Entailment and Basis Optimality in Confidence-Bounded Association Rules*

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

ABSTRACT

We study the set of association rules which reach a certain confidence threshold in a given dataset. By splitting them into exact rules, that is, rules of confidence 1 characterized in a standard way by a closure operator, and partial rules, of confidence less than 1, we show how to obtain a basis of the partial rules (a nonredundant subset of these rules such that all the true partial rules can be derived from them) of absolutely minimum size with respect to a natural notion of semantic entailment. We develop this result from characterizations of the entailment property. Then, we propose and analyze an extension of this notion of entailment, identifying exactly the cases where a partial association rule is entailed jointly by two partial association rules, again in the presence of a closure operator capturing exact rules.

1. INTRODUCTION

Few, if any, data mining tasks have the relative importance within that field of research as association rule mining. Since the publication of the first proposal [2], many algorithms for this task have been designed. Currently, the amount of knowledge related to association rules has grown to the extent that creating complete surveys or websites from where to maintain pointers to literature on association rules become daunting tasks; a recent survey is [6] but, for instance, at the time of writing, additional materials are to be found in `http://wwwai.wu-wien.ac.at/~hahsler/research/association_rules/`). A deterministic notion of association rule, in the form of exact implication, had been studied long before in the research area of closure spaces ([8], [10], [21]).

Among many studies in this topic, quite a few focus on the problem that, on large datasets, and for sensible settings of the parameters, huge amounts of association rules are obtained, much beyond what any user of the data mining process may be expected to look at. A natural question is that of finding a smallish set of association rules, called a "basis", from which all the mined rules can be somehow inferred. Our work here adds to this line of research.

We assume that a dataset $\mathcal{D}$ is given on which a rule mining process has been run. The dataset consists of transactions, each itself a set of items (or attributes) labeled by a unique transaction identifier. From the given dataset we obtain a notion of support of an itemset: the cardinality of the set of transactions that include it; sometimes, abusing language, we also refer to that set of transactions itself as support. The identification of the parameter specified by a support bound [2], [3] made association rules widely applicable. We immediately obtain by standard means (see, for instance, [10] or [22]) 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. Several quite good algorithms exist to find the closed sets and their supports. Like some previous works, we employ the closure operator in the analysis of association rules.

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, each such rule has a confidence associated; the confidence of that association rule is the ratio of the joint support of X and Y, to the support of X: the ratio by which transactions having X have also Y; or again the observed empirical approximation to a conditional probability of Y given X. This view suggests a form of correlation that allows for statistical analyses (see for one [15]), and which, 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 [9]).

Implications, also known as exact association rules or also deterministic association rules in the literature, are association rules with confidence 1; that is, in all the cases where X occurs in the dataset, Y occurs as well and, thus, the support of XY and the support of X do coincide. As a form of knowledge gathered from the dataset, implications have several advantages: a clear notion of redundancy, explicit or implicit correspondence with Horn logic, therefore a tight parallel with functional dependencies, and thus a clear and well-known calculus through the Armstrong rules. Several ways have been proposed in the literature to construct bases from datasets, that is, sets of exact rules from where precisely all the exact rules true in the dataset can be derived,

*This work is supported in part by MCYT TIC2004-07925-C03-02 (Trangram), and by the IST Programme of the European Community, under the DELIS Project IST-001907 and the PASCAL Network of Excellence, IST-2002-506778

and some of them discuss absolute or relative minimality properties ([8], [10], [19], [20], [21]).

Often, however, it is inappropriate to require full confidence in the analysis of real world datasets. They might contain keying or transmission errors in some transactions, or these might come from different large fractions of the population, with different characteristics. Thus, absolute implication analysis becomes too limited for many practical cases. The first attempts at "partial rules", defined in relation to a confidence threshold (corresponding to a lower bound on the empirical conditional probability) were proposed in [16], where a first, rather preliminary solution was proposed and the problem of minimizing the size of the basis was raised; better solutions appeared later in [19] and [22], among others. The additional notion of support bound made the approach applicable to large datasets, and gave rise to the most widely discussed and applied algorithm, Apriori [3]. Further explanations appear below, and additional discussion is to be found in [23]. Instead of confidence, a large variety of ways to measure intensity of implication when the rules are not exact have been proposed; most of them are surveyed in [12] (see also [11]).

Thus, partial rules with confidence lower than 1 are to be considered. However, for partial rules, specific natural inference schemes are not valid anymore: for instance, enlarging the antecedent of a rule of confidence γ may give a rule with much smaller confidence (even zero); that is, the standard augmentation of Armstrong implications is not valid. Applying transitivity between two rules of confidence γ may give as well a resulting rule of confidence less than γ (only γ^2 is guaranteed). Additionally, a rule with several items in the consequent is *not* equivalent to the conjunction of the Horn-style rules with the same antecedent and each item of the consequent separately.

There arises the question of how to choose partial rules so that the confidences of other rules can be inferred. Luxemburger [16], in a context of Formal Concepts, proposes a calculus based on the unicity of the solution of a system of nonlinear inequations (of which we know of no further study) and a basis from where confidences of all remaining rules can be derived; but it is not known how to reach a minimal basis (the minimality problem is explicitly posed by Luxemburger, and also by Zaki in [22]). Many later works have adopted this approach for partial association rules, and have defined notions of covering and redundancy in order to avoid redundant rules. Minimum size is not yet guaranteed. Moreover, the very notion of redundancy can be (and has actually been) formalized in several ways.

We progress along this direction, by adopting a somewhat different setting. First, as is done already in [16] and several later works, we handle separately implications from partial rules; but, then, instead of working as in [16] in a setting where all rules, with their respective confidences, are to be inferred, we formalize a situation closer to the practitioner's process, where a confidence threshold γ is fixed beforehand and the rules with confidence at least γ are to be treated (a support threshold is usually also in force but we choose to study its effect separately).

Thus, instead of considering a rule redundant when its confidence can be somehow inferred from others, we take the slightly different view that it is redundant when "the fact that its confidence is above γ can be inferred from others"; also, these other rules are known to have confidence about γ, but we refrain from using the actual confidence value (even though it is available to show to the user upon request).

Our setting is, therefore, logical in nature, in that only implications between sets of attributes are manipulated. There are several alternative settings where the manipulation of the rules to find other rules employs semantic properties, such as "the support of set X being k" or "the supports of sets X and Y being the same integer value", that cannot be discussed without reference to the fixed dataset at hand (such as [7] or [5], which were inspiring and crucial to our research anyway); these are not comparable to our fully syntactic approach, in which, along the process of constructing a small basis of implications, nothing is used from the dataset after the two collections of rules (those of confidence 1 and those of confidence γ or more) have been mined.

In a previous paper (submitted—see below for more details), we have shown that this setting is amenable to formal discussion, to the extent that we have provided:

- a natural notion of a partial rule being redundant with respect to another partial rule in the presence of exact rules,
- a syntactic calculus corresponding exactly to this notion of redundancy, and
- a specific nonredundant basis, with respect to the same notion of redundancy, for which we have validated empirically that it gives smaller bases than comparable syntactical alternatives.

Here, we follow up with a much deeper theoretical study, which includes, besides a brief survey of the previous points,

- additional characterizations of entailment of a partial rule with respect to a set of partial rules and of the basis proposed previously,
- a slight adjustment improving our previous basis, for which we can prove absolute optimality with respect to our notion of redundancy, and
- substantial progress towards characterizing, by means of logical calculi, the entailment of a partial rule with respect to a *set* of partial rules, rather than with respect to a single one.

2. PRELIMINARIES

$\mathcal{U}$ is a universe of binary attributes, or items. $\mathcal{D}$ denotes a dataset over $\mathcal{U}$. It consists of transactions. Each transaction is a pair: it has a unique identifier and a set of items. 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. We denote $t \models X$ the fact that transaction t has as true all the attributes of X; equivalently, X is a subset of the itemset corresponding to t. The support of X in $\mathcal{D}$, $s_{\mathcal{D}}(X)$, is the cardinality of the set $\{t \in \mathcal{D} \mid t \models X\}$. Whenever $\mathcal{D}$ is clear, we drop the subindex: $s(X)$.

The closure operator $\lambda X.\overline{X}$ maps each itemset X to the largest itemset that has the same support as X; 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. The family of closed itemsets is denoted $\mathcal{X}$

(thus leaving again implicit the dependence on the dataset). 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 [22], [19], and others.

The confidence of an association rule $X \to Y$ in a dataset is $c_{\mathcal{D}}(X \to Y) = \frac{s(XY)}{s(X)}$. Again, often we drop the subindex. We resort to the convention that, if $s(X) = 0$ (which implies $s(XY) = 0$) we redefine the undefined confidence as 1. Implications are association rules of confidence 1. They are closely related to the closure operator indicated above, since $c(X \to Y) = 1$ if and only if $Y \subseteq \overline{X}$. We abbreviate $X \to Y/\gamma$ the implicit mention that $c(X \to Y) \geq \gamma$, and likewise for $X \to Y/1$.

The remainder of this section reviews briefly some of the contributions in the (currently submitted) previous paper of the author.

2.1 Redundancy for Individual Partial Rules

The following notion of redundancy is quite natural and, in the same form or in slight variants, appears in several sources:

Definition 1. Given a set $\mathcal{B}$ of exact rules, rule $X_0 \to Y_0/\gamma$ is γ-redundant with respect to rule $X_1 \to Y_1/\gamma$, denoted $\mathcal{B} \cup \{X_1 \to Y_1/\gamma\} \models X_0 \to Y_0/\gamma$, if every dataset in which the rules of $\mathcal{B}$ hold exactly and the confidence of $X_1 \to Y_1$ is at least γ must satisfy as well $X_0 \to Y_0$ with confidence at least γ.

That is, redundancy, as defined here, corresponds to a form of entailment in which a set of premises, all but one of which have confidence 1, the remaining one with confidence at least γ, entails a conclusion: itself a new rule that holds also with confidence γ (or more). Recently, we have proposed a calculus to derive association rules, which we have proved to match exactly this notion of redundancy as stated below. The calculus consists of four inference schemes involving rules with confidence γ or 1.

(rA) $\frac{X \to Y/\gamma, \quad X \to Z/1}{X \to YZ/\gamma}$

(rI) $\frac{X \to Y/\gamma, \quad Y \to Z/1}{X \to Z/\gamma}$

(ℓA) $\frac{X \to YZ/\gamma}{XY \to Z/\gamma}$

(ℓI) $\frac{X \to Y/\gamma, \quad Z \subseteq X, \quad Z \to X/1}{Z \to Y/\gamma}$

We allow as well to state empty right hand side rules without any premise: $\frac{}{X \to \emptyset/\gamma}$. Note that (rA), employed with empty Y, allows us to transform an exact rule into the corresponding partial rule. The names of the rules indicate with the lower-case letter whether they operate at the right or left hand side, and with the upper-case letter whether their effect is a sort of Augmentation or a sort of composition with an Implication. We use the standard derivation symbol $\vdash$ to denote derivability in this system. The main property of this calculus is:

Theorem 1. *Let $\mathcal{B}$ consist of exact rules. Then, $\mathcal{B} \cup \{X \to Y/\gamma\} \vdash X' \to Y'/\gamma$ if and only if $\mathcal{B} \cup \{X \to Y/\gamma\} \models X' \to Y'/\gamma$.*

2.2 Characterizing Redundancy

We give here a half-new characterization; for a known sufficient condition for redundancy, we show that the necessary condition is also true.

Theorem 2. *Let $\mathcal{B}$ be a set of exact rules, with associated closure operator mapping Z to its closure $\overline{Z}$. Let $X_2 \to Y_2$ be a rule not implied by $\mathcal{B}$, that is, where $Y_2 \not\subseteq \overline{X_2}$. Then, the following are equivalent:*

- $X_1 \subseteq \overline{X_2}$ *and* $X_2Y_2 \subseteq \overline{X_1Y_1}$
- $\mathcal{B} \cup \{X_1 \to Y_1/\gamma\} \models X_2 \to Y_2/\gamma$

Proof. In the particular case where the closure operator is the mere identity, the sufficient condition for redundancy is covered already jointly by theorems 4.1 and 4.2 in [1], which handle respectively "simple" redundancy, which is $X_1 \subseteq X_2$ and $X_2Y_2 = X_1Y_1$ (that is, scheme (ℓA) in our calculus) and "strict" redundancy, which is $X_1 \subseteq X_2$ and $X_2Y_2 \subset X_1Y_1$. Their proof can be extended to a proof of the direct implication of our statement; we show now the reverse implication. We prove that if either of $X_1 \subseteq \overline{X_2}$ and $X_2Y_2 \subseteq \overline{X_1Y_1}$ fails, together with $Y_2 \not\subseteq \overline{X_2}$, then there is a dataset where $\mathcal{B}$ holds with confidence 1 and $X_1 \to Y_1$ holds with confidence at least γ, but the confidence of $X_2 \to Y_2$ is strictly below γ.

First, we point out that all supports are always nonnegative integers, and, therefore, we lose no generality by assuming that γ is a rational number, say $\gamma = \frac{m}{n}$; the inequalities $0 < \gamma < 1$ then translate into $1 \leq m \leq n - 1$, which makes sense only for $n \geq 2$. Then we observe 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 first that $X_1 \not\subseteq \overline{X_2}$: then a dataset consisting only of one or more transactions with itemset $\overline{X_2}$ satisfies (vacuously) $X_1 \to Y_1$ but, given that $Y_2 \not\subseteq \overline{X_2}$, leads to confidence zero for $X_2 \to Y_2$. (A slightly more complex construction allows one also to argue the same fact without resorting to vacuous satisfaction.)

Then consider the case where $X_1 \subseteq \overline{X_2}$, whence the other inclusion fails: $X_2Y_2 \not\subseteq \overline{X_1Y_1}$. Consider a dataset with one transaction for the itemset $\overline{X_2}$ and $n-1$ transactions for the itemset $\overline{X_1Y_1}$. The confidence of $X_1 \to Y_1$ is $\frac{n-1}{n}$, whereas the presence of at least one X_2 and no transaction at all containing X_2Y_2 gives confidence zero to $X_2 \to Y_2$. Thus, in either case, we see that entailment does not hold.

2.3 Characterizing a Basis

To each notion of redundancy, there corresponds a notion of basis. Given a dataset and a confidence threshold, consider all the association rules that hold in that dataset with that confidence: a basis is a subset of them that makes all the others redundant; in our case, since we have a derivability notion corresponding exactly to redundancy via our four inference schemes, a basis is a subset from which all the other rules can be derived. Bases must be irredundant, that is, minimal: removing a rule loses information. But minimality does not imply minimum size; in principle, there could be other, fully unrelated, smaller bases, depending on the notion of redundancy.

Our previous paper proposes a basis construction process that combines a variant, slightly more efficient, of the choice of generators for closed sets as in [19], [22], and several others, with the approach of [1]. An empirical evaluation of the sizes of the bases constructed, compared to other bases in the literature, was also provided. We review it now with some additional notation that we will need below.

Definition 2. 1. Given Y and $X \subseteq Y$, X is a *γ-antecedent* for Y if $c(X \to Y) \geq \gamma$, that is, $s(Y) \geq \gamma s(X)$.

2. We denote

$$\mathcal{I}_\gamma(Y) = \{X \mid X \subseteq Y,\, s(Y) < \gamma s(X)\}$$

the family of subsets that are *not* γ-antecedents of Y.

That is, γ-antecedents are the left hand sides of all the rules of the form $X \to XZ$ with high enough confidence, as obtained by the classical approach [3]. As such, the definition depends only on the set of rules that hold in the original dataset $\mathcal{D}$ with confidence either 1 or at least γ.

Regarding $\mathcal{I}_\gamma(Y)$, inference scheme (ℓA) is, in fact, a form of monotonicity property among the subsets of a fixed itemset Y: if $X \subseteq X' \subseteq Y$ and $c(X \to Y) \geq \gamma$, then $c(X' \to Y) \geq \gamma$; and, if $X' \in \mathcal{I}_\gamma(Y)$ and $X \subseteq X'$ then $X \in \mathcal{I}_\gamma(Y)$. Thus, "apriori-like" algorithms can be used [22] to compute the minimal γ-antecedents. An interesting alternative, that we have used successfully in our preliminary implementation, is given by the following caracterization in terms of transversals of hypergraphs, corresponding to the same intuitions as in a number of previous papers (notably [17], [14], [20]).

PROPOSITION 3. *For a set $\mathcal{F}$ of itemsets, denote $Tr\mathcal{F}$ its family of minimal transversals (or minimal hitting sets): all itemsets that intersect each of the sets in $\mathcal{F}$, and do it minimally, that is, their proper subsets no longer have that property. Let Y be a closed itemset, and $X \subseteq Y$; the following are equivalent:*

1. *X is a* γ-antecedent *for Y,*
2. *X intersects all the sets of the form $Y - Y'$ for $Y' \in \mathcal{I}_\gamma(Y)$,*
3. *X intersects all the sets of the form $Y - Y'$ for $Y' \in \mathcal{I}_\gamma(Y)$ and Y' closed;*

additionally, when this is the case, there is some $X' \subseteq X$ such that $X' \in Tr\{Y - Y' \mid Y' \in \mathcal{I}_\gamma(Y), Y' \in \mathcal{X}\}$.

Proof (sketch). The equivalence of the first two statements is easy to see since I_γ is downwards closed by inclusion and, for $X, Y' \subseteq Y$, X not intersecting some $Y - Y'$ is equivalent to $X \subseteq Y'$. Clearly the second statement implies the third. Conversely, considering any Y', it is straightforward to check that if X intersects $Y - \overline{Y'}$ then X intersects $Y - Y'$. The final remark simply amounts to observing that a nonempty family of finite sets must have inclusion-minimal elements.

The next condition is essentially from [1] (where it was applied to a much larger lattice than here), and parallels one given in [21] for exact rules.

Definition 3. In the same conditions, the valid γ-antecedents of X are

$$\mathcal{A}_\gamma(X) = \mathrm{Tr}\{X - X' \mid X' \in \mathcal{I}_\gamma(X)\} \cap \bigcap \{\mathcal{I}_\gamma(X_1) \mid X \subset X_1 \in \mathcal{X}\}$$

That is, this set consists of the minimal γ-antecedents of X that are *not* γ-antecedents of itemsets strictly larger than X. Thus, let $\mathcal{B} \cup \mathcal{R}$ the set of association rules mined at confidence γ from a dataset, split into exact rules in $\mathcal{B}$ and the rest in $\mathcal{R}$. Consider the set $\mathcal{B}_\gamma$ of association rules, constructed as follows: first, valid γ-antecedents are found for all *closed* sets; then, for each closed set X, we pick a subset $X_0 \subseteq X$ such that $\overline{X_0} = X$ and is of minimum size. Also, for each closed set X, we organize all the valid γ-antecedents in equivalence classes: valid γ-antecedents Y and Y' of X are equivalent if $\overline{Y} = \overline{Y'}$. Finally, we gather rules together into basis $\mathcal{B}_\gamma$ as follows: for each closed set X, for each equivalence class of valid γ-antecedents of X, we check whether the $X - \overline{Y} \neq \emptyset$ and, if so, we keep a rule $Y \to X_0 - \overline{Y}$ where Y is a minimum-size representative of its own equivalence class.

For this basis, a form of completeness with respect to our calculus for association rules in $\mathcal{D}$ is already known, as well as soundness:

THEOREM 4. *1. All the implications in $\mathcal{B}_\gamma$ hold with confidence at least γ in $\mathcal{D}$.*

2. If the implication $Y \to Z$ holds in $\mathcal{D}$ with confidence at least γ, then, taken together with the exact rules, $\mathcal{B}_\gamma \vdash Y \to Z/\gamma$.

A variation of this theorem for our newly proposed, better basis will be proved below.

3. AN OPTIMUM-SIZE BASIS

We move on now into the first main contribution of this paper. We give a more refined basis, similar to $\mathcal{B}_\gamma$ but with a slightly different construction; we show that it is still sound and complete, that is, consists of rules that hold and make redundant all other rules that hold; and then prove that it has a minimum number of rules. Any rule set consisting of rules that do hold in the dataset (with the corresponding confidence), and for which all the rules that hold become redundant according to definition 1, must necessarily have at least as many rules as our basis.

The suboptimality of the basis already constructed lies in a single point. It is possible, for two minimal valid antecedents of the same closed set X, to have comparable closures, say, $Y \subseteq Y'$. Equality is appropriately handled by the construction in the previous section, but proper inclusion is not; the rule with larger closure as antecedent is redundant due to the other one and the inference scheme (ℓA). Examples can be easily constructed.

Our new, more refined basis is constructed as follows. For each closed set X, we will consider a number of closed sets Y properly included in X to act as antecedents. They follow a similar pattern to the one of valid antecedents; but, instead of minimal antecedents, we will pick just minimal closed antecedents. That is:

Definition 4. For each closed set X, a closed set $Y \subset X$ (proper inclusion) is a *basic γ-antecedent* if the following holds:

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

Then we use these antecedents for our basis:

Definition 5. The basis $B^\star_\gamma$ consists of all the rules $Y \to X - Y$ for all closed sets X and all basic γ-antecedents Y of X.

Later on we will discuss how to get smaller rules, to minimize the number of total symbols; for now, we will study the number of rules in $B^\star_\gamma$. This variant of the basis in the previous section is also sound and complete, in the sense that it entails exactly the rules that hold with respect to the supports given by the dataset.

THEOREM 5. *1. All the implications in $B^\star_\gamma$ hold with confidence at least γ in $\mathcal{D}$.*

2. If the implication $Y \to Z$ holds in $\mathcal{D}$ with confidence at least γ, then, taken together with the exact rules, $B^\star_\gamma \vdash Y \to Z/\gamma$.

Proof. All these rules must hold because all the left hand sides are actually antecedents. To prove that all the rules that hold are entailed by $B^\star_\gamma$, assume that indeed $Y \to Z$ holds with confidence γ, that is, $s(YZ) \geq \gamma s(Y)$; thus Y is a γ-antecedent of $\overline{YZ}$. Let X be closed, including YZ (and thus $\overline{YZ}$), and maximal with respect to these two conditions: then $\overline{Y}$ is a γ-antecedent of X but not of any strictly larger closed itemset. Let $Y' \subseteq \overline{Y} \subseteq X$ be closed, a γ-antecedent of X, and minimal with respect to these properties; it is straightforward to check that Y' is then a basic γ-antecedent of X and, hence, $Y' \to X - Y' \in B^\star_\gamma$. We can apply in turn (rA), (ℓA), (rA) again, (ℓI) and, finally (rI) to get the following deductive chain:

$Y' \to X - Y' \vdash Y' \to X \vdash \overline{Y} \to X \vdash Y \to X \vdash Y \to Z$ therefore inferring $Y \to Z$ from $B^\star_\gamma$.

Now we can prove the main result of this section: this basis is of optimum size.

THEOREM 6. *Let $\mathcal{B}'$ be an arbitrary basis, that is, a set of rules such that $\mathcal{B}' \vdash Y \to Z/\gamma$ for all implications $Y \to Z$ that hold in $\mathcal{D}$ with confidence at least γ. Also, assume that the rules in $\mathcal{B}'$ do hold with confidence at least γ. Then, for each implication $Y \to Z \in B^\star_\gamma$, there is in $\mathcal{B}'$ an implication of the form $Y' \to Z'$ with $\overline{Y'Z'} = \overline{YZ}$ and $\overline{Y'} = Y$.*

That is, for each $Y \to X - \overline{Y} \in B^\star_\gamma$, there is a corresponding partial rule in $Y' \to Z' \in \mathcal{B}'$; this rule in $\mathcal{B}'$ provides us with $X = \overline{Y'Z'}$ and $Y = \overline{Y'}$. Thus, both X and Y are univocally determined by $Y' \to Z'$ and, hence, the same rule in $\mathcal{B}'$ cannot yield but one of the rules in $B^\star_\gamma$, so that $\mathcal{B}'$ must have at least as many rules as $B^\star_\gamma$. Therefore, $B^\star_\gamma$ has a minimum number of rules, in an absolute sense, among all bases that are complete according to our inference scheme (whence, by theorem 1, according to the redundancy notion of definition 1).

Proof. We pick any rule $Y \to X - \overline{Y} \in B^\star_\gamma$, that is, where Y is a basic γ-antecedent of X; we consider a derivation of this rule from $\mathcal{B}'$, and we prove inductively (backwards) that all partial rules $Y' \to Z'$ along the derivation fulfill $\overline{Y'Z'} = X$ and $\overline{Y'} = Y$. Note that soundness of the calculus (theorem 1), jointly with the fact that the partial rules in $\mathcal{B}'$ hold, ensures that all the rules along the derivation also have confidence at least γ.

Consider a step $Y_i \to Z_i \vdash Y_{i+1} \to Z_{i+1}$ along the derivation. Assuming $\overline{Y_{i+1}Z_{i+1}} = X$ and $\overline{Y_{i+1}} = Y$, we argue the same properties for Y_i and Z_i, in four different manners depending on what inference scheme is applied at this step.

If the step is (ℓI), then $Z_i = Z_{i+1}$ and $Y_{i+1} \subseteq Y_i \subseteq \overline{Y_{i+1}}$; applying closures in these inequalities, $\overline{Y_{i+1}} = \overline{Y_i} = Y$ follows, and from $Y_{i+1}Z_{i+1} \subseteq Y_iZ_i \subseteq \overline{Y_{i+1}}Z_{i+1}$ and again taking closures we get $\overline{Y_iZ_i} = \overline{Y_{i+1}Z_{i+1}} = X$.

If the step is (rA), then $Y_i = Y_{i+1}$ and $Z_{i+1} = Z_iZ$ for some $Z \subseteq \overline{Y_i}$, which makes it immediate to prove our claim.

If the step is (rI), then again $Y_i = Y_{i+1}$, with $Z_{i+1} \subseteq \overline{Z_i}$. Then one equality is immediate, as well as the inclusion $X = \overline{Y_{i+1}Z_{i+1}} \subseteq \overline{Y_iZ_i}$. To argue equality, since $c(Y_i \to Z_i) \geq \gamma$ as indicated above, and $s(Y_i) = s(\overline{Y_i}) = s(Y)$, it holds $s(\overline{Y_iZ_i}) = s(Y_iZ_i) \geq \gamma s(Y_i) = \gamma s(Y)$, and from $Y \subseteq \overline{Y_iZ_i}$ we find that Y is a γ-antecedent of $\overline{Y_iZ_i}$, a closed set that contains X; thus, it cannot be strictly larger than X for Y to be a basic γ-antecedent of X.

Finally, we consider the case where the step is (ℓA). Then, clearly, $Y_iZ_i = Y_{i+1}Z_{i+1}$, and $Y_i \subseteq Y_{i+1}$, so that $\overline{Y_i} \subseteq \overline{Y_{i+1}} = Y$ and $s(Y) \leq s(\overline{Y_i})$. Again we use the fact that $c(Y_i \to Z_i) \geq \gamma$ to attain the remaining equality $\overline{Y_i} = Y$: Y_i is in fact a γ-antecedent of Y_iZ_i, and of $\overline{Y_iZ_i}$ as well, but $\overline{Y_iZ_i} = \overline{Y_{i+1}Z_{i+1}} = X$, so that Y_i is a γ-antecedent of X. Suppose that, for some closed X' strictly larger than X, Y_i was also a γ-antecedent: then $s(X') \geq \gamma s(Y_i) = \gamma s(\overline{Y_i}) \geq \gamma s(Y)$, which cannot be the case since Y is a basic γ-antecedent of X. Therefore, $\overline{Y_i}$ is a closed subset of Y that fulfills the first two conditions in the definition of basic γ-antecedent of X; then, by the minimality property of basic γ-antecedent Y, necessarily $Y = \overline{Y_i}$. This completes the proof.

3.1 Reducing Rule Sizes

As discussed in several references, such as [19] or [22], we can use generators instead of closed sets in the rules. In the [19] style, we would replace the left hand sides Y in each rule of $B^\star_\gamma$ by a generator of Y. This allows one to show to the user minimum sets of antecedents together with all their nontrivial consequents. In the style of [22], we would replace both Y by one of its generators, and $X - Y$ by $X' - Y$ where X' is a generator of X, thus reducing the total number of symbols if minimum-size generators are used, since we can pick any generator. We find that further discussion of this issue is unnecessary.

4. OTHER PREMISE SETS

In this section we report on partial progress towards a full characterization of entailment, both in semantic terms and by adding extra rules to the calculus. This analysis is nontrivial: the reader is invited to come up, on him/herself, with an example of three association rules such that whenever the first two hold with confidence, say, 0.65, also the third one does. (An example answer is provided next in proposition 7.)

Let us first discuss the following issue: should we assume $\gamma \geq 1/2$? A condition like that would be very natural. Whereas usual support thresholds may be rather low, confidence below half (or not higher enough than 1/2) may be quite misleading. For instance, if it so happens that attribute A and attribute B are almost complementary, in the sense that almost all transactions have exactly one of them, for some antecedents of support not too small, both might be consequents with about 50% confidence, whereas at most one of them would be found as a consequent for higher thresholds. Thus, low confidence thresholds may result in the suggestion that a certain antecedent implies both an attribute and its “negation”.

But the knowledge that $\gamma \geq 1/2$ enlarges the set of valid entailments. For instance, consider the following fact:

PROPOSITION 7. *Let $\gamma \geq 1/2$. Assume 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.*

(However, for $\gamma < 1/2$, this entailment does not hold, as will be discussed in the Conclusions.)

We omit the proof of this proposition since it is just the simplest particular case of theorem 8 below. We will propose later on a more general notion of redundancy; we have a complete analysis of a clearly-defined case encompassing proposition 7 above. In fact, the first purpose of this section is to try and clarify when the following holds, where we still assume a separate set of exact rules $\mathcal{B}$ with its corresponding closure operator:

$\mathcal{B} \cup \{X_1 \to Y_1/\gamma,\ X_2 \to Y_2/\gamma\} \models X_0 \to Y_0/\gamma$

a condition that fully parallels definition 1, only that with two partial rules among the premises, instead of only one.

THEOREM 8. *Let $\mathcal{B}$ be a set of implications, with $\lambda X.\overline{X}$ being its corresponding closure operator, and let $1/2 \leq \gamma < 1$. Then, $\mathcal{B} \cup \{X_1 \to Y_1/\gamma,\ X_2 \to Y_2/\gamma\} \models X_0 \to Y_0/\gamma$ if and only if either:*

1. *$X_0 \to Y_0/1 \in \mathcal{B}$, or*
2. *$\mathcal{B} \cup \{X_1 \to Y_1/\gamma\} \models X_0 \to Y_0/\gamma$, or*
3. *$\mathcal{B} \cup \{X_2 \to Y_2/\gamma\} \models X_0 \to Y_0/\gamma$, 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) $Y_0 \subseteq \overline{X_0 Y_2}$

(It is easy to see that proposition 7 is a particular case of this characterization.)

Proof. 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. 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 bound $s(X_1Y_1)$ (and $s(X_2Y_2)$ 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:

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

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

Summing them up, and then 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, simplifying, $a \geq \gamma(a + b)$, so that

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

as was to be shown.

Now we prove the rightwards direction, and we warn ahead of time, for later use, that the bound $\gamma \geq \frac{1}{2}$ is not necessary for this part. As in the proof of theorem 2, we assume, though, that $\gamma = \frac{m}{n}$, so that we can count on $n - m > 0$ and $1 \leq m \leq n - 1$.

Arguing the contrapositive, we assume that we are in neither of the four cases, and show that then 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 counterexamples, 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, $Y_0 \not\subseteq \overline{X_0}$; since cases (2) and (3) do not happen either, by theorem 2 we know that $X_1 \subseteq \overline{X_0}$ implies $X_0Y_0 \not\subseteq \overline{X_1Y_1}$ and $X_2 \subseteq \overline{X_0}$ implies $X_0Y_0 \not\subseteq \overline{X_2Y_2}$. Later on we will refer to these properties explained in this paragraph as the "extra 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 $X_1 \to Y_1$ into $X_2 \to Y_2$ 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 extra 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 $X_0 \to Y_0$. However, the antecedent 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 γ as follows:

$$\frac{mn-1}{mn} \geq \frac{mn-m}{mn} = \frac{n-1}{n} \geq \frac{m}{n}$$

whereas the support of X_2 is n^2, that of X_2Y_2 is nm, and therefore the confidence is $m/n = \gamma$.

Case B. This corresponds to both of (i) and (ii) failing. Then, for a dataset consisting only of $\overline{X_0}$'s, the antecedent rules hold vacuously whereas the consequent rule 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^2m-1 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 extra 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 extra 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 $n+1$ and the support of X_1Y_1 is 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 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 $X_1 \to Y_1$ and $X_2 \to Y_2$.

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 extra 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 that (i), (ii), and (v) hold, and also the extra 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 $\frac{m-1}{n-1} < \gamma$, the consequent does not hold. However, the antecedents 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.

4.1 Extending the calculus

We work now towards a rule form, in order to enlarge our calculus with entailment from larger sets of premises. 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:

$$\frac{X_1 \to Y_1/\gamma,\ X_2 \to Y_2/\gamma,\ X_1Y_1 \to X_2/1,\ X_2Y_2 \to X_1/1,\ X_1Y_1X_2Y_2 \to Z/1}{X_1X_2Z \to \overline{X_1Y_1Z} \cap \overline{X_2Y_2Z}}$$

and state the following properties:

THEOREM 9. *Given a threshold γ and a set $\mathcal{B}$ of exact rules,*

1. *this rule is sound, and*
2. *together with the rules in section 2.1, 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 8, 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 8 hold. Completeness is argued by considering any rule $X_0 \to Y_0$ entailed by $X_1 \to Y_1/\gamma$ and $X_2 \to Y_2/\gamma$ jointly; if the entailment is improper, apply theorem 1, otherwise just apply this new rule with $Z = \overline{X_0}$ to get $\overline{X_0} \to \overline{X_0Y_1} \cap \overline{X_0Y_2}$ and apply (ℓI) and (rI) to obtain $X_0 \to Y_0$.

5. CONCLUSIONS AND FURTHER WORK

The old open problem [16] of finding a basis for confidence-bounded association rules offers two sensible ways to simplify the statement and, thus, tackle it progressively. One of them is to follow the practitioner's process and fix a single confidence threshold to discuss, or, as later works did, fixing the threshold at two points, one of them 1, in order to discuss exact rules separately, due to their different algebraic properties. We still maintain all of our analysis within this simplification. The other was to define notions of redundancy where, besides exact rules, partial rules as premises were considered singletonwise. We have provided a basis of optimum size for this case.

We wish to progress now towards removing the second simplification, that is, characterizing and providing bases for the notion of redundancy that corresponds to just the first simplification. A formal proposal for that notion could be: Given a set $\mathcal{B}$ of exact rules, and a set $\mathcal{R}$ of partial rules, rule $X_0 \to Y_0/\gamma$ is γ-redundant with respect to them, $\mathcal{B} \cup \mathcal{R} \models X_0 \to Y_0/\gamma$, if every dataset in which the rules of $\mathcal{B}$ hold exactly 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.

It turns out that the number of partial premises that can be actually used in a derivation is tightly related to the value of γ itself. Note that, in theorem 8, the proof of the necessity of at least one of (1) to (4) for entailment to hold only requires $0 < \gamma < 1$. We are orienting our further work towards the following scheme (paper in preparation):

- For $0 < \gamma < 1/2$, entailment holds if and only if either of cases 1, 2, or 3 in theorem 8 holds; that is, proper entailment from a set of two (or more) antecedents never holds, and redundancy is characterized by the calculus in section 2.1;
- For $1/2 \leq \gamma < 2/3$, no proper entailment holds from a set of more than two antecedents, and redundancy is characterized by the same calculus, extended with the rule in theorem 9.

This is extremely suggestive of a general pattern, which could as well provide optimum-size bases for arbitrary values of γ. We are currently developing this line of research. We have also started some experimental comparisons to show the usefulness of our approach in practice. There remains to study also a comparison with other "semantic" redundancy schemes based on the actual values of the support, such as those in [5], [13], [18], or, along a different avenue, [7].

And, then, go on: extending the development to sequences, partial orders, and trees, is not fully trivial, because, as demonstrated in [4], the combinatorial structures may make redundant rules that would not be redundant in a propositional (item-based) framework; also, a precise study of the influence of the support threshold in our proposal is yet to be done, and we believe that the whole approach may yield principled ways of choosing the actual convenient values of the thresholds (γ and the support/frequency lower bound) for specific applications. This author believes that this progress will allow for lifting also the current simplification and solving, eventually, the open problem from [16], having provided to society useful algorithms along the way.

6. 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, 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, 307–328.

[4] J L Balcázar, A Bifet, A Lozano: Mining implications from lattices of closed trees. To appear in Extraction et Gestion des Connaissances 2008.

[5] T Calders, B Goethals: Mining all non-derivable frequent itemsets. PKDD 2002, LNCS 2431, 74–85.

[6] A Ceglar, J F Roddick: Association mining. *ACM Computing Surveys* 38 (2006).

[7] L Cristofor, D Simovici: Generating an informative cover for association rules. ICDM 2002, 597–613.

[8] V Duquenne, J-L Guigues: Famille minimale d'implications informatives résultant d'un tableau de données binaires. *Mathématiques et Sciences Humaines* 24 (1986), 5–18.

[9] A Freitas: Understanding the crucial differences between classification and discovery of association rules. *SIGKDD Explorations*, 2 (2000), 65–69.

[10] B Ganter, R Wille: *Formal Concept Analysis.* Springer 1999.

[11] G C Garriga: Statistical Strategies for Pruning All the Uninteresting Association Rules. ECAI 2004, 430–434.

[12] L Geng, H J Hamilton: Interestingness measures for data mining: a survey. *ACM Computing Surveys* 38 (2006).

[13] B Goethals, J Muhonen, H Toivonen: Mining non-derivable association rules. SDM 2005.

[14] 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.

[15] B Liu, W Hsu, Y Ma: Pruning and summarizing the discovered associations. KDD 1999, 125–134.

[16] M Luxenburger: Implications partielles dans un contexte. *Mathématiques et Sciences Humaines* 29 (1991), 35–55.

[17] H Mannila, H Toivonen: Levelwise search and borders of theories in knowledge discovery. *Data Mining and Knowledge Discovery* 1 (1997), 241–258.

[18] J Muhonen, H Toivonen: Closed non-derivable itemsets. PKDD 2006, 601–608.

[19] 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.

[20] J L Pfaltz, C M Taylor: Closed set mining of biological data. BIOKDD 2002, 43–48.

[21] M Wild: A theory of finite closure spaces based on implications. *Advances in Mathematics* 108 (1994), 118–139.

[22] M Zaki: Mining non-redundant association rules. *Data Mining and Knowledge Discovery* 9 (2004), 223–248.

[23] M Zaki, M Ogihara: Theoretical foundations of association rules. DMKD Workshop on research issues in DMKD (1998).