times) relies on handling separately implications from partial rules. Indeed, implications can be summarized better, because they allow for Transitivity and Augmentation to apply in order to find redundancies; moreover, they can be combined in a certain form of transitivity with a partial rule of confidence, say, γ to give rules of confidence at least γ. The best way to handle them is through a closure operator ([7], [9], [16], [18]). Specifically, given a dataset $\mathcal{D}$, the closure operator associated to $\mathcal{D}$ maps each itemset X to the largest itemset $\overline{X} \supseteq X$ that has the same support as X in $\mathcal{D}$; it can be defined in several alternative ways. A set is closed if it coincides with its closure. When

$\overline{X} = Y$ we also say that X is a generator of Y. Our definition gives directly that always $s(X) = s(\overline{X})$. We will make liberal use of this fact, which is easy to check also with other definitions of the closure operator, as stated in [16], [18], and others. Implications are intimately related to this closure operator: $c(X \to Y) = 1$ if and only if $Y \subseteq \overline{X}$. Several quite good algorithms exist to find the closed sets and their supports. In the literature, there are proposals of basis constructions out of closed sets with respect to a notion of redundancy based on closures, a natural generalization of equivalence by reflexivity that works as follows ([18], see also section 4 in [16]): given a dataset and a closure operator corresponding to implications that have confidence 1 in the dataset, two partial rules $X_0 \to Y_0$ and $X_1 \to Y_1$ such that $\overline{X_0} = \overline{X_1}$ and $\overline{X_0 Y_0} = \overline{X_1 Y_1}$ turn out to be equivalent in terms of support and confidence; the reason is that $s(X_0) = s(\overline{X_0}) = s(\overline{X_1}) = s(X_1)$, and $s(X_0 Y_0) = s(\overline{X_0 Y_0}) = s(\overline{X_1 Y_1}) = s(X_1 Y_1)$, by the property stated above that always $s(X) = s(\overline{X})$.

Let $\mathcal{B}$ be the set of implications, of confidence 1, in the dataset $\mathcal{D}$; alternatively, $\mathcal{B}$ can be the basis already known for implications in a dataset [7]. From here on, we require $0 < \gamma < 1$, leaving the rules of confidence 1 to be handled from $\mathcal{B}$. Our previous work has contributed the following property, along the same lines as Theorem 1:

Theorem 2. [3] *Let $\mathcal{B}$ be a set of implications. Let $X_2 \to Y_2$ be a rule not implied by $\mathcal{B}$, that is, where $Y_2 \not\subseteq \overline{X_2}$. Then, the following are equivalent:*

1. *$X_1 \subseteq \overline{X_2}$ and $X_2 Y_2 \subseteq \overline{X_1 Y_1}$*
2. *Every dataset $\mathcal{D}$ in which all the rules in $\mathcal{B}$ hold with confidence 1 gives $c_{\mathcal{D}}(X_2 \to Y_2) \geq c_{\mathcal{D}}(X_1 \to Y_1)$.*

In either case we say that rule $X_2 \to Y_2$ has closure-based redundancy relative to $\mathcal{B}$ with respect to rule $X_1 \to Y_1$.

3 Deduction Schemes for Redundancy

Redundancy is, in principle, more restrictive than entailment: so far, in the literature, redundancy has been taken mostly to signify a relationship between two association rules, as just described. We start our discussion of deduction systems along the same lines.

We give now a calculus consisting 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.

(rR) $\quad \dfrac{X \to Y, \qquad Z \subseteq Y}{X \to Z}$

(rA) $\quad \dfrac{X \to Y}{X \to XY}$

(ℓA) $\quad \dfrac{X \to YZ}{XY \to Z}$

We also allow always to state trivial rules: $\overline{X \to \emptyset}$, which, combined with (rA) and (rR), allows us to infer $X \to Y$ whenever $Y \subseteq X$.

Scheme (ℓA) is exactly simple redundancy from Definition 2. The Reduction Scheme (rR) allows us to "lose" information and find inequivalent rules (whose confidence may be larger).

It is not difficult to see that the calculus is sound, that is, for every dataset, the confidence of the rule below the line in any one of the three deduction schemes is, at least, the same as the confidence of the rule above the line. For the scheme (ℓA), this claim is Theorem 4.1 in [1]: correctness of simple redundancy. Likewise, soundness for (rR) it is the next result in [1], Theorem 4.2. The soundness of (rA) follows from equivalence by reflexivity. Also, trivial rules with empty right hand side are always sound.

Of course, this extends to chained applications of these schemes. In fact, if we start with a rule $X_1 \to Y_1$, and keep applying these three inference schemes to obtain new association rules, the rules we obtain are all plainly redundant with respect to $X_1 \to Y_1$.

The interesting property of these schemes, and our first contribution, is that the converse also holds; that is: whenever two rules are related by plain redundancy, it is always possible to prove it using just those inference schemes.

Theorem 3. *Rule $X_0 \to Y_0$ is plainly redundant with respect to rule $X_1 \to Y_1$ if and only if $X_0 \to Y_0$ can be derived from $X_1 \to Y_1$ by repeated application of the inference schemes (rR), (rA), and (ℓA).*

Proof. That all rules derived are plainly redundant has just been argued above. For the converse, assume that rule $X_0 \to Y_0$ is plainly redundant with respect to rule $X_1 \to Y_1$. By Theorem 1, we know that this implies that $X_1 \to Y_1$ covers $X_0 \to Y_0$, that is, by Definition 2, $X_1 \subseteq X_0$ and $X_0Y_0 \subseteq X_1Y_1$. Now, to infer $X_0 \to Y_0$ from $X_1 \to Y_1$, we chain up applications of our schemes as follows:

$$X_1 \to Y_1 \vdash_{(rA)} X_1 \to X_1Y_1 \vdash_{(rR)} X_1 \to X_0Y_0 \vdash_{(\ell A)} X_0 \to Y_0$$

where the second step makes use of the inclusion $X_0Y_0 \subseteq X_1Y_1$, 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 rule indicated as a subscript. $\square$

3.1 Calculus for Closure-Based Redundancy

The calculus just given is not appropriate to handle closure-based redundancy, because it does not contemplate any form of Transitivity. The stronger calculus we provide now is sound and complete with respect to closure-based redundancy. We will use two different symbols for rules: we will keep $X_0 \to Y_0$ to denote association rules, of which we will lower-bound the confidence, and we will use the notation $X_0 \Rightarrow Y_0$ to denote implications. 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.

(rA) $\frac{X \to Y, \qquad X \Rightarrow Z}{X \to YZ}$

(rI) $\frac{X \to Y, \qquad Y \Rightarrow Z}{X \to Z}$

(ℓA) $\frac{X \to YZ}{XY \to Z}$

(ℓI) $\frac{X \to Y, \qquad Z \subseteq X, \qquad Z \Rightarrow X}{Z \to Y}$

Again we allow as well to state directly rules with empty right hand side: $\overline{X \to \emptyset}$. Note that this opens the door to using (rA) with an empty Y, and this allows us to transform an implication into the corresponding partial rule. Also, (ℓA) could be stated equivalently with $XY \to YZ$ below the line, by (rA).

The connection with the previous, simpler calculus should be easy to understand: first, observe that the (ℓA) rules are identical. Now, if implications are not considered separately, 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, (ℓI) becomes fully trivial since $X \subseteq Z$ makes $X = Z$, and the partial rules above and below the line would coincide.

In the remaining of this section, we denote as $\mathcal{B} \cup \{X \to Y\} \vdash X' \to Y'$ the fact that, in the presence of the implications in the set $\mathcal{B}$, rule $X' \to Y'$ can be derived from rule $X \to Y$ using zero or more applications of the four deduction schemes.

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. *Let $\mathcal{B}$ consist of implications. Then, $\mathcal{B} \cup \{X_1 \to Y_1\} \vdash X_2 \to Y_2$ if and only if rule $X_2 \to Y_2$ has closure-based redundancy relative to $\mathcal{B}$ with respect to rule $X_1 \to Y_1$.*

Proof (Sketch). Proofs given (within a slightly different framework) in [18] provide directly the soundness of (rI) and the soundness of a combination of (rA) with (ℓI) that can be extended quite easily to obtain soundness of our four schemes. To prove completeness, we must see that all redundant rules can be derived. We assume the closure-based redundancy of $X_2 \to Y_2$ and resort to Theorem 2: we know that the inclusions $X_1 \subseteq \overline{X_2}$ and $X_2Y_2 \subseteq \overline{X_1Y_1}$ must hold. From the second inclusion, and the properties of the closure operator, we have that $\overline{X_2}Y_2 \subseteq \overline{X_1Y_1}$.

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

$X_1 \to Y_1 \vdash_{(rA)} X_1 \to X_1Y_1 \vdash_{(rI)} X_1 \to \overline{X_2}Y_2 \vdash_{(\ell A)} \overline{X_2} \to Y_2 \vdash_{(\ell I)} X_2 \to Y_2$

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

4 Closure-Based Entailment

We move on towards the main contribution of this paper. The following question naturally arises: all these notions of redundancy only 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? Along this section, we are after a calculus for all partial rules whose confidence is above some fixed but arbitrary threshold γ. We will fully answer this question by, first, characterizing precisely the case where a partial rule follows from exactly two partial rules, the simplest case where our previous calculus becomes incomplete; and, second, proving that a sound and complete calculus can be constructed by adding one extra rule that allows us to conclude a consequent partial rule from two antecedent partial rules. We present the whole setting on top of closure-based redundancy, but an analogous development can be made on top of plain redundancy. We consider the following definition:

Definition 3. *Given a set $\mathcal{B}$ of implications, and a set $\mathcal{R}$ of partial rules, rule $X_0 \to Y_0$ is γ-redundant with respect to them, $\mathcal{B} \cup \mathcal{R} \models_\gamma X_0 \to Y_0$, 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 $X_0 \to Y_0$ with confidence at least γ.*

As an interesting example that does not need the presence of implications, consider the following fact, mentioned in the Introduction (the analogous statement for $\gamma < 1/2$ does not hold, as discussed below):

Proposition 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 $A \to BC$ and $A \to BD$ is above γ in dataset $\mathcal{D}$. Then, the confidence of the rule $ACD \to B$ in $\mathcal{D}$ is also above γ.*

We omit the proof, since it is just the simplest particular case of our main result in the paper, which clarifies exactly these situations, and reads as follows:

Theorem 5. *Let $\mathcal{B}$ be a set of implications, and let $1/2 \leq \gamma < 1$. Then, $\mathcal{B} \cup \{X_1 \to Y_1, X_2 \to Y_2\} \models_\gamma X_0 \to Y_0$ if and only if either:*

1. $Y_0 \subseteq \overline{X_0}$, *or*
2. $\mathcal{B} \cup \{X_1 \to Y_1\} \models_\gamma X_0 \to Y_0$, *or*
3. $\mathcal{B} \cup \{X_2 \to Y_2\} \models_\gamma X_0 \to Y_0$, *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 (Sketch). Let us discuss first the leftwards implication. In case (1) the consequent rule holds trivially. Clearly cases (2) and (3) also give the entailment, though in a somehow "improper" way. For case (4), we must argue that, if all the seven conditions hold, then the entailment happens. Thus, fix any dataset $\mathcal{D}$ where the confidences of the antecedent 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 $X_0 \to Y_0$ in $\mathcal{D}$ is also at least γ. Consider the following four sets of transactions from $\mathcal{D}$:

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

and let a, b, c, and d be the respective cardinalities. Using condition (v), it can be seen that all four sets are mutually 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 $s(X_2) \geq a + b + c + d$.

If we split all the tuples that satisfy X_1Y_1 into two sets, those that additionally satisfy X_0, and those that don't, using conditions (i) and (vi) it can be argued that $s(X_1Y_1) \leq a + c$ and, symmetrically, resorting to (ii) and (vii), $s(X_2Y_2) \leq a + d$.

Thus we can write the following inequations:

$$a + c \geq s(X_1Y_1) \geq \gamma s(X_1) \geq \gamma(a + b + c + d)$$

$$a + d \geq s(X_2Y_2) \geq \gamma s(X_2) \geq \gamma(a + b + c + d)$$

Adding them up, using $\gamma \geq \frac{1}{2}$, and simplifying, we get $a \geq \gamma(a + b)$, so that $c(X_0 \to Y_0) = s(X_0Y_0)/s(X_0) = a/(a + b) \geq \gamma$ as was to be shown.

For the converse, we first point out that the bound $\gamma \geq \frac{1}{2}$ is not necessary for this part. The proof goes on by arguing 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, omitted here for lack of space. □

4.1 Extending the calculus

We work now towards a rule form. Let us say that an entailment is proper if the consequent follows from the given set of antecedents but does not follow from any proper subset thereof. We propose the following additional rule:

$$\frac{X_1 \rightarrow Y_1, \quad X_2 \rightarrow Y_2, \quad X_1 Y_1 \Rightarrow X_2, \quad X_2 Y_2 \Rightarrow X_1, \quad X_1 Y_1 X_2 Y_2 \Rightarrow Z}{X_1 X_2 Z \rightarrow \overline{X_1 Y_1 Z} \cap \overline{X_2 Y_2 Z}}$$

and state the following properties:

Theorem 6. *Given a threshold γ and a set $\mathcal{B}$ of implications,*

1. *this deduction scheme is sound, and*
2. *together with the deduction schemes in Subsection 3.1, it gives a calculus complete with respect to all entailments with two partial rules in the antecedent for $\gamma \geq 1/2$.*

Proof (sketch). This follows easily from Theorem 5, 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 hold. 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, otherwise just apply this new rule with $Z = \overline{X_0}$ to get $\overline{X_0} \rightarrow \overline{X_0 Y_1} \cap \overline{X_0 Y_2}$ and apply (ℓI) and (rI) to obtain $X_0 \rightarrow Y_0$. □

5 Conclusions

Our study here belongs to a larger program of research on the fundamentals of the Logic of Association Rules. We have described sound and complete variants of a deductive calculus for redundancy and entailment notions defined in terms of models, that is, datasets that assign a confidence value to each partial rule. The notions of redundancy correspond to already existing proposals, which discuss redundancy of a partial rule only with respect to another single partial rule; in our Theorem 5 we have moved beyond into the use of two partial rules. We believe this last step has been undertaken for the first time here since the early attempts of [15].

On the basis of our results as described here, it turns out that the following holds: for $0 < \gamma < 1/2$, there is γ-entailment if and only if either of cases 1, 2, or 3 in Theorem 5 is true; existing notions of redundancy are, therefore, fully appropriate for confidence below 1/2, but insufficient beyond. For larger confidence thresholds, our preliminary analyses are very suggestive of a general pattern, which we expect to be able to develop and apply to arbitrary values of γ: further values of the threshold correspond, successively, to the ability of using more and more partial premises. Namely, up to two partial rules can be used as premise, but not three, if $\gamma < 2/3$; up to three, but not four, if $\gamma < 3/4$; and so on. However, the combinatorics to fully characterize the case of two premises are already difficult enough for the current state of the art, and progressing along this line requires to build intuition to much further a degree. There remains to study also a comparison with other "semantic" redundancy schemes based on the actual values of the supports, such as those in [10] or, along a different track, [6].

Finally, we wish to discuss also the basis constructed in [3]: there we proved that the size of that basis is minimum with respect to closure-based redundancy.

It is worth to point out that our initial proofs of the results reported both here and there were much more involved, and it was thanks to the calculus described here that the minimum-size bases described in [3] were found, although further work has allowed us to simplify the analysis for expository purposes. It is also interesting to observe that, given the minimal basis described there, one can scan it to check the existence of pairs of rules that generate a third rule in the basis according to Theorem 5: then, removing these third rules gives a smaller basis with respect to general entailment. We expect to be able to establish a more general similar mechanism depending of the value of the threshold γ to reach absolutely minimum-size bases with respect to general entailment.

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.
2. 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, 307–328.
3. J L Balcázar: Minimum-Size bases of Association Rules. To appear in: PKDD'08, Antwerp.
4. C Borgelt. Efficient Implementations of Apriori and Eclat Workshop on Frequent Itemset Mining Implementations (2003).
5. A Ceglar, J F Roddick: Association mining. *ACM Computing Surveys* 38 (2006).
6. L Cristofor, D Simovici: Generating an informative cover for association rules. ICDM 2002, 597–613.
7. J-L Guigues, V Duquenne: Famille minimale d'implications informatives résultant d'un tableau de données binaires. *Math. et Sciences Humaines* 24 (1986), 5–18.
8. A Freitas: Understanding the crucial differences between classification and discovery of association rules. *SIGKDD Explorations*, 2 (2000), 65–69.
9. B Ganter, R Wille: *Formal Concept Analysis*. Springer 1999.
10. B Goethals, J Muhonen, H Toivonen: Mining non-derivable association rules. SDM 2005.
11. R Khardon, D Roth: Reasoning with models. *Artif. Intell.* 87 (1996), 187–213.
12. M Kryszkiewicz: Representative Association Rules. Pacific-Asia KDD Conference, PAKDD'98, LNCS 1394, 198–209.
13. M Kryszkiewicz: Representative Association Rules and Minimum Condition Maximum Consequence Association Rules. PKDD'98, LNCS 1510, 361–369.
14. G Li, H Hamilton: Basic association rules. SDM 2004.
15. M Luxenburger: Implications partielles dans un contexte. *Mathématiques et Sciences Humaines* 29 (1991), 35–55.
16. N Pasquier, R Taouil, Y Bastide, G Stumme, L Lakhal: Generating a condensed representation for association rules. *J. Intell. Inform. Sys.* 24 (2005), 29–60.
17. J Ullman, J Widom: A First Course in Database Systems. 1997.
18. M Zaki: Mining non-redundant association rules. *Data Mining and Knowledge Discovery* 9 (2004), 223–248.
19. M Zaki, M Ogihara: Theoretical foundations of association rules. DMKD Workshop on research issues in DMKD (1998).