FILTERING ASSOCIATION RULES WITH NEGATIONS ON THE BASIS OF THEIR CONFIDENCE BOOST

José L Balcázar, Cristina Tîrnăucă and Marta E. Zorrilla
Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Santander, Spain

Keywords: Association mining, Transactional dataset, Negated items, Confidence, Support, Confidence boost.

Abstract: We consider a recent proposal to filter association rules on the basis of their novelty: the confidence boost. We develop appropriate mathematical tools to understand it in the presence of negated attributes, and explore the effect of applying it to association rules with negations.
We show that, in many cases, the notion of confidence boost allows us to obtain reasonably sized output consisting of intuitively interesting association rules with negations.

1 INTRODUCTION

Among the many data mining techniques widely available nowadays, association rules are a major tool. Association rules are basically defined on transactional data, where there is a global set of items, and the dataset is structured in transactions, each of which is an itemset, that is, a subset of the global set of items. Many standard applications of association mining (e.g. market basket data) obey this syntax, and many association algorithms, both proprietary and open source (such as the `apriori` implementation of (Borgelt, 2003), for instance) work on it. An enormous amount of literature is connected to this topic: see `http://michael.hahsler.net/research/bib/association_rules/` where almost a hundred of the most cited references are enumerated.

In relational data, the dataset consists of tuples where each transaction maps each attribute into one of a number of values available for the attribute. Relational data can be casted into transactional data by considering each potential attribute-value pair as an item. Often, associations on relational data are computed in that way. Conversely, one can consider transactional data as relational in several ways. The one employed in several association miners, such as the associators of Weka as of version 3.6, consists in considering each item as a boolean-valued attribute. But some of these transformations do not preserve all semantics. For instance, the information that different values for the same attribute in the same tuple are incompatible is lost upon performing a conversion into transactional. The associators of Weka, applied to transactional data, give different results than applying a standard transactional associator: Weka will find associations not only among items (attributes of the form "item = true") but also among their negations (as attributes of the form "item = false"), and rules that mix them arbitrarily. These rules can be found on a standard associator as well by "composing" both transformations, that is, preprocessing the transactional data to compute the set of all items and adding to each transaction the corresponding negations: this is the outcome of transforming the transactional dataset into relational and back. We will call this transformation the *neg-expansion* of the dataset. In fact, there may be datasets where considering negated attributes is convenient (Boulicaut et al., 2000; Kryszkiewicz, 2005; Kryszkiewicz, 2009).

However, this "relational-like" expansion of a transactional dataset, where each item is transformed into a pair of items, namely, the positive and the negative versions of the same original item, exhibits an important drawback on many datasets. It is often the case that the universe of items becomes much, much larger than the average size of the transactions. In that case, many transactions "have" the negative versions of many attributes. In standard terminology, the dataset becomes "dense", with the additional algorithmic and conceptual difficulties associated to all dense datasets. Then, rules consisting of negative information easily reach very high confidence and support thresholds. The outcome is a large amount

Balcázar J., Tîrnăucă C. and Zorrilla M..
FILTERING ASSOCIATION RULES WITH NEGATIONS ON THE BASIS OF THEIR CONFIDENCE BOOST.
DOI: 10.5220/0003095802630268
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval* (KDIR-2010), pages 263-268
ISBN: 978-989-8425-28-7

of "noisy" rules that make it difficult to extract knowledge.

Specifically, in previous, recent works a notion of "confidence boost" for association rules has been proposed (Balcázar, 2010) that is able to select a reasonably sized set of output rules, in many cases, in the transactional setting. Confidence boost measures a form of objective "novelty", quantifying to what extent the information in each association rule "looks different" from that of the rest of the rules. Two variants of this intuition can be deployed on closure spaces and several notions of bases. Specifically, we start indeed from the notion of confidence boost for the $\mathcal{B}^*$ basis, and we lift both the basis and its corresponding variant of boost into confidence-boost bounded $\mathcal{B}^*$ rules with negated attributes. We define the neg-expanded closure space, and state and prove mathematically its exact connection with the original closure space: the neg-expansion does not alter the mathematical structure of the original closures, but just extends it. Then we demonstrate that there are interesting cases of association rules in neg-expanded datasets that can be handled efficiently by this form of closure-aware confidence boost.

2 PRELIMINARIES

A given set of available items $\mathcal{U}$ is assumed; subsets of it are called itemsets. We will 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 n 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. An alternative rendering of support is its normalized version $s(X)/n$. The *confidence* of a rule $X \to Y$ is $c(X \to Y) = s(XY)/s(X)$, and its *support* is $s(X \to Y) = s(XY)$. We assume that no transaction of $\mathcal{D}$ includes all items.

Definition 1. *Given a dataset $\mathcal{D}$ over universe of items $\mathcal{U}$, the* neg-expanded *dataset $\mathcal{D}^\ominus$ over universe $\mathcal{U}^\ominus$ is formed by adding to $\mathcal{U}$ a "negative copy" $A^\ominus$ of each item $A \in \mathcal{U}$ to obtain $\mathcal{U}^\ominus$, and adding to each transaction $t \in \mathcal{D}$ all the negative items $A^\ominus$ for all $A \notin t$, to obtain $\mathcal{D}^\ominus$.*

Definition 2. *Given a set $X \subseteq \mathcal{U}$, the* closure of X with respect to $\mathcal{D}$, *denoted by $cl_\mathcal{D}(X)$, is the maximal set (with respect to the set inclusion) $Y \subseteq \mathcal{U}$ such that $X \subseteq Y$ and $s(X) = s(Y)$. Similarly, given $X \subseteq \mathcal{U}^\ominus$, the* closure of X with respect to $\mathcal{D}^\ominus$, *denoted by $cl_{\mathcal{D}^\ominus}(X)$, is the maximal set $Y \subseteq \mathcal{U}^\ominus$ such that $X \subseteq Y$ and $s(X) = s(Y)$.*

It is well-known that the closure of a set is well-defined since there is a single such maximal; closure operators are characterized by the three properties of monotonicity $X \subseteq cl_\mathcal{D}(X)$, idempotency $cl_\mathcal{D}(cl_\mathcal{D}(X)) = cl_\mathcal{D}(X)$, and extensivity, $cl_\mathcal{D}(X) \subseteq cl_\mathcal{D}(Y)$ if $X \subseteq Y$.

Definition 3. *1. Given a dataset $(\mathcal{D}, \mathcal{U})$, we say that $X \subseteq \mathcal{U}$ is* $(\mathcal{D}, \mathcal{U})$-closed *(or simply* closed *when the dataset and the set of attributes is understood from the context) if $cl_\mathcal{D}(X) = X$.*
2. Given the neg-expanded dataset $(\mathcal{D}^\ominus, \mathcal{U}^\ominus)$, we say that $X \subseteq \mathcal{U}^\ominus$ is $(\mathcal{D}^\ominus, \mathcal{U}^\ominus)$-closed *if $cl_{\mathcal{D}^\ominus}(X) = X$.*

Table 1: Dataset $\mathcal{D}$ with $\mathcal{U} = \{a, b, c\}$.

a	b	c
1	0	1
1	0	1
1	0	0
0	1	0
0	1	1

Example 1. *Imagine we have the dataset $\mathcal{D}$ represented in Table 1. The neg-expanded dataset $\mathcal{D}^\ominus$ is represented in Table 2. It is easy to check that although the set $X = \{a\}$ is $(\mathcal{D}, \mathcal{U})$-closed ($s(\{a\}) = 3, s(\{a,b\}) = 0, s(\{a,c\}) = 2$), it is not $(\mathcal{D}^\ominus, \mathcal{U}^\ominus)$-closed ($s(\{a, b^\ominus\}) = s(\{a\}) = 3$).*

Table 2: Dataset $\mathcal{D}^\ominus$ with $\mathcal{U}^\ominus = \{a, b, c, a^\ominus, b^\ominus, c^\ominus\}$.

a	b	c	$a^\ominus$	$b^\ominus$	$c^\ominus$
1	0	1	0	1	0
1	0	1	0	1	0
1	0	0	0	1	1
0	1	0	1	0	1
0	1	1	1	0	0

The following property trivially follows:

Proposition 1. *A set $X \subseteq \mathcal{U}$ is $(\mathcal{D}, \mathcal{U})$-closed if and only if there is no $X' \subseteq \mathcal{U}$ such that $X \subset X'$ and $s(X') = s(X)$. Likewise, a set $X \subseteq \mathcal{U}^\ominus$ is $(\mathcal{D}^\ominus, \mathcal{U}^\ominus)$-closed if and only if there is no $X' \subseteq \mathcal{U}^\ominus$ such that $X \subset X'$ and $s(X') = s(X)$.*

Closure operators are also characterized by the property that any intersection of closed sets is closed. The empty set is closed ($(\mathcal{D}^\ominus, \mathcal{U}^\ominus)$-closed) if and only if no item appears in each and every transaction (and all items appear in at least one transaction, respectively). Note that $s(X) = 0$ implies $cl_\mathcal{D}(X) = \mathcal{U}$ for any $X \subseteq \mathcal{U}$ (since $s(\mathcal{U}) = 0$ by our assumptions). $\mathcal{U}$ is always closed. Again, $cl_{\mathcal{D}^\ominus}(X) = \mathcal{U}^\ominus$ for any $X \subseteq \mathcal{U}^\ominus$ with $s(X) = 0$. Note also that a set X that is $(\mathcal{D}, \mathcal{U})$-closed is not necessarily $(\mathcal{D}^\ominus, \mathcal{U}^\ominus)$-closed, as shown in Example 1.

2.1 Confidence Boost

Motivated by a number of previous works, in (Balcázar, 2010) we proposed to filter rules according to confidence boost, which measures to what extent the rule at hand has higher confidence than rules related to it.

Some notions of redundancy allow for characterizing irredundant bases of absolutely minimum size. A basis for a dataset at a given confidence is a set of rules that hold in the dataset at least at that confidence, and such that every rule that holds in the dataset at the same confidence or higher is made redundant by some rule in the basis; we wish the basis also to be as small as possible. Specifically, the variant of redundancy that takes into account the closure space defined by the dataset leads to the $\mathcal{B}^*$ basis (Balcázar, 2010). This basis offers several computational advantages over its main competitor (called representative rules) at a little price: occasionally it is slightly larger but can be computed much faster. In fact, in the tests we made on educational data, the $\mathcal{B}^*$ rules did coincide exactly with the representative rules.

One can push the intuition of redundancy further in order to gain a perspective of novelty of association rules. Intuitively, an irredundant rule is so because the actual value of its confidence is higher than the value that the rest of the rules would suggest; then, one can ask: "how much higher?". If other rules suggest, say, a confidence of 0.8 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 rules, is not high; whereas, in case its confidence is 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. For instance, in the shopping dataset discussed below, one could consider a rule indicating that market baskets with canned vegetables and frozen meals tend to include beer, with a confidence of 0.84; it turns out that the rule that says that such baskets not only tend to contain beer but they are also bought by a male person has confidence 0.81. We may not want to reduce the right hand side, removing the sex attribute from it, if it is to gain just about a 0.03 percent of improvement of the confidence: more likely, we wish to keep the larger rule and postpone (or altogether remove from consideration) the slightly more confident but less informative rule having only beer as consequent.

For an association rule $X \to Y$, denote $\mathcal{C}(X \to Y)$ the set of all rules $X' \to Y'$ for which the following three conditions are true: $(cl_{\mathcal{D}}(X') \neq cl_{\mathcal{D}}(X) \vee cl_{\mathcal{D}}(XY) \neq cl_{\mathcal{D}}(X'Y'))$, $X' \subseteq cl_{\mathcal{D}}(X)$, $Y \subseteq cl_{\mathcal{D}}(X'Y')$. The first condition corresponds to both rules not being equivalent to each other. This is very close to a nontrivial mathematical characterization of a specific form of closure-based redundancy, as discussed in (Balcázar, 2010).

Definition 4. *The* confidence boost *of an association rule $X \to Y$ (always with $X \cap Y = \emptyset$) is* $\beta(X \to Y) =$

$$= \frac{c(X \to Y)}{\max\{c(X' \to Y') \mid X' \to Y' \in \mathcal{C}(X \to Y)\}}$$

From the definition of the confidence boost, it follows immediately that $X \cap Y = \emptyset$ implies $\beta(X \to Y) = \frac{c(X \to Y)}{\max\{A_{(X,Y)}, B_{(X,Y)}\}}$ where $A_{(X,Y)} = \max_{a \notin XY} \frac{s(XY\{a\})}{s(X)}$ and $B_{(X,Y)} = \max_{X_0 \subset X} \frac{s(X_0 Y)}{s(X_0)}$. Intuitively, this means the following: there may be two reasons to assign a low confidence boost to a rule, one of them due to low relative confidence improvement over some alternative rule having larger right-hand side (corresponding to the added item a); and the other due to low relative confidence improvement over some alternative rule having a smaller left-hand side X_0.

The fact that a low confidence boost corresponds to a low novelty is argued in (Balcázar, 2010). Also, it is proved there that all rules that would be pruned off due to low lift will be pruned as well due to low confidence boost; actually, for rules of the form $a \to b$ for single items a and b, low confidence boost may be due to a possibility of extending the right hand side, but, if this is not so, then the confidence boost is exactly equal to the lift, which is, therefore, also low.

3 MATHEMATICAL TOOLS

We can apply the facts explained after the definition of confidence boost as follows:

Proposition 2. *Let XY and $X'Y'$ be two closed sets such that $XY \subset X'Y'$ and $X \cap Y = X' \cap Y' = \emptyset$. Then* $\beta(X \to Y) \leq \frac{s(XY)}{s(X'Y')}$.

Proof. Let us first notice that from $\beta(X \to Y) = \frac{c(X \to Y)}{\max\{A_{(X,Y)}, B_{(X,Y)}\}}$ we can deduce $\beta(X \to Y) \leq \frac{c(X \to Y)}{A_{(X,Y)}}$. On the other hand,

$$\max_{a \notin XY} s(XY\{a\}) \geq s(XYZ)$$

for any non-empty Z such that $Z \cap XY = \emptyset$, as it suffices to pick any $a \in Z$; hence,

$$\max_{a \notin XY} s(XY\{a\}) \geq s(X'Y').$$

$$\text{Therefore, } \beta(X \to Y) \leq \frac{c(X \to Y)}{s(X'Y')/s(X)} = \frac{s(XY)}{s(X'Y')}.$$

This property is heavily employed in our algorithmics. Given a dataset $\mathcal{D}$ over universe $\mathcal{U}$, let

$(\mathcal{D}^{\ominus}, \mathcal{U}^{\ominus})$ be the neg-expansion as defined previously. In order to ease the readability of this section we will denote $cl_{\mathcal{D}}(X)$ by $\bar{X}$ and $cl_{\mathcal{D}^{\ominus}}(X)$ by $\tilde{X}$.

Lemma 1. *Let $X_1, X_2 \subseteq \mathcal{U}$ be such that $\bar{X}_1 = \bar{X}_2$. Then $\tilde{X}_1 = \tilde{X}_2$.*

Proof. Note that it is enough to show that if $Y = \bar{X}$ for some $X, Y \subseteq \mathcal{U}$ then $\tilde{X} = \tilde{Y}$. From $Y = \bar{X}$ we deduce $s(Y) = s(X)$ and $X \subseteq Y$. On the other hand, $s(Y) = s(\tilde{Y})$, so $s(X) = s(\tilde{Y})$. Moreover $X \subseteq Y \subseteq \tilde{Y}$. Hence, it must be that $\tilde{X} \supseteq \tilde{Y}$. And since $\tilde{Y}$ is a closed set, we get the desired equality.

Lemma 2. *For any $X \subseteq \mathcal{U}$, $\bar{X} \subseteq \tilde{X}$ and $\bar{X} = \tilde{X} \cap \mathcal{U}$.*

Proof. The inclusion $\bar{X} \subseteq \tilde{X}$ is obvious given the way the two closures were defined. Let us now show that the closure of X in $(\mathcal{D}, \mathcal{U})$ can be obtained by taking all and only the positive elements from the closure of X in $(\mathcal{D}^{\ominus}, \mathcal{U}^{\ominus})$. Clearly, $\bar{X} \subseteq \tilde{X} \cap \mathcal{U}$, so we only need to show that $\bar{X} \supseteq \tilde{X} \cap \mathcal{U}$. On one hand, we have $\tilde{X} \cap \mathcal{U} \subseteq \tilde{X}$, so $s(\tilde{X} \cap \mathcal{U}) \geq s(\tilde{X}) = s(X)$. On the other hand, $X \subseteq \tilde{X} \cap \mathcal{U}$, so $s(X) \geq s(\tilde{X} \cap \mathcal{U})$. Therefore, $s(\tilde{X} \cap \mathcal{U}) = s(X)$. We have obtained a set $\tilde{X} \cap \mathcal{U}$ in $\mathcal{U}$ that includes X and has the same support as X. It follows immediately that $\tilde{X} \cap \mathcal{U} \subseteq \bar{X}$.

Let us define $\varphi : cl_{\mathcal{D}}(\mathcal{P}(\mathcal{U})) \rightarrow cl_{\mathcal{D}^{\ominus}}(\mathcal{P}(\mathcal{U}^{\ominus}))$ by $\varphi(\bar{X}) = \tilde{X}$. We prove that this function is an injective homomorphism from the lattice of original closures into the lattice of neg-expanded closures.

Proposition 3. *φ is well-defined, injective and it preserves the inclusion relation (i.e., if $\bar{X}_1 \subseteq \bar{X}_2$ then $\varphi(\bar{X}_1) \subseteq \varphi(\bar{X}_2)$).*

Proof. Obviously, for any closed set $Y \in \mathcal{P}(\mathcal{U})$ there might be several X with $\bar{X} = Y$, so for φ to be well-defined we need to show that given X_1, X_2 with $\bar{X}_1 = \bar{X}_2$, we have $\tilde{X}_1 = \tilde{X}_2$. But this is clear from Lemma 1.

To see that φ is injective, let us take $\bar{X}_1, \bar{X}_2$ in $cl_{\mathcal{D}}(\mathcal{P}(\mathcal{U}))$ such that $\varphi(\bar{X}_1) = \varphi(\bar{X}_2)$. From $\tilde{X}_1 = \tilde{X}_2$ we get $\bar{X}_1 = \tilde{X}_1 \cap \mathcal{U} = \tilde{X}_2 \cap \mathcal{U} = \bar{X}_2$ (by Lemma 2).

Let $X_1, X_2 \subseteq \mathcal{U}$ be such that $\bar{X}_1 \subseteq \bar{X}_2$. We have $X_1 \subseteq \bar{X}_1 \subseteq \bar{X}_2 \subseteq \tilde{X}_2$. But $X_1 \subseteq \tilde{X}_2$ implies $\tilde{X}_1 \subseteq \tilde{X}_2$.

This proposition explains how a "copy" of the original closure structure appears inside the neg-expanded closures, but does not give any information about "the rest", that is, what is added to the lattice by the neg-expansion process. Now, define $\psi : cl_{\mathcal{D}^{\ominus}}(\mathcal{P}(\mathcal{U}^{\ominus})) \rightarrow \mathcal{P}(\mathcal{U})$ by $\psi(\tilde{Y}) = \tilde{Y} \cap \mathcal{U}$. We prove that this function "explains" that each neg-expanded closure is a variation of an original closure.

Proposition 4. *The following properties hold:*
- $\psi(cl_{\mathcal{D}^{\ominus}}(\mathcal{P}(\mathcal{U}^{\ominus}))) = cl_{\mathcal{D}}(\mathcal{P}(\mathcal{U})) \cup \{\emptyset\}$
- If $X \neq \emptyset$ is $(\mathcal{D}, \mathcal{U})$-closed then $\psi(\varphi(X)) = X$
- If $Y \cap \mathcal{U} \neq \emptyset$ is $(\mathcal{D}^{\ominus}, \mathcal{U}^{\ominus})$-closed then $\varphi(\psi(Y)) \subseteq Y$

Proof. Note that ψ is well-defined because even if there might be various $Y' \subseteq \mathcal{U}^{\ominus}$ such that $\tilde{Y} = \tilde{Y'}$, the intersection of their closure with the set of positive attributes will always coincide. So, let us start by determining the image of the function ψ. Let $Y \subseteq \mathcal{U}^{\ominus}$ be an arbitrary $(\mathcal{D}^{\ominus}, \mathcal{U}^{\ominus})$-closed set. If $Y \cap \mathcal{U} = \emptyset$, then $\psi(Y)$ equals by definition the empty set, and we are done. Assume now that $Y \cap \mathcal{U} \neq \emptyset$, and take $Y = Y_1 Y_2$ such that $\emptyset \neq Y_1 \subseteq \mathcal{U}$ and $Y_2 \subseteq \mathcal{U}^{\ominus} \setminus \mathcal{U}$ (where "\" denotes set-theoretic difference). We need to show that Y_1 is $(\mathcal{D}, \mathcal{U})$-closed. Suppose by contrary that it is not. This means that there exists Z such that $Z \supset Y_1$ and $s(Z) = s(Y_1)$. It follows immediately that $s(ZY_2) = s(Y_1 Y_2) = s(Y)$, and we found a set ZY_2 that strictly includes Y and has the same support, a contradiction.

Let $X \neq \emptyset$ be a $(\mathcal{D}, \mathcal{U})$-closed set. We have, $\psi(\varphi(X)) = \psi(\tilde{X}) = \tilde{X} \cap \mathcal{U} = \bar{X} = X$.

As for the inequality $\varphi(\psi(Y)) \subseteq Y$, let us take Y a $(\mathcal{D}^{\ominus}, \mathcal{U}^{\ominus})$-closed set such that $Y \cap \mathcal{U} \neq \emptyset$. We have to show that $\varphi(Y \cap \mathcal{U}) \subseteq Y$. That is, if we take $Y = Y_1 Y_2$ such that $\emptyset \neq Y_1 \subseteq \mathcal{U}$ and $Y_2 \subseteq \mathcal{U}^{\ominus} \setminus \mathcal{U}$ we need to prove that $\tilde{Y}_1 \subseteq Y$. But this follows immediately from $Y = \tilde{Y}$ and $Y_1 \subseteq Y$.

4 EMPIRICAL VALIDATION

We analyzed the role played by the $\mathcal{B}^*$ basis and the closure-based confidence boost on several datasets, comparing association rules mined from the original dataset with the one obtained from its neg-expansion. An important observation is that even if the size of $\mathcal{B}^*$ is exponentially bigger for the neg-expanded dataset, filtering it on the basis of the confidence boost reduces the number of output rules to a very reasonable size. We describe here two such cases. As we shall see, even strict confidence bounds, which lose interesting rules of lower confidence, fail to reduce the output into an usable and sufficiently irredundant outcome. As indicated above, lift is not competitive either: all rules pruned off by lift are also pruned off by confidence boost, but many intuitive redundancies are undetected by lift.

The first example dataset that we analyze in this paper is a typical market basket dataset, taken from the Clementine data mining workbench (Clementine, 2005). The goal is to discover groups of customers who buy similar products and can be characterized demographically, such as by gender or home ownership. The dataset has 1000 transactions over 13 attributes, 11 of them truly boolean (fruitveg, freshmeat, dairy, cannedveg, cannedmeat, frozenmeal, beer, wine, softdrink, fish, confectionery), representing the existing

types of products, and two other binary attributes, gender and homeown. Due to the non transactional nature of these last two attributes, we decided to include both of their values in the mining process of the original database. The report in (Clementine, 2005) identifies three customer profiles: those who buy beer, frozen meals, and canned vegetables (the "beer, beans, and pizza" group), those who buy fish, fruits and vegetables (the "healthy eaters"), and those who buy wine and confectionery.

Table 3: Basket Dataset: Number of Rules.

S	C	$\mathcal{B}^*$ pos | neg	$\beta = 1.05$ pos | neg	$\beta = 1.20$ pos | neg
0.10	0.7	12 | 31708	10 | 208	7 | 14
	0.8	5 | 20517	5 | 43	5 | 4
	0.9	4 | 72	4 | 5	0 | 0
0.15	0.7	6 | 13350	6 | 68	6 | 14
	0.8	2 | 8529	2 | 20	2 | 5
	0.9	0 | 1	0 | 0	0 | 0
0.30	0.7	0 | 1686	0 | 28	0 | 8
	0.8	0 | 1015	0 | 16	0 | 3
	0.9	0 | 0	0 | 0	0 | 0

The results obtained with different values for confidence and support parameters are shown in Table 3, which reports the number of rules passing the thresholds for each case. The five $\mathcal{B}^*$ rules obtained with support 0.10 and confidence 0.80 when mining the original dataset reveal that those who buy beer, frozen meals, and canned vegetables are mostly men, those who buy sweets and wine are usually women, and healthy eaters (fish, fruits and vegetables) do not own a house.

[c: 0.89 s: 0.12] fish fruitveg ⇒ nohomeowner
[c: 0.82 s: 0.14] beer frozenmeal ⇒ male cannedveg
[c: 0.84 s: 0.14] beer cannedveg ⇒ male frozenmeal
[c: 0.81 s: 0.14] cannedveg frozenmeal ⇒ male beer
[c: 0.86 s: 0.12] confectionery wine ⇒ female

On the other hand, allowing "negative" instances, the size of the basis, that is, *after redundancy removal*, increases dramatically to 20517 rules, a number that would discourage any human trying to figure out some correlations in the dataset. It is only due to the confidence boost that this number is decreased to reasonable values. However, it is important to allow "negated" items, and applying the confidence boost bound actually allows us to do so and remain within reasonable figures: using negations, one may discover other interesting facts that are hidden in the data. For example, the 43 rules obtained with support 0.10, confidence 0.80 and confidence boost 1.05 reveal, besides some of the facts we knew from the "positive" case, the following tendencies: - Those that own a house do not buy fish:

[c: 0.81 s: 0.40] homeowner ⇒ nofish

- There are products very seldom bought, like fresh meat, dairy and soft drinks:

[c: 0.81 s: 0.81] ⇒ nofreshmeat
[c: 0.82 s: 0.82] ⇒ nodairy
[c: 0.81 s: 0.81] ⇒ nosoftdrink

- Women do not buy beer or frozen meals:

[c: 0.81 s: 0.41] female ⇒ nobeer
[c: 0.81 s: 0.41] female ⇒ nofrozenmeal
[c: 0.85 s: 0.12] confectionery wine ⇒ nofrozenmeal
[c: 0.82 s: 0.11] confectionery wine ⇒ nobeer

- Men do not buy sweets:

[c: 0.80 s: 0.39] male ⇒ noconfectionery

The second dataset we analyze deals with real data from a virtual course entitled "Introduction to multimedia methods". It is a subject of 6 ECTS which was taught in the first semester of 2009 at the largest virtual campus in Spain, called G9. It is a practical course having as final objective teaching the students how to use a particular multimedia tool. The goal is to discover the resources which are commonly used together in each session, thus allowing the instructors to find out which collaborative tools are used more frequently (wiki, chat, forum, etc.) by their students, which ones are rather ignored, and which is the profile of the learning process followed by the students. This information is very valuable in order to propose tasks according to the learner's learning style.

Table 4: E-learning Dataset: Number of Rules

S	C	$\mathcal{B}^*$ pos | neg	$\beta = 1.05$ pos | neg	$\beta = 1.20$ pos |neg
0.3	0.7	4 | 7152	3 | 34	1 | 9
	0.8	3 | 6269	3 | 16	1 | 2
	0.9	1 | 6943	1 | 1	0 | 0
0.4	0.7	2 | 4155	1 | 6	1 | 0
	0.8	1 | 3605	1 | 4	1 | 0
	0.9	0 | 3969	0 | 0	0 | 0
0.5	0.7	1 | 2131	1 | 14	1 | 0
	0.8	1 | 2103	1 | 6	1 | 0
	0.9	0 | 2275	0 | 0	0 | 0

This dataset contains 6206 transactions over 14 attributes, each of them having value 1 or 0, indicating whether the respective course resource was visited or not in that session. The attributes are: content-page, mail, forum, chat, web-link, organizer, learning-objectives, assignment, calendar, filemanager, who-is-online, announcement, my-grades and student-bookmark.

The results obtained with different values for confidence and support parameters are shown in Table 4.

Observing it, one can see that the number of rules obtained with support 0.30 and confidence 0.70 for the original dataset (without negated attributes) is 4:

[c: 0.74 s: 0.45] forum ⇒ organizer
[c: 0.97 s: 0.34] content-page ⇒ organizer
[c: 0.89 s: 0.32] assignment ⇒ organizer
[c: 0.81 s: 0.81] ⇒ organizer

The rules obtained are not very informative, basically saying that whatever they do, students also visit the organizer page. But this does not come as a surprise, given the fact that "organizer" is the main front page of the course.

Clearly, these rules cannot offer information about which are the resources less used, like the ones obtained by using the neg-expanded dataset. Here are some of the 34 rules mined at support 0.3, confidence 0.7 and confidence boost 1.05 with negated attributes:

[c: 0.70 s: 0.70] ⇒ announcement=0 file-manager=0 calendar=0 learning-objectives=0 student-bookmark=0 web-link=0 who-is-online=0 chat=0

[c: 0.86 s: 0.30] content-page=1 ⇒ announcement=0 chat=0 organizer=1 student-bookmark=0

[c: 0.84 s: 0.30] content-page=1 ⇒ chat=0 student-bookmark=0 mail=0

[c: 0.86 s: 0.30] content-page=1 ⇒ my-grades=0 organizer=1 student-bookmark=0

[c: 0.88 s: 0.31] content-page=1 ⇒ who-is-online=0 organizer=1 student-bookmark=0

[c: 0.86 s: 0.30] content-page=1 ⇒ organizer=1 web-link=0 student-bookmark=0

[c: 0.85 s: 0.30] content-page=1 ⇒ calendar=0 chat=0 organizer=1 student-bookmark=0

[c: 0.85 s: 0.30] content-page=1 ⇒ chat=0 organizer=1 student-bookmark=0 file-manager=0

[c: 0.85 s: 0.30] content-page=1 ⇒ chat=0 organizer=1 who-is-online=0

In the first rule, one can see that there are many resources that are scarcely used, like the chat or the announcement page; therefore, if the instructor has something important to communicate to the students, the best option would be to put it in the forum (a resource known to be accessed more often from the positive rules). Furthermore, one may note that when the students connect to the platform in order to study (i.e., when they visit content-page resources), they do not visit the chat, their bookmark or email.

A further, similar analysis of this dataset, also by comparison with a different one with similar origin and quite different characteristics, in terms of association rules with negations and high confidence boost, and including an additional pruning heuristic, is described in (Balcázar et al., 2010).

5 CONCLUSIONS AND FUTURE WORK

In many practical applications, the output of a data mining process could greatly benefit from adding to the dataset the "negated" versions of the attributes. One of the problems that arises though is that the resulting set of rules mined is huge, making human interpretation unfeasible. In this paper we propose to use a recently introduced notion called confidence boost that is able to filter out those rules that are not "novel", by quantifying to what extent the information in each association rule "looks different" from that of the rest of the rules. Our implementation employs the open-source closure miner from (Borgelt, 2003), and is available at `slatt.googlecode.com`.

As future work, we would like to look into (mathematical and practical) ways of pushing the confidence boost constraint at an earlier stage of the algorithm, thus avoiding the vast amount of time dedicated to compute closed sets that will not be used, or to generate thousands of rules that will be later on discarded based on their low confidence boost.

REFERENCES

Balcázar, J. L. (2010). Formal and computational properties of the confidence boost in association rules. Available at: [http://personales.unican.es/balcazarjl].

Balcázar, J. L., Tîrnăucă, C., and Zorrilla, M. (2010). Mining educational data for patterns with negations and high confidence boost. Accepted for TAMIDA'2010; available at: [http://personales.unican.es/balcazarjl].

Borgelt, C. (2003). Efficient implementations of apriori and eclat. In Goethals, B. and Zaki, M. J., editors, *FIMI*, volume 90 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Boulicaut, J.-F., Bykowski, A., and Jeudy, B. (2000). Towards the tractable discovery of association rules with negations. In *FQAS*, pages 425–434.

Clementine (2005). Clementine 10.0 desktop user guide.

Kryszkiewicz, M. (2005). Generalized disjunction-free representation of frequent patterns with negation. *J. Exp. Theor. Artif. Intell.*, 17(1-2):63–82.

Kryszkiewicz, M. (2009). Non-derivable item set and non-derivable literal set representations of patterns admitting negation. In Pedersen, T. B., Mohania, M. K., and Tjoa, A. M., editors, *DaWaK*, volume 5691 of *LNCS*, pages 138–150. Springer.