Galois Lattices and Bases for M_{GK}-valid Association Rules

Jean Diatta[1], Daniel R. Feno[1,2], and A. Totohasina[2]

[1] IREMIA, Université de La Réunion, 15, Avenue René Cassin - B.P. 7151
97715, Saint-Denis. Messag Cedex 9, France
{jdiatta,drfeno}@univ-reunion.fr
[2] ENSET, Université d'Antsiranana - B.P. 0 - Antsiranana 201 Madagascar
{fenodaniel2,totohasina}@yahoo.fr

Abstract. We review the main properties of the quality measure M_{GK}, which has been shown to be the normalized quality measure associated to most of the quality measures used in the data mining literature, and which enables to handle negative association rules. On the other hand, we characterize bases for M_{GK}-valid association rules in terms of a closure operator induced by a Galois connection. Thus, these bases can be derived from a Galois lattice, as do well known bases for Confidence-valid association rules.

Keywords. Closure operator, Basis, Galois connection, Negative association rule, Quality measure.

1 Introduction

Association rules reveal attributes (or attribute values) that occur together frequently in a data set. Their relevance is commonly assessed by means of quality measures. Several quality measures have been proposed in the literature [1], the most popular of them being the well-known Support and Confidence [2]. A major problem faced in association rule mining is the large number of valid rules, *i.e.*, rules that meet specific constraints relative to a given (set of) quality measure(s). Such a situation is generally due to the presence of many redundant and/or trivial rules in the set of valid ones. A way to cope with these redundant and trivial rules is to generate a basis, *i.e.*, a minimal set of rules from which all the valid rules can be derived, using some inference axioms.

In this paper, we consider the quality measure M_{GK} independently introduced in [3] and in [4], and which has been shown to be the normalized quality measure associated to most of the quality measures used in the data mining literature [5]. On the one hand, we review its main properties. On the other hand, we characterize bases for M_{GK}-valid association rules in terms of a closure operator induced by a Galois connection [6]. This result shows that these bases can be derived from a Galois lattice, as do well known bases for Confidence-valid association rules [7, 8]. The rest of the paper is organized as follows.

Basic concepts relative to association rules and Galois lattices, and the main properties of the quality measure $\mathrm{M_{GK}}$ are presented in Section 2. Section 3 is devoted to two known bases for Confidence-valid association rules, whereas the bases we propose for $\mathrm{M_{GK}}$-valid rules are dealt with in Section 4. Finally, a short conclusion is included in the end of the paper.

2 Association rules, Quality measures, Galois lattices

2.1 Association rules

In this paper, we place ourselves in the framework of a binary context $(\mathcal{E}, \mathcal{V})$, where $\mathcal{E}$ is a finite entity set and $\mathcal{V}$ a finite set of boolean variables (or items) defined on $\mathcal{E}$. The subsets of $\mathcal{V}$ will be called *itemsets*, and an entity "e" will be said to contain an item "x" if $x(e) = 1$.

Definition 1 *An association rule of* $(\mathcal{E}, \mathcal{V})$ *is an ordered pair* (X, Y) *of itemsets, denoted by* $X \to Y$*, where* Y *is required to be nonempty. The itemsets* X *and* Y *are respectively called the "*premise*" and the "*consequent*" of the association rule* $X \to Y$*.*

Given an itemset X,

- X' will denote the set of entities containing all the items of X, *i.e.*,

$$X' = \{e \in \mathcal{E} : \forall x \in X[x(e) = 1]\}, and$$

- $\overline{X}$ will denote the negation of X, *i.e.*, $\overline{X}(e) = 1$ if and only if there exists $x \in X$ such that $x(e) = 0$; it may be noticed that $\overline{X}' = \mathcal{E} \setminus X'$.

Table 1 presents a binary context $\mathbb{K} = (\mathcal{E}, \mathcal{V})$, where $\mathcal{E} = \{e_1, e_2, e_3, e_4, e_5\}$ and $\mathcal{V} = \{A, B, C, D, E\}$. If we let $X = \{B, C\}$ then $X' = \{e_2, e_3, e_5\}$ and $\overline{X}' = \{e_1, e_4\}$.

Table 1. A binary context

	A	B	C	D	E
e_1	1	0	1	1	0
e_2	0	1	1	0	1
e_3	1	1	1	0	1
e_4	0	1	0	0	1
e_5	1	1	1	0	1

According to the definition above, the binary context $(\mathcal{E}, \mathcal{V})$ contains $2^{|\mathcal{V}|}(2^{|\mathcal{V}|} - 1)$ association rules among which several are certainly irrelevant. To cope with this, quality measures, also called interestingness measures, are used to capture only those association rules meeting some given constraints [1]. In the sequel, $\mathcal{E}$ will denote a finite entity set, $\mathcal{V}$ a finite set of boolean variables defined on $\mathcal{E}$, and $\mathbb{K}$ the binary context $(\mathcal{E}, \mathcal{V})$.

2.2 Quality measures for association rules

Let Σ denote the set of association rules of the binary context $\mathbb{K}$.

Definition 2 *A* quality measure *for the association rules of* $\mathbb{K}$ *is a real-valued map* μ *defined on* Σ*.*

There are several quality measures introduced in the literature, the most popular of them being Support and Confidence [2].

The support of an itemset X, denoted by $\text{Supp}(X)$, is the proportion of entities in $\mathcal{E}$ containing all the items belonging to X; it is defined by $\text{Supp}(X) = \frac{|X'|}{|\mathcal{E}|}$, where, for any finite set W, $|W|$ denotes the number of its elements. Denoting by p the intuitive probability measure defined on $(\mathcal{E}, \mathcal{P}(\mathcal{E}))$ by $p(E) = \frac{|E|}{|\mathcal{E}|}$ for $E \subseteq \mathcal{E}$, the support of X can be written in terms of p as $\text{Supp}(X) = p(X')$.

The support of an association rule $X \rightarrow Y$ is defined by:

$$\text{Supp}(X \rightarrow Y) = \text{Supp}(X \cup Y) = p((X \cup Y)') = p(X' \cap Y').$$

The confidence of $X \rightarrow Y$, denoted by $\text{Conf}(X \rightarrow Y)$, is the proportion of entities containing all the items belonging to Y, among those entities containing all the items belonging to X; it is defined by:

$$\text{Conf}(X \rightarrow Y) = \frac{\text{Supp}(X \rightarrow Y)}{\text{Supp}(X)} = \frac{p(X' \cap Y')}{p(X')} = p(Y'|X'),$$

where $p(Y'|X')$ is the conditional probability of Y' given X'.

The two following straightforward inequalities involving conditional probabilities may help to understand the definition of the quality measure M_{GK} below.

(i) if $p(Y'|X') \geq p(Y')$, then $0 \leq p(Y'|X') - p(Y') \leq 1 - p(Y')$;
(ii) if $p(Y'|X') \leq p(Y')$, then $-p(Y') \leq p(Y'|X') - p(Y') \leq 0$.

The quality measure M_{GK} independently introduced in [3] and in [4], is defined by:

$$\text{M}_{\text{GK}}(X \rightarrow Y) = \begin{cases} \frac{p(Y'|X') - p(Y')}{1 - p(Y')} & \text{if } p(Y'|X') \geq p(Y'); \\ \frac{p(Y'|X') - p(Y')}{p(Y')} & \text{if } p(Y'|X') \leq p(Y'). \end{cases}$$

In this paper, we will be mainly concerned with the quality measures Confidence and M_{GK}. The quality measure Confidence is clearly a probability measure and its properties are more or less well known. For instance, $\text{Conf}(X \rightarrow Y) = 0$ if and only if X and Y are incompatible. Moreover, the Confidence measure is not symmetric (*i.e.* $\text{Conf}(X \rightarrow Y)$ is not always equal to $\text{Conf}(Y \rightarrow X)$), and

$\mathrm{Conf}(X \to Y) = 1$ if and only if $X' \subseteq Y'$, *i.e.*, if X logically implies Y. However, the Confidence measure does not reflect the independence between the premise and the consequent of an association rule. Indeed, in case of independence between X and Y, $p(Y'|X') = p(Y')$ and, equivalently, $p(X'|Y') = p(X')$. Furthermore, as quoted in [9], Confidence does not satisfy the logical principle of contraposition, *i.e.*, $\mathrm{Conf}(\overline{Y} \to \overline{X})$ is not always equal to $\mathrm{Conf}(X \to Y)$.

On the other hand, it can be easily checked that $\mathrm{M_{GK}}$ satisfies the five following properties:

1. $\mathrm{M_{GK}}(X \to Y) = -1$ if and only if X and Y are incompatible, *i.e.*, if $p(X' \cap Y') = 0$;
2. $-1 \leq \mathrm{M_{GK}}(X \to Y) < 0$ if and only if X disfavors Y (or X and Y are negatively dependent), *i.e.*, if $p(Y'|X') < p(Y')$;
3. $\mathrm{M_{GK}}(X \to Y) = 0$ if and only if X and Y are independent, *i.e.*, if $p(X' \cap Y') = p(X')p(Y')$;
4. $0 < \mathrm{M_{GK}}(X \to Y) \leq 1$ if and only if X favors Y (or X and Y are positively dependent), *i.e.*, if $p(Y'|X') > p(Y')$;
5. $\mathrm{M_{GK}}(X \to Y) = 1$ if and only if X logically implies Y, *i.e.*, if $p(Y'|X') = 1$.

This shows that the values of $\mathrm{M_{GK}}$ lie into the interval $[-1, +1]$ as well as they reflect references situations such as incompatibility, negative dependence, independence, positive dependence, and logical implication between the premise and the consequent. Thus, according to [5], $\mathrm{M_{GK}}$ is a normalized quality measure. Moreover, it has been shown in [5] that $\mathrm{M_{GK}}$ is the normalized quality measure associated to most of the quality measures proposed in the literature, including Support and Confidence [2], ϕ- coefficient [10], Laplace, Rule interest, Cosine and Kappa (cf. [11]), and Lift [12]. That is, if we normalize such a quality measure by transforming its expression in order to make its values both lie into the interval $[-1, +1]$ and reflect the five reference situations mentioned above, then we obtain the quality measure $\mathrm{M_{GK}}$. In other words, all these quality measures can be written as affine functions of $\mathrm{M_{GK}}$, with coefficients depending on the support of the premise and/or the support of the consequent. Furthermore, unlike several other quality measures, $\mathrm{M_{GK}}$ satisfies the logical principle of contraposition in case of positive dependence, *i.e.*, $\mathrm{M_{GK}}(\overline{Y} \to \overline{X}) = \mathrm{M_{GK}}(X \to Y)$ when X favors Y [13]. In addition, the greater the absolute value of $\mathrm{M_{GK}}(X \to Y)$, the stronger the (positive or negative) dependence between X and Y.

The following result provides us with relationships between positive dependence and negative dependence.

Proposition 1 *Let X and Y be two itemsets. Then the three following conditions are equivalent.*

(1) X disfavors Y.
(2) X favors $\overline{Y}$.
(3) $\overline{X}$ favors Y.

This result shows that the so-called right-hand side negative (RHSN) rule $X \to \overline{Y}$ and/or the so-called left-hand side negative (LHSN) rule $(\overline{X} \to Y)$ can be of interest when X disfavors Y. This is an additional motivation for our choice of $\mathrm{M_{GK}}$ because $\mathrm{M_{GK}}$ enables to handle negative rules as well as positive ones, *i.e.*, those which do not involve negation of itemsets.

It should be noticed that only a rule whose premise favors its consequent is interesting, regardless if it is a positive or a negative rule. Thus, let ${\mathrm{M_{GK}}}^f(X{\rightarrow}Y)$ denote the value of $\mathrm{M_{GK}}(X{\rightarrow}Y)$ when X favors Y, *i.e.*, ${\mathrm{M_{GK}}}^f(X{\rightarrow}Y) = \frac{p(Y'|X')-p(Y')}{1-p(Y')}$, and let ${\mathrm{M_{GK}}}^{df}(X{\rightarrow}Y)$ denote the value of $\mathrm{M_{GK}}(X{\rightarrow}Y)$ when X disfavors Y, *i.e.*, ${\mathrm{M_{GK}}}^{df}(X{\rightarrow}Y) = \frac{p(Y'|X')-p(Y')}{p(Y')}$. The next result shows that the value of $\mathrm{M_{GK}}$ for a RHSN rule is equal to that of $\mathrm{M_{GK}}$ for the corresponding positive rule, on the one hand, and, on the other hand, this value both determines and can be determined from that for the corresponding LHSN rule.

Proposition 2 *Let X and Y be two itemsets. Then the two following properties hold.*

(1) ${\mathrm{M_{GK}}}^f(X{\rightarrow}\overline{Y}) = -{\mathrm{M_{GK}}}^{df}(X{\rightarrow}Y)$.
(2) ${\mathrm{M_{GK}}}^f(\overline{X}{\rightarrow}Y) = \frac{P(X')}{1-P(X')}\frac{P(Y')}{1-P(Y')}{\mathrm{M_{GK}}}^f(X{\rightarrow}\overline{Y})$.

Definition 3 *Let μ be a quality measure, and let $\alpha > 0$ be a positive real number. Let $X{\rightarrow}Y$ be a positive or a (right-hand side, left-hand side or both side) negative association rule. Then $X{\rightarrow}Y$ is said to be* valid *w.r.t. α in the sense of μ or, simply, (μ,α)*-valid *if $\mu(X{\rightarrow}Y) \geq \alpha$. When the meaning is clear from the context, we omit the validity threshold α and/or the quality measure μ, and talk about μ-valid or, simply, valid association rules.*

In the sequel, α will denote a minimum validity threshold belonging to the interval $]0,1[$. As a consequence of Proposition 2 above, LHSN $\mathrm{M_{GK}}$-valid rules can be obtained from RHSN ones, as stated in the next corollary.

Corollary 1 *If X and Y are two itemsets such that X disfavors Y, then* ${\mathrm{M_{GK}}}^f(\overline{X}{\rightarrow}Y) \geq \alpha$ *if and only if* ${\mathrm{M_{GK}}}^f(X{\rightarrow}\overline{Y}) \geq \alpha(\frac{1}{\mathrm{Supp}(X)} - 1)(\frac{1}{\mathrm{Supp}(Y)} - 1)$.

To summarize, we need to consider negative rules as well as positive ones. However, LHSN $\mathrm{M_{GK}}$-valid rules can be derived from RHSN ones w.r.t. a corresponding validity threshold, so that we can restrict ourselves to generate only RHSN $\mathrm{M_{GK}}$-valid rules. Moreover, as $\mathrm{M_{GK}}$ satisfies the logical principle of contraposition when the premise favors the consequent, the both side negative $\mathrm{M_{GK}}$-valid rules can also be derived from their corresponding positive $\mathrm{M_{GK}}$-valid ones w.r.t. the same validity threshold. Therefore, we will consider only positive rules and RHSN rules in the sequel. Hence, we will simply use the term negative rule to mean RHSN rule.

One of the major problems faced in association rule mining is the huge number of generated rules. Indeed, despite the fact that a (set of) quality measure(s) is used in order to capture only those rules meeting some given constraints, the set of generated rules can still be of a very large size, due to the presence of redundant and/or trivial rules. Indeed, for a given quality measure μ, the set of μ-valid association rules often contains many rules that are redundant in the sense that they can be derived from other μ-valid rules. For instance, if $\mathrm{Conf}(X\to Y) = 1$ and $\mathrm{Conf}(Y\to Z) = 1$, then $\mathrm{Conf}(X\to Z) = 1$. Thus, if we look for Confidence-exact association rules, *i.e.* rules whose confidence is equal to 1, then the rule $X\to Z$ is redundant when the rules $X\to Y$ and $Y\to Z$ are given, since it can be derived from these ones.

On the other hand, some rules are valid whatever the validity threshold is, and thus, are not informative at all. For instance, for any itemsets X and Y with $Y\subseteq X$, the rule $X\to Y$ is Confidence-exact. Therefore, if we are interested in informative Confidence-exact association rules, then the rules of the form $X\to Y$ with $Y\subseteq X$ are not worth generating.

A way to cope with redundant or non informative association rules without loss of information is to generate a basis for the set of valid rules. Indeed, a basis is a set of rules from which any valid rule can be derived using given inference axioms, and which is minimal (w.r.t. set inclusion) among the rule sets having this property. In this paper, we characterize bases for $\mathrm{M_{GK}}$-valid association rules of a binary context, in terms of the closure operator induced by a Galois connection. Thus, these bases can be derived from a Galois lattice, as do bases for positive Confidence-valid rules.

2.3 The Galois lattices of a binary context

The binary context $\mathbb{K}$ induces a Galois connection between the partially ordered sets $(\mathcal{P}(\mathcal{E}), \subseteq)$ and $(\mathcal{P}(\mathcal{V}), \subseteq)$ by means of the maps

$$f : X \mapsto \bigcap_{x\in X}\{v \in \mathcal{V} : v(x) = 1\} = X'$$

and

$$g : Y \mapsto \bigcap_{v\in Y}\{x \in \mathcal{E} : v(x) = 1\},$$

for $X \subseteq \mathcal{E}$ and $Y \subseteq \mathcal{V}$ [14]. Moreover, the Galois connection (f, g) induces, in turn, a closure operator $\varphi := f \circ g$ on $(\mathcal{P}(\mathcal{V}), \subseteq)$ [6]. That is, for $X, Y \subseteq \mathcal{V}$:

(C1) $X \subseteq \varphi(X)$ (extensivity);
(C2) $X \subseteq Y$ implies $\varphi(X) \subseteq \varphi(Y)$ (isotony);
(C3) $\varphi(\varphi(X)) = \varphi(X)$ (idempotence).

Let $G(\mathbb{K})$ denote the set of all pairs $(X, Y) \in \mathcal{P}(\mathcal{E}) \times \mathcal{P}(\mathcal{V})$ such that $\varphi(Y) = Y$ and $g(Y) = X$. Then $G(\mathbb{K})$, endowed with the order defined by $(X_1, Y_1) \leq (X_2, Y_2)$ if and only if $X_1 \subseteq X_2$ (or, equivalently $Y_2 \subseteq Y_1$), is a complete lattice called the *Galois lattice* of the binary context $\mathbb{K}$ [14], also known as the concept lattice of the formal context $(\mathcal{E}, \mathcal{V}, \mathcal{R})$, where $\mathcal{R}$ is the binary relation from $\mathcal{E}$ to $\mathcal{V}$ defined by $x\mathcal{R}v$ if and only if $v(x) = 1$ [15].

Example 1 *Consider the binary context $\mathbb{K}$ given in Table 1. Then, the pair $(\{e_2, e_3, e_5\}, \{B, C\})$ is a member of $G(\mathbb{K})$. But though $\varphi(\{B, C\}) = \{B, C\}$, $(\{e_2, e_3\}, \{B, C\})$ does not belong to $G(\mathbb{K})$ since $g(\{B, C\}) \neq \{e_2, e_3\}$.*

3 Bases for Confidence-valid association rules

This section is intended to remind two known bases for positive Confidence-valid association rules, namely, the Luxemburger basis for approximate rules and the Guigues-Duquenne basis for exact ones.

The set of positive Confidence-exact association rules is a full implicational system, *i.e.*, it satisfies the following Armstrong's inference axioms for all itemsets X, Y, Z [16]:

(PE1) $Y \subseteq X$ implies $X \rightarrow Y$;
(PE2) $X \rightarrow Y$ and $Y \rightarrow Z$ imply $X \rightarrow Z$;
(PE3) $X \rightarrow Y$ and $Z \rightarrow T$ imply $X \cup Z \rightarrow Y \cup T$.

Thus, the Guigues-Duquenne basis [7] for full implicational systems is by the way a basis for positive Confidence-exact rules. To define this basis, we need to recall the notion of a critical set of a closure operator.

Consider the closure operator φ, induced on $\mathcal{P}(\mathcal{V})$ by the Galois connection (f, g) defined above. An itemset X is said to be φ-*closed* if $\varphi(X) = X$; it is said to be φ-*quasi-closed* if it is not φ-closed and for all $Y \subset X$, either $\varphi(Y) \subset X$ or $X \subset \varphi(Y)$ [17]; it is said to be φ-*critical* if it is minimal among the φ-quasi-closed itemsets Y such that $\varphi(Y) = \varphi(X)$ [18]. A definition of quasi-closed sets in terms of Moore families can be found in [19–21], as well as other characterizations of $\varphi-$critical sets.

The Guigues-Duquenne basis [7] for positive Confidence-exact association rules is the set BPE defined by

$$\text{BPE} = \{X \rightarrow \varphi(X) \setminus X : X \text{ is } \varphi\text{-critical}\}.$$

This basis has been adapted to Support-and-Confindence-exact association rules by [22] and [23], who placed association rule mining problem within the theoretic framework of Galois lattices.

Example 2 *The rules $B \rightarrow E$ and $D \rightarrow AC$ are two rules belonging to* BPE*, from which many other positive* M_{GK}*-exact rules such as, for instance, $BD \rightarrow ACE$, $AB \rightarrow E$ and $AD \rightarrow ACE$ can be derived, using (PE1), (PE2) and (PE3).*

The Luxemburger basis [8] for Confidence-approximate association rules is the set LB defined by

$$\text{LB} = \{X \rightarrow Y : X = \varphi(X), Y = \varphi(Y), X \prec Y \text{ and } \text{Conf}(X \rightarrow Y) \geq \alpha\},$$

where $X \prec Y$ means that $X \subset Y$ and there is no φ-closed set Z such that $X \subset Z \subset Y$.

4 Bases for M_{GK}-valid association rules

In this section, we characterize a basis for $(\text{M}_{\text{GK}}, \alpha)$-valid association rules. This basis is in fact the union of four bases: a basis for positive exact rules (*i.e.* the rules $X \to Y$ such that $\mathrm{M_{GK}}(X \to Y) = 1$), a basis for negative exact rules (*i.e.* the rules $X \to \overline{Y}$ such that $\mathrm{M_{GK}}(X \to \overline{Y}) = 1$), a basis for positive approximate rules (*i.e.* the rules $X \to Y$ such that $\alpha \leq \mathrm{M_{GK}}(X \to Y) < 1$), and a basis for negative approximate rules (*i.e.* the rules $X \to \overline{Y}$ such that $\alpha \leq \mathrm{M_{GK}}(X \to \overline{Y}) < 1$).

4.1 Basis for positive $\mathbf{M_{GK}}$-exact association rules

The set of positive $\mathrm{M_{GK}}$-exact rules coincides with that of positive Confidence-exact ones. Thus, the basis BPE for Confidence-exact association rules is by the way a basis for positive $\mathrm{M_{GK}}$-exact rules.

4.2 Basis for negative $\mathbf{M_{GK}}$-exact association rules

Recall that negative association rules are rules of the form $X \to \overline{Y}$, where X and Y are itemsets. The following straightforward but instrumental properties define their support and confidence.

Proposition 3 *Let X and Y be two itemsets. Then the three following conditions hold.*

(1) $\mathrm{Supp}(\overline{X}) = 1 - \mathrm{Supp}(X)$.
(2) $\mathrm{Supp}(X \to \overline{Y}) = \mathrm{Supp}(X) - \mathrm{Supp}(X \to Y)$.
(3) $\mathrm{Conf}(X \to \overline{Y}) = 1 - \mathrm{Conf}(X \to Y)$.

Negative $\mathrm{M_{GK}}$-exact association rules are those negative rules $X \to \overline{Y}$ such that $\mathrm{M_{GK}}(X \to \overline{Y}) = 1$. The next easily-checked result characterizes them in terms of the support or the confidence of their corresponding positive rules.

Proposition 4 *Let X and Y be two itemsets such that* $\mathrm{Supp}(X) \neq 0$ *and* $\mathrm{Supp}(Y) \neq 0$. *Then the following conditions are equivalent:*

(1) $\mathrm{M_{GK}}(X \to \overline{Y}) = 1$. *(2)* $\mathrm{M_{GK}}(X \to Y) = -1$.
(3) $\mathrm{Conf}(X \to Y) = 0$. *(4)* $\mathrm{Supp}(X \to Y) = 0$.

In the sequel, for $x \in \mathcal{V}$ and $X, Y \subseteq \mathcal{V}$, we will sometimes denote $\{x\}$ by x, $X \cup Y$ by XY and $\{x\} \cup X$ by $x + X$. Proposition 4 leads us to consider the following inference axioms for any itemsets X, Y, Z:

(NE1) $X \to \overline{Y}$ and $\mathrm{Supp}(YZ) > 0$ imply $X \to \overline{YZ}$;
(NE2) $X \to \overline{Y}$, $Z \subset X$ and $\mathrm{Supp}(ZY) = 0$ imply $Z \to \overline{Y}$.

The next result shows that every association rule derived from negative $\mathrm{M_{GK}}$-exact ones using (NE1) and (NE2) is also negative $\mathrm{M_{GK}}$-exact.

Proposition 5 *The inference axioms (NE1) and (NE2) are sound for negative* $\mathrm{M_{GK}}$*-exact association rules.*

Proposition 4 also leads us to consider the positive border of the set of itemsets having a null support [24], *i.e.*, the set

$$\mathrm{Bd}^{+}(0) = \{X \subseteq \mathcal{V} : \mathrm{Supp}(X) > 0 \text{and for all } x \notin X [\mathrm{Supp}(x+X) = 0]\}$$

consisting of maximal itemsets (w.r.t. set inclusion) having a non null support.

Example 3 *For the context given in Table 1,* $\mathrm{Bd}^{+}(0) = \{ACD, BCE, ABCE\}$.

We now go on to characterize the basis we propose for the set of negative $\mathrm{M_{GK}}$-exact association rules.

Theorem 1 *The set* BNE *defined by*

$$\mathrm{BNE} = \{X \rightarrow \overline{x} : X \in Bd^{+}(0) \textit{ and } x \notin X\}$$

is a basis for negative $\mathrm{M_{GK}}$*-exact association rules w.r.t. the inference axioms (NE1) and (NE2).*

Example 4 *For the context given in Table 1,* $\mathrm{BNE} = \{ABCE \rightarrow \overline{D}, ACD \rightarrow \overline{B}, ACD \rightarrow \overline{E}, BCE \rightarrow \overline{A}, BCE \rightarrow \overline{D}\}$. *Moreover, the eleven rules* $ABCE \rightarrow \overline{D}$, $ABCE \rightarrow \overline{AD}$, $ABCE \rightarrow \overline{CD}$, $ABE \rightarrow \overline{ACD}$, $BE \rightarrow \overline{AD}$, $E \rightarrow \overline{AD}$, $B \rightarrow \overline{AD}$, $E \rightarrow \overline{CD}$, $B \rightarrow \overline{CD}$, $E \rightarrow \overline{ACD}$, $B \rightarrow \overline{ACD}$ *can be derived from the rule* $ABCE \rightarrow \overline{D}$*, using (NE1) and (NE2).*

It may be noticed that the positive border $\mathrm{Bd}^{+}(0)$ is nothing else than the set of maximal φ-closed itemsets having a strictly positive support. Thus, the basis BNE is clearly characterized in terms of the closure operator φ. It may also be noticed that $Y \rightarrow \overline{X}$ is a negative $\mathrm{M_{GK}}$-exact rule whenever $X \rightarrow \overline{Y}$ is. However these two rules are not always equally informative. Indeed, if, for instance, $|X_1| > |X_2| > |Y_1| > |Y_2|$, then the rule $X_2 \rightarrow \overline{Y_2}$ is more informative than any other negative rule involving the itemsets X_1, X_2, Y_1, Y_2.

4.3 Basis for positive $\mathrm{M_{GK}}$-approximate association rules

Positive $(\mathrm{M_{GK}}, \alpha)$-approximate association rules are those positive rules $X \rightarrow Y$ such that $\alpha \leq \mathrm{M_{GK}}(X \rightarrow Y) < 1$. The following straightforward result characterizes them in terms of their confidence.

Proposition 6 *Let X and Y be two itemsets such that X favors Y. Then* $\alpha \leq \mathrm{M_{GK}}(X \rightarrow Y) < 1$ *if and only if* $\mathrm{Supp}(Y)(1-\alpha) + \alpha \leq \mathrm{Conf}(X \rightarrow Y) < 1$.

This result leads us to consider the following inference axiom for any itemsets X, Y, Z, T:

(PA) $X \rightarrow Y$, $\varphi(X) = \varphi(Z)$ and $\varphi(Y) = \varphi(T)$ imply $Z \rightarrow T$.

The two following technical lemmas will be helpful for proving the soundness of the axiom (PA). The first lemma shows that every itemset has the same support as its φ-closure [25].

Lemma 1 *For any itemset* X*,* $\mathrm{Supp}(\varphi(X)) = \mathrm{Supp}(X)$.

The second lemma is a characterization of closure operators by means of path independence property [26, 21].

Lemma 2 *An extensive function* ϕ *on a finite powerset, say* $\mathcal{P}$*, is a closure operator on* $\mathcal{P}$ *if and only if it satisfies the path independence property:* $\phi(X \cup Y) = \phi(\phi(X) \cup \phi(Y))$*, for any* $X, Y \in \mathcal{P}$.

The next proposition shows that every association rule derived from a positive $\mathrm{M_{GK}}$-approximate one, using the inference axiom (PA), is also positive $\mathrm{M_{GK}}$-approximate.

Proposition 7 *The inference axiom (PA) is sound for positive* $(\mathrm{M_{GK}}, \alpha)$*-approximate association rules.*

We now go on to characterize the basis we propose for the set of positive $\mathrm{M_{GK}}$-approximate association rules.

Theorem 2 *The set* $\mathrm{BPA}(\alpha)$ *defined by*

$$\mathrm{BPA}(\alpha) = \{X{\rightarrow}Y : \varphi(X) = X, \varphi(Y) = Y, \mathrm{Supp}(Y)(1-\alpha)+\alpha \leq \mathrm{Conf}(X{\rightarrow}Y) < 1\}$$

is a basis for positive $(\mathrm{M_{GK}}, \alpha)$*-approximate association rules w.r.t. the inference axiom (PA).*

Example 5 *Consider the context given in Table 1 and let the minimum validity threshold* α *be set to* $\frac{1}{10}$*. Then, the rule* $AC{\rightarrow}BCE$ *is a member of* $\mathrm{BPA}(\alpha)$ *from which can be derived the five rules* $A{\rightarrow}BC$*,* $A{\rightarrow}CE$*,* $A{\rightarrow}BCE$*,* $AC{\rightarrow}BC$ *and* $AC{\rightarrow}CE$*, using the inference axiom (PA).*

4.4 Basis for negative $\mathrm{M_{GK}}$-approximate association rules

Negative $(\mathrm{M_{GK}}, \alpha)$-approximate association rules are those negative rules $X{\rightarrow}\overline{Y}$ such that $\alpha \leq \mathrm{M_{GK}}(X{\rightarrow}\overline{Y}) < 1$. The next straightforward result characterizes them in terms of the confidence of their corresponding positive rules.

Proposition 8 *Let* X *and* Y *be two itemsets such that* X *disfavors* Y*. Then* $\alpha \leq \mathrm{M_{GK}}(X{\rightarrow}\overline{Y}) < 1$ *if and only if* $0 < \mathrm{Conf}(X{\rightarrow}Y) \leq \mathrm{Supp}(Y)(1-\alpha)$.

This result leads us to consider the following inference axiom for any itemsets X, Y, Z, T:

(NA) $X{\rightarrow}\overline{Y}$, $\varphi(X) = \varphi(Z)$ and $\varphi(Y) = \varphi(T)$ imply $Z{\rightarrow}\overline{T}$.

The next result shows the soundness of the inference axiom (NA).

Proposition 9 *The inference axiom (NA) is sound for negative* $(\mathrm{M_{GK}}, \alpha)$*-approximate association rules.*

Theorem 3 below characterizes the basis we propose for the set of negative M_{GK}-approximate association rules.

Theorem 3 *The set* $\mathrm{BNA}(\alpha)$ *defined by*

$$\mathrm{BNA}(\alpha) = \{X \rightarrow \overline{Y} : \varphi(X) = X, \varphi(Y) = Y, 0 < \mathrm{Conf}(X \rightarrow Y) \leq \mathrm{Supp}(Y)(1-\alpha)\}$$

is a basis for negative (M_{GK}, α)*-approximate association rules w.r.t. the inference axiom (NA).*

Example 6 *Consider the context given in Table 1 and let the minimum validity threshold* α *be set to* $\frac{1}{10}$*. Then, the rule* $AC \rightarrow \overline{BE}$ *is a member of* $\mathrm{BNA}(\alpha)$ *from which can be derived the five rules* $A \rightarrow \overline{B}$, $A \rightarrow \overline{E}$, $A \rightarrow \overline{BE}$, $AC \rightarrow \overline{B}$ *and* $AC \rightarrow \overline{E}$*, using the inference axiom (NA).*

5 Conclusion

We reviewed the main properties of the quality measure for association rules, M_{GK}, independently introduced in [3] and in [4], and which has been shown to be the normalized quality measure associated to most of the quality measures proposed in the data mining literature [5]. On the other hand, we characterized bases for M_{GK}-valid association rules in terms of a closure operator induced by a Galois connection [6]: two bases for positive rules (exact and approximate) and two bases for negative rules (exact and approximate). Thus, these bases can be derived from a Galois lattice, as do well known bases for Confidence-valid association rules [7, 8].

References

1. Hilderman, R.J., Hamilton, H.J.: Knowledge discovery and interestingness measures: A survey. Technical Report CS 99-04, Department of Computer Science, University of Regina (1999)
2. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In Buneman, P., Jajodia, S., eds.: Proc. of the ACM SIGMOD International Conference on Management of Data. Volume 22., Washington, ACM press (1993) 207–216
3. Guillaume, S.: Traitement des données volumineuses. Mesures et algorithmes d'extraction des règles d'association et règles ordinales. PhD thesis, Université de Nantes, France (2000)
4. Wu, X., Zhang, C., Zhang, S.: Mining both positive and negative rules. ACM J. Information Systems **22** (2004) 381–405
5. Feno, D., Diatta, J., Totohasina, A.: Normalisée d'une mesure probabiliste de qualité des règles d'association : étude de cas. In: Actes du 2nd Atelier Qualité des Données et des Connaissances, Lille, France (2006) 25–30
6. Birkhoff, G.: Lattice theory. 3rd edition, Coll. Publ., XXV. American Mathematical Society, Providence, RI (1967)

7. Guigues, J.L., Duquenne, V.: Famille non redondante d'implications informatives résultant d'un tableau de données binaires. Mathématiques et Sciences humaines **95** (1986) 5–18
8. Luxemburger, M.: Implications partielles dans un contexte. Math. Inf. Sci. hum. **113** (1991) 35–55
9. Brin, S., Motwani, R., Ullman, J.D., Tsur, S.: Dynamic itemset counting and implication rules for market basket data. In: Proc. of the ACM SIGMOD Conference. (1997) 255–264
10. Lerman, I., Gras, R., Rostam, H.: Elaboration et évaluation d'un indice d'implication pour des données binaires. Math Sc. Hum **74** (1981) 5–35
11. Huynh, X., Guillet, F., Briand, H.: Une plateforme exploratoire pour la qualité des règles d'association : Apport pour l'analyse implicative. In: Troisièmes Rencontres Internationales A.S.I. (2005) 339–349
12. Brin, S., Motwani, R., Silverstein, C.: Beyond market baskets: Generalizing association rules to correlation. In: Proc. of the ACM SIGMOD Conference. (1997) 265–276
13. Totohasina, A., Ralambondrainy, H.: Ion: A pertinent new measure for mining information from many types of data. In: IEEE SITIS'05. (2005) 202–207
14. Barbut, M., Monjardet, B.: Ordre et classification. Hachette, Paris (1970)
15. Wille, R.: Restructuring lattice theory: an approach based on hierarchies of concepts. In Rival, I., ed.: Ordered sets. Ridel, Dordrecht-Boston (1982) 445–470
16. Armstrong, W.W.: Dependency structures of data base relationships. Information Processing **74** (1974) 580–583
17. Diatta, J.: Charactérisation des ensembles critiques d'une famille de Moore finie. In: Rencontres de la Société Francophone de Classification, Montréal, Canada (2005) 126–129
18. Day, A.: The lattice theory of functional dependencies and normal decompositions. Internat. J. Algebra Comput. **2** (1992) 409–431
19. Caspard, N.: A characterization theorem for the canonical basis of a closure operator. Order **16** (1999) 227–230
20. Caspard, N., Monjardet, B.: The lattices of closure systems, closure operators, and implicational systems on a finite set: a survey. Discrete Applied Mathematics **127** (2003) 241–269
21. Domenach, F., Leclerc, B.: Closure systems, implicational systems, overhanging relations and the case of hierarchical classification. Mathematical Social Sciences **47** (2004) 349–366
22. Zaki, M.J., Ogihara, M.: Theoretical Foundations of Association Rules. In: 3rd SIGMOD'98 Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD). (1998) 1–8
23. Pasquier, N., Bastide, Y., Taouil, R., Lakhal, L.: Closed set based discovery of small covers for association rules. In: Proc. 15emes Journees Bases de Donnees Avancees, BDA. (1999) 361–381
24. Mannila, H., Toivonen, H.: Levelwise search and borders of theories in knowledge discovery. Data Mining Knowledge Discovery **1** (1997) 241–258
25. Pasquier, N., Bastide, Y., Taouil, R., Lakhal, L.: Efficient mining of association rules using closed itemset lattices. Information Systems **24** (1999) 25–46
26. Plott, C.R.: Path independence, rationality and social choice. Econometrica **41** (1973) 1075–1091