Two Measures of Objective Novelty in Association Rule Mining⋆

José L. Balcázar[1]

Departamento de Matemáticas, Estadística y Computación
Universidad de Cantabria
Santander, Spain
joseluis.balcazar@unican.es

Abstract. Association rule mining is well-known to depend heavily on a support threshold parameter, and on one or more thresholds for intensity of implication; among these measures, confidence is most often used and, sometimes, related alternatives such as lift, leverage, improvement, or all-confidence are employed, either separately or jointly with confidence. We remain within the support-and-confidence framework in an attempt at studying complementary notions, which have the goal of measuring relative forms of objective novelty or surprisingness of each individual rule with respect to other rules that hold in the same dataset. We measure novelty through the extent to which the confidence value is robust, taken relative to the confidences of related (for instance, logically stronger) rules, as opposed to the absolute consideration of the single rule at hand. We consider two variants of this idea and analyze their logical and algorithmic properties. Since this approach has the drawback of requiring further parameters, we also propose a framework in which the user sets a single parameter, of quite clear intuitive semantics, from which the corresponding thresholds for confidence and novelty are computed.

1 Introduction and Related Work

Association rule mining is a process by which a transactional or relational dataset is explored in an attempt at identifying implications among its elementary components (items or attribute values). The syntax of implications is very suggestive of cause-effect relationships; therefore, such syntax is welcome by human decision-makers and domain experts, who can analyze actions to be taken on the basis of the causality intuitively suggested by the implications found.

The idea of expressing knowledge extracted from data in a form of implications has been proposed in a myriad of contributions, many of these in a manner independent of each other. An early development, largely unknown, that already offered the current notion of association rules as a mere part of a much more expressive logic-based system is described in [20]. The research area of Machine

Learning has contributed also many algorithms to "learn rules from examples", which, often, amounts to identifying implications or variants thereof. Purely logical implications have been explored in many contributions (see [16], [35], and the references there for one of the settings, and [23], [25] for closely related perspectives); a proposal that gave the topic of research and applications of association rules inmense momentum, was the description in [2] of the usefulness of parameterizing the association mining process according to a support constraint and a confidence constraint (or "precision" in [30]). In fact, the support constraint opened the door to the design of practically feasible algorithms, starting with [3]; in fact, different datasets often require different algorithmics: see the outcomes of the FIMI competition [14] and the alternatives described in the survey [12]. On the other hand, there is a clear need of quantifying "degrees of implication" because purely logical implications turn out not to match exactly the needs of practical association mining projects. However, several criticisms could be put forward about confidence as a measure of "degree of implication", and a large number of alternatives have been proposed, evaluated, and studied; the literature about these notions is, in fact quite large [17], [19], [21], [37]. A good survey with many references is [18].

Yet, we prefer to develop our proposal in the context of support and confidence bounds, for several reasons. First, conditional probability is a concept known to many educated users from a number of scientific and engineering disciplines, so that communication with the data mining expert is simplified if our measure is confidence. Second, as a very elementary concept, it is the best playground to study other proposals, such as our contribution here, which could be then lifted to other similar parameters. Third, we believe that, in fact, our relative measures will make up for many of the objections raised against confidence. Additionally, it must be taken into account that the quantity of data is usually insufficient to test the extremely large number of hypotheses given by the set of all possible rules, even if schemes more efficient than the Bonferroni guarantees are employed; and it has been observed and argued that the combination of support and confidence is already very good at discarding rules that are present only as statistical artifacts and do not really correspond to correlations in the phenomenon at the origin of the dataset [32].

Now, let us put forward the following considerations. The outcome of a data mining project is expected to offer some degree of novelty. A wide spectrum of subjective considerations regarding the user's previous knowledge can be considered, and, of course, novelty with respect to knowledge existing previously to the data mining process is hard to formalize. But one fact is clear: novelty cannot be evaluated in an absolute form; it refers to knowledge that is somehow unexpected, and therefore some expectation, lower than actually found, must exist, due to some alternative prediction mechanism. Additionally, an intuitive "rule of thumb" is that the amount of novel facts must be low in order that novelty is actually useful.

We propose to measure the novelty of each rule with respect to the rest of the outcome of the same data mining process. To do this, we resort to recent advances

in the construction of irredundant bases and in mathematical characterizations of the most natural notion of redundancy. As we shall see, a redundant rule is so because we can know beforehand, from the information in a basis, that its confidence will be above the threshold. Pushing this intuition further, an irredundant rule in the basis is so because its confidence is higher than what the rest of the basis would suggest: this opens the door to asking, "how much higher?". If the basis suggests, say, a confidence of 0.8 (or 80%) for a rule, and the rule has actually a confidence of 0.81, the rule is indeed irredundant and brings in additional information, but its novelty, with respect to the rest of the basis, is not high; whereas, in case its confidence is actually 0.95, quite higher than the 0.8 expected, the fact can be considered novel, in that it states something really different from the rest of the information mined. We provide a new notion that formalizes this intuition, and show that it indeed refines very much the data mining process, but has a limitation due to being too close to a fully logical approach. Then we relax slightly the definition into a more useful variant, and we study both concepts.

The main notions to be defined below are similar to the "pruning" proposal from [29], in that the intuition is the same; two major differences are, first, that we will work on an already heavily reduced basis, so that a large portion of the pruning becomes unnecessary, and that for what remains, the pruning in [29] is based on the χ^2 statistic, whereas we will look instead into the confidence thresholds that would make the rule logically redundant. Our notions are also similar to the notion of *improvement*, proposed in [7] (and also briefly discussed in [29], although we are not aware of that proposal having received further attention); this quantity also attempts at discarding uninteresting rules due to the same intuitions as ours; but it is a measure of an absolute, additive confidence increase, with no reference to representative rules or standard redundancy, and it only allows for varying the antecedent into a smaller one, keeping the same consequent. Our quotient-based definitions are more powerful, enjoy better algorithmic properties than those currently known for the analogous difference-based alternative, and are also, in our opinion, more natural.

Our notions have some surface similarity as well with the notion of all-confidence [33] and the related concept of m-patterns [31]. However, these notions are rather restrictive, and provide only strong "niches" where all the sets of attributes within an output pattern depend heavily pairwise among them. We wish to depart in a lesser degree from the standard association rule setting. On the other hand, a strong point of these notions is that they bring in an antimonotonicity property to prune the search space. Instead, we just employ a support bound for its antimonotonicity property, and discuss our contribution in terms of postprocessing the output of a standard frequent closed set miner.

Each of these additional measures, and ours are not exceptions, raises an additional difficulty. For a vast majority of datasets, already the setting of a support and confidence value by a human requires enormous expertise and intuition, and/or insistingly repeated runs of the computation process with different values. Few works discuss the setting of the support threshold for association

rules; worth mentioning are the works [6], [13], [22], and [24], all of which provide interesting advances for the case where the association rules are to be used as a classifier (which is not our case here), using the additional information that one of the attributes will be a as target class; this opens the door to using coverage analysis or criteria related to the ROC curve to orient the decision of which support threshold to use. Many algorithms related to Machine Learning have a similar criticism; say, the parameter corresponding to the box constraint of the soft-margin support vector machines, as one mere example. Many successful algorithms are so through the identification of some sort of autonomic or semi-autonomic self-adjustment of the parameters, thus freeing the user from having to choose a value for them.

However, fully removing all parameters would not be the best choice either. It is clear that different characteristics of datasets (largish or smallish transactions, large or small deviations from the average in the transaction sizes, large or small universe of items, more uniform or less uniform distributions of the individual items) are likely to call for somewhat tailored explorations. Therefore, the data mining process needs some way of tuning the exploration to the dataset at hand. We propose here an interpretation of confidence that allows us to suggest values for the bounds on our new novelty parameters, automatically from the confidence bound.

1.1 Redundancy among Association Rules

We start our analysis from one of the notions of redundancy defined formally first in [1], but employed also, generally with no formal definition, in several papers on association rules; thus, we will qualify sometimes this redundancy notion as "standard". We give up front two equivalent characterizations of the notion: the second one was proposed, as nearly identical "covering"-like simplifications, in several independent sources ([1], [26], [36]); the fact that they are equivalent to standard redundancy, instead of being a simplified variant of it, is quite recent [5].

We denote itemsets by capital letters from the end of the alphabet, and use juxtaposition to denote union, as in XY. For a given dataset $\mathcal{D}$, consisting of transactions, each of which is an itemset labeled with a unique transaction identifier, we can count the *support* $s(X)$ of an itemset X, which is the cardinality of the set of transactions that contain X. The *confidence* of a rule $X \rightarrow Y$ is $c(X \rightarrow Y) = s(XY)/s(X)$.

Lemma 1. *[5] Consider two association rules, $X_0 \rightarrow Y_0$ and $X_1 \rightarrow Y_1$. The following are equivalent:*

1. *The confidence and support of $X_0 \rightarrow Y_0$ are always larger than or equal to those of $X_1 \rightarrow Y_1$, in* all *datasets; that is, for every dataset $\mathcal{D}$, $c(X_0 \rightarrow Y_0) \geq c(X_1 \rightarrow Y_1)$ and $s(X_0Y_0) \geq s(X_1Y_1)$.*
2. *$X_1 \subseteq X_0 \subseteq X_0Y_0 \subseteq X_1Y_1$.*

The fact that 2 implies 1 is easy to see and was pointed out in the references indicated. The fact that 1 implies 2 is nontrivial and much more recently shown.

Whenever rules $X_0 \to Y_0$ and $X_1 \to Y_1$ fulfill either of the two equivalent conditions, we say that $X_0 \to Y_0$ is *redundant* with respect to $X_1 \to Y_1$. As an example, for items A, B, C, and D, the rule $AB \to C$ is redundant with respect to rule $A \to BC$, and is also redundant with respect to $AB \to CD$. The first of the two equivalent forms of definition is akin to the definition of entailment in purely logic-based studies, and we will use sometimes the phrase "logically stronger" to refer to a rule that makes another one redundant with respect to standard redundancy.

Note that the rules $X \to Y$ and $X \to XY$ are mutually redundant, in fact fully equivalent because their confidence $s(XY)/s(X)$ and support $s(XY)$ always coincide. Therefore we consider all association rules where *the right-hand side always includes the left-hand side*, although for the purpose of showing them to the user the repeated items of the left-hand side will be removed from the right-hand side. This simple convention greatly simplifies the mathematical development.

There are several alternative notions of redundancy in the literature; see [5] for further comparisons among a few of them. For this particular notion we have just given, the aim is clear: whatever the dataset under analysis, and the support and confidence parameters, if we find that rule $X_1 \to Y_1$ appears among the mined rules by passing the support and confidence thresholds, any other rule $X_0 \to Y_0$ showing standard redundancy with respect to it is known to be also in the set of mined rules without need to inspect them to check out. This is because the support and confidence must be at least the same as those of rule $X_1 \to Y_1$, whence it passes the thresholds as well.

1.2 Representative Rules

The fact that the output of association rule miners tends to be far larger than desired has been widely reported; it is also self-apparent to anyone that has tried any of the association miners in data mining packages or implementations freely available on the web, e.g. [9].

Our implementation builds on the *representative rules* for association rules, proposed independently and in different but equivalent ways, in [1], in [26], and in [36]. Recently, several new mathematical properties of this basis have been proved, including a form of optimality [5].

Definition 1. *Fix a dataset $\mathcal{D}$ and confidence and support thresholds. The corresponding basis of* representative rules *consists of all the rules that hold in $\mathcal{D}$, passing both thresholds, which are* not *redundant with respect to any other rule that holds in $\mathcal{D}$ for the same thresholds.*

Among several equivalent possibilities to define representative rules, we have chosen a definition so that the following claim becomes intuitively clear: every rule that passes the thresholds for $\mathcal{D}$ is either a representative rule, or is redundant with respect to a representative rule. Indeed, any given rule that is not among the representative rules must be redundant with respect to some other

rule, which again must be redundant with respect to a third, and so on, until finiteness enforces termination that can be only reached by finding a rule in the basis, making redundant all the others found along the way. The formalization of this argument can be found in [26] (Lemma 1 must be taken into account to complete the proof).

Thus, every rule that passes the thresholds for $\mathcal{D}$ is either a representative rule, or is redundant with respect to a representative rule. Moreover, any basis, that is, any set of rules that makes redundant all the rules mined from $\mathcal{D}$ at the given thresholds, must include all the representative rules, since there is no other way of making them redundant. Thus, the representative rules form the unique smallest basis with respect to standard redundancy. (This is not true of rules of confidence 100%; for these absolute implications, the representative basis from [1], [26], [36] can be constructed as well and coincides with the "canonical iteration-free basis" of [38], the nonredundant implications of [40], the proposal in [35] and the "generic" (or "exact min-max") basis of [34]; but all these equivalent proposals fail to reach a minimum size, since there is a more economical alternative [15]. Full discussion can be found in [5], where all these facts, and also the equivalence of our formulation with the original ones, are studied in detail.)

In a sense, representative rules are sort of a required starting point, since they give demonstrably the best basis size one can hope for with no loss of information, with respect to redundancy as defined. Representative rules turn out to be intimately related to closed itemsets and minimal generators. These two notions play an important role in rule mining ([11], [27], [34], [40], [41]). A set is *closed* if there is no proper superset with the same support. A set is a *minimal generator* (or also a *free set*) if there is no proper subset with the same support. In the presence of a support threshold, frequent closed sets are closed sets whose support clears the threshold. Frequent closed sets are very crucial to the algorithmics of association rules and to the identification of irredundant bases. Specifically, in [27] we find a proof of the following nonobvious fact: all representative rules have a minimal generator as antecedent and a closed itemset as consequent (however, not all such pairs give representative rules). Good algorithms and implementations to find them already exist. Absolute optimality of certain versions of these bases is shown in [5].

2 Confidence Width

This section describes the foundations of our proposal. Our intuition is as follows: consider a rule $X \to Y$ of a given confidence, say $c(X \to Y) = c_0 \in [0, 1]$, in a given dataset $\mathcal{D}$. Assume that a fixed support threshold is enforced throughout the discussion, and consider what happens as we vary the confidence threshold γ.

If we set it higher than c_0, that is, $c_0 < \gamma$, the rule at hand will not play any role at all, being of confidence too low for the threshold. As we lower the threshold and reach exactly $\gamma = c_0$, the rule becomes part of the output of any standard association mining process, but two different things may happen:

the question is whether, at the same confidence, some other "logically stronger" rule appears. If not, $X \to Y$ will belong to the representative rules basis for that threshold; but it may be that, at the same threshold, some other logically stronger rule is found. For instance, it could be that both rules $A \to B$ and $A \to BC$ have confidence c_0: then $A \to B$ is redundant and will not belong to the basis for that confidence.

Let's then assume that the rule at hand does appear among the representative rules at the confidence threshold given by its own confidence value; and let's keep decreasing the threshold. At some lower confidence, a logically stronger rule may appear. If a logically stronger rule shows up early, at a confidence threshold γ very close to c_0, then the rule $X \to Y$ is not very novel: it is too similar to the logically stronger one, and this shows in the fact that the interval of confidence thresholds where it is a representative rule is short.

To the contrary, a stronger rule may take long to appear: in that case, only rules of much lower confidence entail $X \to Y$, so the fact that it does reach confidence c_0 is novel in this sense. The interval of confidence thresholds where $X \to Y$ is a representative rule is large. For instance, if the confidence of $A \to AB$ is 0.9, and all rules that make it redundant all have confidences below 0.75, the rule is a much better candidate to novelty than it would be if some rule like $A \to ABC$ would have a confidence of 0.88.

This motivates the following definition:

Definition 2. *Fix a dataset $\mathcal{D}$ and a support threshold τ. Consider a rule that has support at least τ in $\mathcal{D}$, say rule $X \to Y$. Consider all rules that are not equivalent to $X \to Y$, but such that $X \to Y$ is redundant with respect to them, and pick one with maximum confidence in $\mathcal{D}$ among them, say $X' \to Y'$ (thus $c(X' \to Y') \leq c(X \to Y)$). The* confidence width *of $X \to Y$ in $\mathcal{D}$ is:*

$$w(X \to Y) = \frac{c(X \to Y)}{c(X' \to Y')}$$

In case $X \to Y$ is representative, only rules of confidence smaller than γ can make it redundant. In order to check for the existence of $X' \to Y'$, one should mine at lower confidence levels (but see comments after Theorem 1 below). The confidence width can be defined equivalently as the ratio between the extremes of the interval of confidence thresholds that allow the rule to be representative. That is: the highest value where the rule can belong to the representative rule basis is the confidence of the rule; and the denominator is the highest value where there is a different representative rule that makes it redundant, thus forcing it out of the representative basis.

Observe that when $X \to Y$ is redundant with respect to $X' \to Y'$, its confidence must be at least the confidence of the latter, which implies that the confidence width is always greater than or equal to 1. For a rule $X'' \to Y''$, the confidence width is exactly 1 if and only if there is a rule making redundant $X'' \to Y''$ and having the same confidence: this is the same as saying that

$X'' \to Y''$ is never among the representative rules. Regarding upper bounds, in principle there is none, as it may happen that a rule of as large confidence as desired is only redundant with respect to rules of as low confidence as desired.

2.1 Properties and Algorithms

We proceed to study some properties of the confidence width; by combining them with known properties of the standard redundancy and of the representative rules, we will obtain reasonably efficient ways to compute the width of the rules in the basis. We will need a preliminary fact:

Proposition 1. *Consider a rule* $X \to Y$ *and a different rule* $X' \to Y'$ *that makes it redundant; assume* $X' \to Y'$ *has maximum confidence as in the definition of width, say* δ. *Then* $X' \to Y'$ *can be chosen among the representative rules for confidence* δ.

This proposition can be proved easily by resorting to the known fact [26] that every rule of confidence δ is redundant with respect to a representative rule of the same confidence (possibly itself). As indicated in the previous section, rules not in the representative basis have minimum width, namely 1. Thus, to know the confidence width of all the rules it suffices to find it for representative rules.

We do not need to scan all frequent sets, since, as indicated above, it is known that if $X \to Y$ is a representative rule, then XY is a closed set and X is a minimal generator [27]. There are several published algorithms that compute the frequent closed sets and the minimal generators (see the survey [12]); in one form or another, all of them employ the key and well-known fact of the antimonotonicity of the frequent itemsets. These closures and minimal generators can be used to find the representative rules whose width is to be computed, by using the algorithm in [27].

A naive algorithm follows immediately: construct the representative rules and scan them repeatedly, applying Proposition 1 to find, for each rule $X \to Y$, the largest confidence c of any representative rule that makes $X \to Y$ redundant; use Lemma 1 to test for standard redundancy. Once this largest confidence c is known, the width of $X \to Y$ is clearly $w(X \to Y) = \frac{c(X \to Y)}{c}$ by definition. However, notice that this algorithm requires time quadratic in the number of representative rules, and that we mean *all* representative rules, that is, for *all* values of the confidence threshold. This is likely to be a large set.

2.2 An Alternative Algorithm

In some cases, we are likely to wish a computational shortcut: consider the usual case of a user having indicated thresholds for support and confidence, so that our proposal would end in answering the user with a set of representative rules that pass both thresholds, maybe ordered according to width, or possibly even pruned once more at a width threshold. In principle, we only need representative rules at the confidence threshold given. However, to compute the width, we need

all representative rules at all threshold levels. If the threshold is somewhat high, say 0.8, it is overkill to find representative rules at all confidence levels, including, say, 0.1, 0.01, 0.001.

We analyze further properties of the confidence width to search for a faster computation. The key is to avoid much of the exploration in the naive algorithm by precomputing a small amount of side information in a single scan of the closures lattice. We explain now what side information would be sufficient; it is the same as used as a heuristic in [28] to compute a large subset of the representative rules faster[1]. The first step is to find out more about the rules $X' \to Y'$ that could be useful to compute the width of $X \to Y$.

Theorem 1. *Let $X \to Y$ (with $X \subset Y$, proper inclusion) be a representative rule for a fixed dataset $\mathcal{D}$ at some fixed values of support and confidence. Let $X' \to Y'$ be a different rule that makes it redundant, with $X' \subseteq Y'$, and assume $X' \to Y'$ has maximum confidence as in the definition of width. Then either $X = X'$ and Y' is a closed set, immediate superset of Y in the lattice of closed sets, and of maximum support among the closed supersets of Y; or else, $Y = Y'$, and X' is a minimal generator properly included in X and having minimum support among the proper subsets of X.*

Proof. First apply Lemma 1, but assume that we are in neither of the two cases, that is: $X' \subset X \subset Y \subset Y'$ where all the inclusions are proper. Consider the rules $X' \to Y$ and $X \to Y'$. Clearly, appealing again at Lemma 1, both make $X \to Y$ redundant as well. However, since Y is closed, $s(Y') < s(Y)$, and this implies that $c(X' \to Y') < c(X' \to Y)$; similarly, since X is a minimal generator, $s(X) < s(X')$, and again $c(X' \to Y') < c(X \to Y')$. Therefore, the confidence of $c(X' \to Y')$ is not maximum as required, and one of the two rules $X' \to Y$ and $X \to Y'$ will be the one having maximum confidence among those making $X \to Y$ redundant. $\square$

Now, the algorithmic alternative consists in modifying a closure lattice miner to maintain the side information we need. Since the resulting algorithm depends on which closed itemset miner is chosen as starting point, we cannot be fully explicit and keep generality at the same time: we just indicate the changes to be made into the closure miner. They are as follows: along the antimonotonicity-based construction of the frequent closures lattice and the minimal generators, we keep track of the largest existing support of the frequent closed supersets of each frequent closed set Y, let us denote it $mxs(Y)$. Similarly, for each minimal generator X, we keep track of the smallest existing support among the minimal generators properly contained in X, let us denote it $mns(X)$. Then the following proposition explains how to compute the width:

Proposition 2. *Consider a rule $X \to Y$, and assume that both $mxs(Y)$ and $mns(X)$ are defined. Then the width of $X \to Y$ is the minimum of the two*

---

[1] The algorithm in [28], actually, may miss rules due to an incompleteness of the heuristic employed, caused by the fact that Property 9, as stated in that paper, is not true in all cases. This observation will be further elaborated in a later paper.

values: $\frac{mns(X)}{s(X)}$ and $\frac{s(Y)}{mxs(Y)}$. If only one of $mxs(Y)$ and $mns(X)$ is defined, then the corresponding quotient gives the width.

This follows directly from Theorem 1 since each of the two cases corresponds to one of the two options for a rule of maximum confidence making $X \to Y$ redundant. Note that we must compute $mns(X)$ for all minimal generators regardless of whether they are also closed, which is something that can happen (for instance, the empty set is often closed, and is always a minimal generator of the smallest closed set, possibly itself). Note also that some closures Y may not have frequent closed proper supersets, in the sense that all larger closures could fall below the support threshold; likewise, some minimal generators X, namely, the empty set, will lack minimal generators as proper subsets. For such cases, we leave $mxs(Y)$ and $mns(X)$ undefined. Rules where both are undefined do not have a confidence width value according to the definition, because no rule at all is able to make them redundant. Their width can be likened to "infinity". They have not arisen in our empirical analysis, probably due to the support threshold, and further theoretical development regarding this marginal case is undergoing.

Thus, algorithmically, we would use width by precomputing, at the time of finding closures from the dataset, or along the reading from a file if these are constructed by a separate closed set miner, the values $mxs(Y)$ for each frequent closed set Y and $mns(X)$ for each minimal generator X; then, for each representative rule $X \to Y$, we resort to Proposition 2 to compute $w(X \to Y)$ and use it either to filter (against a width threshold) or to sort the representative rules to be given as output.

Proposition 2 tells us also something else: we can discuss the confidence width according to two variants, one of them corresponding to a rule becoming redundant due to a larger consequent, and the other corresponding to a rule becoming redundant due to a smaller antecedent. It will be important shortly to take into account that the items discarded from the antecedent in this last case must still be present in the consequent, since we are assuming, as discussed immediately after Lemma 1, that right hand sides include left hand sides.

2.3   Squint-based Threshold Setting

We propose here a way of connecting the confidence bound to the confidence width bound. The guiding intuition is as follows. First, we rephrase the confidence in a way that, informally, we call "squint": the extent to which we "see" small details. For squint q, sets that differ in a size ratio of q or higher will be considered distinguishable: their difference is actually seen. Note that this cannot be taken as a formal definition, since it may happen that one cannot distinguish X from Y nor Y from Z, yet X can be distinguished from Z. We take it just as an intuition.

Correlating the intuition of squint with the confidence threshold is easy. The implication $X \to Y$ (that is, $X \to XY$) is 100% true exactly if the set of transactions having X coincides with the (in principle potentially smaller) set of transactions having XY. Now, apply the guiding intuition for the squint

parameter: if these sets can be distinguished, we discard the implication. For instance: at squint zero, sharpness is maximum, any existing difference is seen, and only absolute implications are accepted as association rules. However, at squint $q > 0$, to distinguish the set of transactions having X from those having XY we need that their difference has a size, relative to the larger of both sets, of at least q. That is: to distinguish enough counterexamples for the implication, $(s(X) - s(XY))/s(X)$ must be larger than q, and, conversely, the implication is accepted if $(s(X) - s(XY))/s(X) \leq q$, which is equivalent by straightforward algebraic manipulation to $s(XY)/s(X) = c(X \to Y) \geq (1-q)$. (Note that this part also works for the case $q = 0$.) The confidence threshold corresponding to squint q is, then, $1 - q$.

Now, through a similar intuition, each of the two quantities, of which the smallest one provides the confidence witdh as per Proposition 2, can be connected to this "squint" parameter. To obtain a large enough width bound that allows us to "see" rule $X \to Y$, given squint q, the supports $s(X)$ and $mns(X)$ must be clearly different, and also the supports $s(XY)$ and $mxs(XY)$. Thus, we model the corresponding intuitions as $(mns(X) - s(X))/s(X) > q$ and $(s(XY) - mxs(XY))/mxs(XY) > q$. Straightforward algebraic manipulations lead, in both cases, to the condition $w(X \to XY) > 1+q$. That is, the confidence threshold, through the intuition associated to the squint parameter, provides us with a natural suggestion regarding how to set the threshold on the quantity under study: if the confidence threshold is $\gamma = 1 - q$, the natural first choice for confidence width threshold is $1 + q = 2 - \gamma$.

3   Blocked Rules

The main disadvantage often argued against confidence is as follows: for a threshold of, say, $2/3$ (or around 66%), consider a representative rule $A \to B$ of confidence slightly beyond the threshold. It is going to be provided as interesting in the output, suggesting that transactions having A tend to have also B. However, in case the actual frequency of B is rather high, say, 80%, the correlation is in fact negative, since B appears *less* often among the transactions having A than in the whole dataset. The natural reaction, consisting of a normalization by dividing the confidence by the support of B, gives in fact (an analogue to) the deviation from independence $s(AB)/s(A)s(B)$, also known as *interest*, *strength*, or *lift*, a natural measure that, however, lacks the ability to orient the rules, because, in it, the roles of A and B are absolutely symmetric, so that no preference is given for $A \to B$ versus $B \to A$. The same objection appears for the randomization-based proposal in [19]. Confidence width comes close to help but falls a bit short of offering a new solution to this problem. In this section, we relax slightly the notion of confidence width into a notion of "rule blocking" that progresses towards an alternative, nice solution to this difficulty.

For a specific motivating example, let us observe the outcome of mining for association rules at 5% support and 100% confidence the ADULT dataset, available at the UCI Repository [4]. The representative basis for these thresholds

consists of 71 rules. In four of them, the consequent consists of the items "Male" and "Married-civ-spouse". In the other 67, the consequent is, in all cases, just "Male". For instance, we find "Craft-repair, Husband → Male" or "Husband, Some-college, United-States, White → Male".

Further examination reveals that all the left-hand sides consist of the item "Husband", together with one to four additional items. Domain knowledge suggests that all these 67 rules should be superseded by a single full-confidence rule "Husband → Male". However, tuple 7110 includes actually the item "Husband", and the item "Female" instead of "Male". Hence, such a rule does not appear due to the 100% confidence threshold and, instead, many rules that enlarge a bit the left-hand side (enough to avoid tuple 7110 so as to reach confidence 100%) show up. But the real information is the unsurprising (but reassuring) rule given by the domain knowledge and, to some extent, the fact that its confidence is below 100% due to the odd tuple; the user gains nothing by seeing 67 slightly different variations of the same fact. Confidence width does not help: the rule given by domain knowledge does not really make redundant, in the strict logical sense, the 67 rules mined, due to the extra items present in them. On the other hand, the presence of such a large family of rules, each of them improving the confidence only slightly over an existing rule, is a potentially very effective approach to outlier detection.

Yet another example, on the same dataset, that does not involve implications of full confidence but still allows for a similar argumentation, is the following rule, which relates family status with native country: "Unmarried → United-States", of rather high confidence (88%); it might be taken as a suggestion that people coming from abroad into the given U.S. community under analysis tend to come after marriage, but it may as well be an artifact due to the very large ratio of the sample that actually consists of U.S. natives, irrespective of their family status: over 89%. Note, however, that whereas this large support makes the high confidence of the rule "Unmarried → United-States" much less surprising, both high values carry related but different information: the distribution of the U.S. natives along the two different populations, the global one and the one of unmarried people, in principle could be different. The task is, then, to put the squint intuition into use in order to distinguish whether we should maintain both rules "Unmarried → United-States" and, so to say, "∅ → United-States" (the latter being essentially the same as to observe the high support of that item), because the slightly different information they carry is of interest, or we should consider the former subsumed by the latter. Note that either could have higher confidence than the other, depending on the dataset. To cater for such situations, we propose to work out a variation of confidence width, and a corresponding threshold obtained from the confidence threshold via the "squint" intuition, as follows.

3.1 Blocking a Rule with Another

Consider a rule $X \rightarrow Y$, and assume $X \cap Y = \emptyset$. We wish to discard it in case we find a rule $Z \rightarrow Y$, with $Z \subset X$, having almost the same confidence, and

the task is to quantify this "almost". We propose to apply the squint intuition to compare the number of tuples having XY with the quantity that would be predicted from the confidence of the rule $Z \rightarrow Y$; if both sets of tuples are close enough in size, we keep $Z \rightarrow Y$ and forget about $X \rightarrow Y$. We will say, then, that Z "blocks" $X \rightarrow Y$.

Let $c(Z \rightarrow Y) = c$. If Y is distributed along the support of X at the same ratio as along the larger support of Z, we would expect $s(XY) \approx cs(X)$. We employ the "squint" intuition described in the previous section, and evaluate $X \rightarrow Y$ as follows:

Definition 3. *Given rule $X \rightarrow Y$, with $X \cap Y = \emptyset$, a proper subset $Z \subset X$ blocks $X \rightarrow Y$ at squint q if*

$$(s(XY) - c(Z \rightarrow Y)s(X))/(c(Z \rightarrow Y)s(X)) \leq q.$$

In case the difference in the numerator is negative, it would mean that $s(XY)$ is even lower than what $Z \rightarrow Y$ would suggest. If it is positive but the quotient is bounded by q, the difference is "not seen" and $X \rightarrow Y$ still does not bring high enough confidence with respect to $Z \rightarrow Y$ to be considered: it remains blocked. But, if the quotient is larger, and this happens for all Z, then $X \rightarrow Y$ becomes interesting since its confidence is higher enough than suggested by the rules of the form $Z \rightarrow Y$.

It can be readily checked that the particular problems of the ADULT dataset alluded to above are actually solved in this way. Namely, a bit of arithmetic with the actual supports in the dataset shows that, indeed, the rules given as example above, namely, "Craft-repair, Husband $\rightarrow$ Male", "Husband, Some-college, United-States, White $\rightarrow$ Male", or "Unmarried $\rightarrow$ United-States" get all blocked at minimally reasonable squint levels: only an extreme acuteness value for squint will be able to distinguish the different information provided by these rules from that of their blocking rules.

By way of comparison, note that we assume that redundancy due to larger consequents is handled by confidence width, whereas smaller antecedents only in some cases are handled appopriately by width, due to the stringent condition of logical consequence. With blocking, we handle similarly the case of smaller antecedents but in a way that is not as strict as logical consequence.

4 Empirical Validation

We describe first some experimentation made with the notion of confidence width. We compute closures using the C implementation provided by Borgelt [9]. On top of the obtained lattice of closures, we precompute the quantities mxs and mns as per the previous section at the time of loading the closures into our system, use hypergraph transversal techniques to find minimal generators [35], and thus obtain all the representative rules for the support and confidence bounds computed from the squint value. Table 1 indicates some parameters of the datasets on which we have tested our approach.

Table 1. Dataset parameters

Dataset	Source	Transactions	Different Items
Chess	FIMI	3196	75
Retail	FIMI	88162	16470
Adult	UCI	32561	269
Cmc	UCI	1473	36

First, we consider two of the standard FIMI benchmarks [14], of very different characteristics: `chess`, which is a small but very dense dataset on which even high support constraints lead to huge amounts of closed sets and of rules, and the largish, much sparser dataset `retail` coming from a standard application domain (market basket analysis). We have computed the representative rules and their widths, and we have plotted the number of rules passing each of a series of width thresholds. In all cases the computation has taken just a few seconds in a mid-range laptop.

If comparatively larger width values are expected to correlate in some sense with novelty, we wish the number of such rules above comparatively larger thresholds to decrease substantially. This is indeed the behavior we have found. With respect to the `chess` dataset, we have constructed rules of confidence 85% out of the closures lattice formed by frequent closed sets at support 80%. Even for such a large support, the number of closures is around 5083 and the representative rules amount to a number of 15067. It is known from the theoretical advances that all of them are fully irredundant, that is, omitting any of them loses information; however, it makes no sense to expect a human analyst to look at fifteen thousand rules.

We propose, instead, to look at the width values: for this dataset, they range in the quite limited interval between 1 and 1.22; and we see that if we impose a very mild bound of width above 1.005, only 2467 out of the 15067 rules reach it. This means that all the others, even if they are indeed irredundant, this is so due to a rather negligible confidence increase. Higher width bounds exhibit an interesting phenomenon of discontinuity, represented by each plateau of the graph in Figure 1 (left): the maximum confidence width of 1.22 is attained by two rules; a third comes close, and all three have high confidences (between 97% and 99%). Then seventeen more rules show up together near width 1.18, and nothing happens until the width bound gets below 1.13 where a bunch of 31 rules show up together. Below 1.11 we are again at a stable figure of 134 rules, and seventy more appear together at the already quite low confidence width bound of 1.075. All the others, up to 15067, have extremely low width. But the same role cannot be filled directly by confidence: the plot in Figure 1 (right) indicates that there are no steep decreases, no plateau suggesting a good cutpoint shows up, no hint that really any novelty is at play, and, above all, the following fact: the 51 rules of width 1.13 or more all have confidence of 90% or higher, but there are around
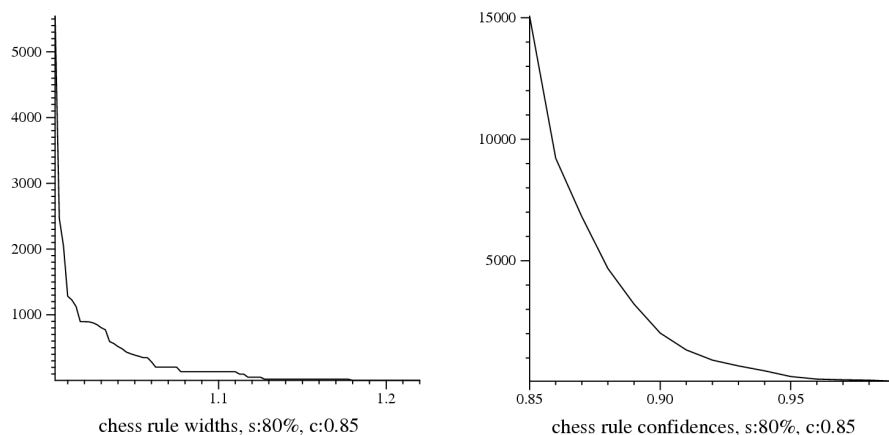
Fig. 1. Chess: Number of rules per width and confidence

1950 other rules, of lower width, attaining the same confidence. Just width is able to focus on the 51 more novel ones.

With respect to `retail`, the behavior of the notion of width is very different, and also very interesting. Huge widths are reached: there are 18 rules whose width is beyond 560 (up to 855.94), whereas the highest next width is just 29: no rule has width between 29 and 560. Another plateau, at width 21, has 7 additional rules, and from there on the number of rules at each width threshold grows steadily.

The FIMI datasets have all their items coded as opaque integers; therefore, the actual rules found cannot be intuitively assessed, in that we do not know their meaning. In order to understand better both the confidence width and the blocked rules, we have performed some further analysis of the very well-known ADULT dataset from the UCI Repository [4]. We use only the train data (we note that the test data has an extra dot in the class attribute). The numeric fields "fnlwgt", "capital-gain" and "capital-loss" were removed, as well as the field "education-num" which is fully redundant with the field "education". In fields "age" and "hours-per-week" the field name was concatenated to the numeric values, in order to distinguish which source to attribute to numeric items. No further cleaning or recoding was done. Table 2 shows the number of rules after the various filtering options. Each row in the table corresponds to a different value of the squint: all thresholds, including support and confidence, are computed from it and used consistently to get each of the figures. Confidence width is computed according to our proposal. Blocking, which in principle should be more powerful, is implemented here in a preliminary form: for each rule of large enough support and confidence, we just test whether it is blocked, at the given squint, by another rule that has also large enough support and confidence. See the Conclusions section for alternatives we wish to explore. The connection of support and squint

Table 2. Number of rules in the ADULT dataset

Squint	Standard	Repr R	Block	Conf Wd	Both
0.10	7916	5706	1509	409	125
0.14	6518	4747	563	290	68
0.18	5270	3730	282	236	40
0.22	4289	2948	162	195	14
0.26	3641	2400	112	156	14
0.30	3024	2012	99	174	11
0.34	2740	1790	71	185	10
0.38	2547	1668	51	219	7
0.42	2255	1486	32	199	6
0.46	2056	1334	24	192	8
0.50	1865	1217	16	196	7

is also very preliminary and under research: in this case we have used the support value $4 * q * r/M$ for squint q, where r is the average transaction size and M is the total number of items; but explanations and variants will be reported in future work.

The columns in Table 2 indicate the number of rules and the effect of filtering the representative rules through thresholds computed according to our proposal on the basis of the squint value. Their meaning is as follows:

– Column "Standard" are the rules found by the standard apriori miner implementation [9]. We must mention that their number is less than the total number of rules since the Apriori rule miner employed only outputs rules with a single item in the consequent, as per the original proposals [3]; our system, and the rest of the figures, do not have this restriction.
– Column "Repr R" is the number of representative rules, which is optimum if we do not want to lose information.
– Column "Block" indicates the number of representative rules clearing the blocked rule condition.
– Column "Conf Wd" indicates the number of representative rules clearing the confidence width threshold.
– Column "Both" indicates the number of representative rules passing both constraints.

For the sake of arguing the interest of our process, we provide in Table 3 the *full* set of rules passing the thresholds at squint 0.32, with their supports, confidences, and confidence widths.

Whereas none is particularly surprising, the advantage is that now we know that, at the corresponding support, everything else is related to these rules through either redundancy, blocking, or lack of novelty; that each of these relationships can be quantified, and that in order to change the level up to which these relationships are computed it suffices to change a single parameter.

Table 3. Rules from the ADULT dataset filtered at squint 0.32

lhs		rhs	Support	Confid	C. Wd
∅	⇒	United-States,White	78.69%	78.69%	1.35
Husband	⇒	Male,Married-civ-spouse,			
		United-States,White	33.91%	83.70%	1.56
Married-civ-spouse	⇒	Husband,Male,			
		United-States,White	33.91%	73.74%	1.56
Not-in-family	⇒	$\leq 50K$,United-States,White	18.06%	70.81%	1.36
Divorced	⇒	$\leq 50K$,United-States,White	9.87%	72.32%	1.39
Black	⇒	$\leq 50K$,United-States	7.62%	79.42%	1.42
hours-per-week:50	⇒	Male,United-States,White	6.37%	73.54%	1.4
Female,Some-college	⇒	$\leq 50K$,United-States,White	6.06%	70.31%	1.37
Adm-clerical,Private	⇒	$\leq 50K$,United-States,White	6.04%	69.43%	1.33
Self-emp-not-inc	⇒	Male,United-States,White	5.71%	73.20%	1.35
$\leq 50K$,Sales	⇒	Private,United-States,White	5.65%	68.95%	1.37

Table 4. Number of rules in the CMC dataset

Squint	Standard	Repr. Rules	Block filter	Conf Wd filter	Both filters
0.10	228	206	120	16	10
0.20	81	67	27	13	8
0.30	33	25	9	7	3
0.40	12	10	2	4	1
0.50	7	5	1	5	1

Finally, we have run our experiments also on an additional dataset: Contraceptive Method Choice, for which the results are displayed in Table 4. This dataset, abbreviated here CMC, is also from [4]; it is similar to ADULT in that it was originally conceived for a prediction task and in that it contains socioeconomic and demographic data where correlations among human factors can be potentially detected; but is very different in terms of size and density. Data come from an actual survey in Indonesia regarding demographic, religious, educational, and offspring data among women, run in 1987. Whereas in ADULT even the representative rules are long to explore manually, in this case the option clearly exists, but it is a frustrating experience: two items ("Good-exposure-to-media", 92%, and "Wife-religion-islam", 85%) are prevalent to such an extent that almost all the rules have just one of these, or both, as consequent, and are therefore uninformative; "High-husband-education" follows closely (61%). Our approach points this out: the rule with empty antecedent "→ Good-exposure-to-media Wife-religion-islam" is clearly singled out beyond squint 0.35, and appears together with the rules "High-wife-education → Good-exposure-to-media High-husband-education" and "High-standard-of-living → Good-exposure-to-media High-husband-education" already at squint 0.25.

This exploration was fast (the closure space consisting of just 1863 closures, thus all rule computations taking just seconds on a mid-range laptop) and immediately suggests to proceed to a more acute exploration, of low squint, to see whether more benign thresholds for support, blocking, and confidence width (automatically compensated by a stricter confidence threshold) provide further information. From such a second phase, also fast, we just note that, at squint around 0.04, the representative rules are several hundred, but our automatically computed thresholds leave just around four dozen rules, most of which are now-unblocked variants with the three very frequent items as consequent (that can be readily discarded at a glance) plus the additional rule "No-children-so-far → No-contraceptive-method", missed in the previous exploration due to low width and support but having large confidence (almost 98%).

5   Conclusions and Further Work

We have proposed two objective approaches to the analysis of the novelty of association rules. A main intuition can be gleaned from the current early developments: it is known that, on the one hand, the standard support-and-confidence bound framework does a good preliminary job for avoiding statistical noise, but, on the other hand, fails somewhat to focus on the really interesting facts; and this is the main reason that has led to a flourishing of variants of notions of "implication degree" to replace confidence, blaming into it the problem. However, we consider now that a viable alternative is to leave the standard support-and-confidence setting on, and complement it, in order to gain further focus, with a measure that does not check the degree of the implications in an alternative way (thus, performing something intuitively analogous to confidence) but which checks a relative intensity of implication compared to the other rules mined in the same process.

Our proposals for this role are confidence width and a related form of blocked rules. Both compare rules among them in search of logical or intuitive redundancy: logical redundancy for the case of width, and a more relaxed, intuitive redundancy for blocking. Our experimental analysis is, admittedly, somewhat limited; but our work so far already suggests several interesting points. It shows that width has the ability to yield wide segments where a width threshold is very robust, and fixing it at a close but different value may select exactly the same rules. It tends to select rules of high confidence but is much more selective.

Also, our proposals open a door to a more human-centered development where one can find ways of evaluating this formal notion of novelty with respect to user-conceived naive notions of novelty. One potential development could be to design an interactive knowledge elicitation tool that, on the basis of the theory described here, could tune in, up to focusing on the user's intuitions for novelty, by showing a handful of unblocked rules of high width, asking the user to label them as novel or not novel: we should develop further the theory to take into account facts such as rules of high width (or support or confidence) being labeled

as not novel, so that the labeling would have consequences on the values of these parameters for the rest of the rules.

We have proposed as well a rule mining framework in which, instead of asking the user to choose, with hardly any guidance, thresholds for all the parameters such as confidence, width, blocking thresholds, and possibly others, a single parameter is chosen, with a degree of semantic intuitive guidance, and then some of the necessary thresholds are autonomously computed by the system from that chosen value. In further work we will analyze the amenability of the support threshold to be treated in the same way; some of our experiments were done according to our preliminary results on that question.

Several major issues need further attention, and are described briefly in the next paragraphs.

5.1   Blocking Rules

Our current implementation only checks for blocking among the rules that have passed the support and confidence thresholds, that is, does not block all "block-able" rules; in most cases this is unproblematic thanks to the confidence width threshold. The analog of Proposition 1 is very easy to prove and we have incorporated this consideration into our experiments. Our further preliminary results suggest the use of a "double-confidence mining" approach analogous to the "double-support mining" approach described in [5], where it is shown how it can be advantageous, in order to distinguish representative rules under a support threshold, to mine closures above a milder threshold than the one set by the user, and employ that information to analyze redundancy of the rules above the user-set threshold. Further mathematical analysis of the formal properties of the blocking process is also necessary to clarify what is the sensible thing to do in case of "transitive blocking", whereby the blocking rule is itself blocked by a third rule: a case that may happen in practice and where we should study the properties we wish for the output rules.

5.2   Robustness

Other parameters instead of those described here may be manageable on the basis of the squint intuition, and possibly with potential advantages. Clearly a large family of candidates is given by the myriad of existing measures of intensity of implication (see [17], [18], [21], [37], among many others). But another family of parameters that could be employed are those whereby the family of closed sets is made more resilient, in the sense of tolerating a small degree of error and considering sets that are "almost" closed (see [8], [10], [11], and the references there). In that approach, the sort of analysis we perform on rules is made in the earlier stage of closure computation. It may become important to understand the potential advantages of this alternative.

In order to safely implement an exploration process as just described, ideally, the main parameter in a system like ours should offer, in as much as possible, *robustness* in some form of continuity: in most cases, slight modifications in the

value of this parameter should not cause extremely big changes of the output. However, the very nature of the discrete sets we work with will impose occassionally abrupt changes. Whereas the squint intuition is already quite good in this respect, one potential way of improvement could be to use the "double-support mining" approach indicated above. The degree of robustness that this approach could contribute to the squint-based analysis is currently under study.

5.3 Alternative Ratios

The relative largeness of a set compared to another, as is considered along the squint-guided intuition, has one additional ambiguity. Namely: as denominator, in the ratio that relates each of our parameters to the squint intuition, we could have chosen the size of the other set. Then, the confidence threshold would result in $1/(1 + q)$ instead of $1 - q$, and the thresholds for the other parameters would end up using a factor $1/(1 - q)$ substituting for the uses of $(1 + q)$. We have chosen the milder, less restrictive, form of the bound for all the cases; further experimentation may suggest, for each threshold, either to stay with the milder bound, to replace it by the more strict one, or to somehow find a way of choosing among both options.

5.4 Revising the Closure Operator

The notion of representative rules is, in fact, only dependent on the dataset; however, the most efficient way to compute it is using the closure operator associated to the dataset. An alternative approach was suggested in [34] and [40], where similar approaches were proposed to treat separately the rules of confidence 100% from the rules of confidence at least γ (a minor variant of the same scheme, which reaches mathematically demonstrable absolute optimality of size for that approach, is described in [5]); all these variants are very tightly coupled to the closure operator, and are better than the representative rules when the confidence threshold is high and there are many rules of confidence 100%. In our preliminary tests we have not detected a major difference in the outcome from using representative rules or from using closure-based redundancy, but further analysis would be in order.

However, the closure operator itself is, essentially, the same mathematical object as the rules of confidence 100%; and, due to the blocked rules and the confidence width bound, we may as well be reluctant to employ anything related to them, since many of these rules may be some sort of artifacts, as we have already discussed in the case of the ADULT dataset. Therefore, we are left with a quandary: should we trust the closure operator when we distrust some of the full-confidence implications that conform it? The effect of this doubt on the representative rules is minor, since they are defined with no reference to closures nor implications and the role of the closure operator in their computation is, essentially, just algorithmic. For this reason, we have developed our approach in terms of representative rules, but further work is necessary to clarify to what extent bases constructed only from the closure space would offer better results.

5.5 Support Bounds versus Itemset Size

We have started to consider some natural heuristics for determining a support threshold. These are based on individual items; however, we can consider briefly here the option of setting different support thresholds for different itemset sizes. This simple idea has, in principle, a serious drawback: if one, generally, already lacks guidance to sensibly set a single support threshold, the problem is exacerbated if we are to set several of them, for the different itemset sizes. Our approach offers a way out: it is conceivable that the squint intuition can be used to suggest supports for different itemset cardinalities.

5.6 Item Distribution

In the computation of support, we have not distinguished among the various items. However, in practical cases, individual items may not be distributed uniformly; Zipfian-like laws or other distributions would be often natural. The effect of this consideration on the computation of the support bound has been discussed in [39]. The extensions of our approach to handle such cases are definitely worth further exploration.

Additional topics become open through our novel proposal: the applicability of the approach to outlier detection has been already hinted at; nowadays, pattern mining on structures more complex than itemsets is necessary in a wide spectrum of application areas, and exporting our approach may not be immediate; we definitely envision the possibility of applying this approach to preference analysis; and other application areas will call for additional developments.

Acknowledgements

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. SIGMOD'93, 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. A Asuncion, D J Newman: UCI Machine Learning Repository. Irvine, CA, University of California, School of Information and Computer Science, 2007; [http://www.ics.uci.edu/~mlearn/MLRepository.html].

5. J L Balcázar: Redundancy, Deduction Schemes, and Minimum-Size Bases for Association Rules; submitted for publication. Pascal Report 4259: [http://eprints.pascal-network.org/archive/00004259]; also available at: [http://www.lsi.upc.edu/~balqui/papers.html].

6. E Baralis, S Chiusano, P Garza: On Support Thresholds in Associative Classification. ACM Symp on Applied Computing, 553–558, 2004.

7. R Bayardo, R Agrawal, D Gunopulos: Constraint-Based Rule Mining in Large, Dense Databases. ICDE'99, 188–197.

8. J Besson, C Robardet, J-F Boulicaut: Mining Formal Concepts with a Bounded Number of Exceptions from Transactional Data. KDID 2004, 33–45

9. C Borgelt: Efficient Implementations of Apriori and Eclat. Workshop on Frequent Itemset Mining Implementations (2003). Available at: [borgelt.net]

10. 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.

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. F Coenen, P Leng, L Zhang: Threshold Tuning for Improved Classification Association Rule Mining. PAKDD 2005, LNAI 3518, 216–225.

14. Frequent Itemset Mining Implementations Repository. Available at: [http://fimi.cs.helsinki.fi/].

15. J-L Guigues, V Duquenne: Famille Minimale d'Implications Informatives Résultant d'un Tableau de Données Binaires. *Math. et Sci. Humaines* 24 (1986), 5–18.

16. B Ganter, R Wille: *Formal Concept Analysis.* Springer 1999.

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

18. L Geng, H J Hamilton: Interestingness Measures for Data Mining: a Survey. *ACM Comp. Surveys* 38, 2006.

19. S Guillaume, F Guillet, J Philippé: "Improving the discovery of association rules with intensity of implication", PKDD 1998, 318–327.

20. P Hájek, M Holeňa: Formal Logics of Discovery and Hypothesis Formation by Machine. *Theoretical Computer Science* 292 (2003), 345–357.

21. C Hébert, B Crémilleux: A Unified View of Objective Interestingness Measures. MLDM 2007, 533–547.

22. C S Kanimozhi Selvi, A Tamilarasi: Association Rule Mining with Dynamic Adaptive Support Thresholds for Associative Classification. IEEE Int. Conf. Comput. Intell. and Multimedia Appl. 76–80, 2007.

23. H Kautz, M Kearns, B Selman: Horn approximations of empirical data. *Artificial Intelligence* 74 (1995), 129–145.

24. M Kawahara, H Kawano: Mining Association Algorithm with Threshold Based on ROC Analysis. IEEE Hawaii International Conference on System Sciences, 2001.

25. R Khardon, D Roth: Reasoning with models. *Artificial Intelligence* 87 (1996), 187–213.

26. M Kryszkiewicz: Representative Association Rules. PAKDD'98, 198–209.

27. M Kryszkiewicz: Fast discovery of representative association rules. RSCTC, 1998, 214–221.

28. M Kryszkiewicz: Closed Set Based Discovery of Representative Association Rules. IDA 2001, 350–359.
29. B Liu, W Hsu, Y Ma: Pruning and Summarizing the Discovered Associations. KDD'99, 125–134.
30. M Luxenburger: Implications Partielles dans un Contexte. *Math. et Sci. Humaines* 29 (1991), 35–55.
31. S Ma, J L Hellerstein: Mining Mutually Dependent Patterns for System Management. IEEE Journal on Selected Areas in Communications 20 (2002), 726–734.
32. N Megiddo, R Srikant: Discovering Predictive Association Rules. KDD'98, 274–278
33. E R Omiecinski: Alternative Interest Measures for Mining Associations in Databases. IEEE Trans. on Knowledge and Data Engineering 15 (2003), 57–69.
34. 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.
35. J L Pfaltz, C M Taylor: Scientific Discovery through Iterative Transformations of Concept Lattices. Workshop Discr. Math. and Data Mining, SDM 2002, 65–74.
36. V Phan-Luong:The Representative Basis for Association Rules. ICDM'01, 639–640.
37. P-N Tan, V Kumar, J Srivastava: Selecting the Right Objective Measure for Association Analysis. *Inf. Syst.* 29(4):293-313 (2004)
38. M Wild: A Theory of Finite Closure Spaces Based on Implications. *Adv. in Math.* 108 (1994), 118–139.
39. H Xiong, P-N Tan, V Kumar: Mining Strong Affinity Association Patterns in Data Sets with Skewed Support Distribution, Third IEEE International Conference on Data Mining ICDM'03, 387.
40. M Zaki: Mining Non-Redundant Association Rules. *Data Mining and Knowledge Discovery* 9 (2004), 223–248.
41. M Zaki, M Ogihara: Theoretical foundations of association rules. Workshop on research issues in DMKD (1998).