

Obtention de règles optimisant un ensemble de mesures

Céline Hébert et Bruno Crémilleux

GREYC, CNRS - UMR 6072, Université de Caen
Campus Côte de Nacre
F-14032 Caen Cédex France
{Prenom.Nom}@info.unicaen.fr

Résumé : Lors d'un processus d'extraction de connaissances dans les bases de données, l'utilisateur est confronté à un nombre considérable de règles d'association. Des mesures objectives permettent alors d'effectuer une sélection parmi ces règles sans connaissance préalable du domaine. Cependant, le choix de la (ou des) mesure(s) à employer est délicat pour l'utilisateur. Dans cet article, nous définissons un ensemble de mesures optimisables (noté \mathcal{M}) dans le sens où ces mesures ont des comportements similaires (i.e., on peut les optimiser simultanément). \mathcal{M} est un ensemble infini regroupant beaucoup de mesures usuelles. Dans le cadre des règles de classe, nous montrons que toute mesure de \mathcal{M} admet un minorant que nous exprimons en fonction de la fréquence minimale et d'une borne pour le nombre d'exceptions. Ce résultat permet de garantir une valeur seuil sur la qualité des règles extraites. Nous identifions la collection complète de toutes les règles de classe optimisant les mesures de \mathcal{M} . Enfin, nous présentons brièvement un prototype qui permet d'extraire toutes les règles de classe informatives optimisant toutes les mesures de \mathcal{M} .

Mots-clés : Règles de classe, Mesures d'intérêt objectives, Optimisation multicritères.

1 Introduction

Explorer et analyser les relations existant au sein d'un ensemble de facteurs est au cœur de tout processus d'extraction de connaissances dans les bases de données (ECBD). La recherche de règles d'association telles que définies dans (Agrawal *et al.*, 1993) vise à découvrir les implications de la forme $X \rightarrow Y$ où X et Y peuvent représenter un ou plusieurs facteurs (aussi dénommés *attributs*). Cependant, dans la pratique, le nombre de règles extraites est souvent considérable et rend difficile voire impossible l'analyse et l'identification de celles qui sont réellement intéressantes (Guillaume, 2000). Il s'avère donc particulièrement utile d'aider l'utilisateur dans sa sélection de règles. Dans cet article, nous nous intéressons plus particulièrement aux règles de classe. Ces règles, qui concluent sur un attribut de classe, sont utilisées par exemple pour la production de motifs émergents (Dong & Li, 1999), de règles de caractérisation et de classification (Liu

et al., 1998; Crémilleux & Boulicaut, 2002) et sont à la base de la construction de nombreux modèles provenant de motifs locaux (i.e., associations entre les attributs) comme les classifieurs fondés sur les associations (Liu *et al.*, 1998; Baralis & Garza, 2003).

Différentes approches tentent de pallier le problème de la sélection de règles afin de cerner l'information la plus pertinente. Citons l'utilisation de templates pour donner des critères de filtrage sur les règles (Klemettinen *et al.*, 1994), la spécification de contraintes portant sur les prémisses ou conclusions des règles (Ng *et al.*, 1998), le regroupement de règles basées sur leur similarité afin de structurer leur présentation à l'utilisateur (Lent *et al.*, 1997), les couvertures qui éliminent une règle selon le contexte global dans lequel elle se trouve (Toivonen *et al.*, 1995; Zaki, 2000), l'utilisation de mesures d'intérêt ayant pour but de faire émerger les règles les plus significatives des autres (McGarry, 2005).

Cet article s'intéresse à l'utilisation des mesures d'intérêt pour les règles de classe. Il existe de nombreuses mesures d'intérêt. La fréquence et la confiance sont certainement les plus célèbres (Agrawal & Srikant, 1994), d'autres sont plus spécifiques comme le lift (Brin *et al.*, 1997) ou la mesure de Sebag et Schoenauer (Sebag & Schoenauer, 1988). Dans la pratique, le choix d'une mesure d'intérêt adéquate ainsi que la détermination éventuelle d'un seuil minimal pour celle-ci se révèlent souvent délicats. Combiner plusieurs mesures relève alors du défi. Il semble donc naturel de chercher à comparer plusieurs mesures afin de mieux comprendre leurs comportements et dégager leurs similarités et différences (Plasse *et al.*, 2006; Bayardo & Agrawal, 1999). Une réflexion sur le comportement des mesures basées sur la fréquence des parties d'une règle (en fait, la plupart des mesures utilisées) a été le point de départ de ce travail. Il nous a conduit à définir de façon formelle un ensemble de mesures nommé *ensemble des mesures optimisables*, noté \mathcal{M} , et ayant des comportements similaires (i.e., on peut les optimiser simultanément). Cette formalisation permet de donner pour toute mesure de \mathcal{M} un minorant en fonction de la fréquence minimale de la prémisse et du nombre maximal d'exceptions de la règle. Ce résultat garantit une valeur seuil sur la qualité des règles extraites. D'autre part, l'ensemble des règles de classe à prémisses fréquentes et ayant un nombre d'exceptions borné est équivalent à l'ensemble des règles de classe ayant pour chaque mesure une valeur supérieure à son minorant. Ce résultat identifie la collection complète de toutes les règles de classe optimisant les mesures de \mathcal{M} .

L'article est organisé de la manière suivante. La section 2 est consacrée à un rapide état de l'art et aux définitions. Dans la section 3, nous définissons les *mesures optimisables*. La section 4 expose les conditions assurant que chaque mesure optimisable admet un minorant et caractérise l'ensemble des règles optimisant toutes les mesures de \mathcal{M} . Nous montrons dans la section 5 comment extraire efficacement les règles de classe informatives optimisant toutes les mesures de \mathcal{M} .

2 Contexte et motivations

2.1 Travaux relatifs

L'extraction de toutes les règles d'association conduit à de nombreuses règles redondantes. Ce constat a motivé la définition de règles non redondantes (Zaki, 2000) appelées *règles informatives* (Bastide *et al.*, 2002). Parmi les règles de mêmes support

et confiance, ces règles sont caractérisées par une prémisse minimale et une conclusion maximale. Elles ne font pas perdre d'information et forment une couverture¹ de toutes les règles d'association quantifiées par le support et la confiance. Nous verrons à la section 5 qu'il est possible de générer efficacement l'ensemble de toutes les règles informatives.

En parallèle à cette approche visant à éliminer la redondance, de nombreux travaux traitent de la sélection des meilleures règles au moyen de mesures d'intérêt. Les mesures objectives, que nous considérons dans cet article, sont basées sur des propriétés intrinsèques aux données. Afin de faciliter la tâche de l'utilisateur final, il est essentiel de (1) pouvoir définir ce qu'est une bonne mesure, (2) comparer les différentes mesures existantes et de dégager les particularités de chacune d'entre elles. Concernant le premier point, il s'agit de traduire de manière formelle des attentes de l'utilisateur ou des arguments relevant du bon sens. De nombreux travaux définissent des propriétés pour évaluer les mesures (Hilderman & Hamilton, 2003; Lallich & Teytaud, 2004). L'un des cadres théoriques les plus utilisés est celui de Piatetsky-Shapiro (Piatetsky-Shapiro, 1991) et nous situons notre travail par rapport à celui-ci à la section 2.3. Pour le deuxième point, on distingue des approches expérimentales (Vaillant *et al.*, 2004) ou théoriques (Tan *et al.*, 2002). Une alternative à ces points de vue consiste à combiner plusieurs mesures (optimisation multi-critères (Francisci & Collard, 2003)).

Notre travail aborde ces différents aspects. D'une part, nous définissons un cadre formel (l'ensemble des mesures optimisables \mathcal{M}) précisant les propriétés d'une mesure de \mathcal{M} . D'autre part, en montrant qu'il est possible d'optimiser simultanément toutes les mesures relevant de ce cadre, nous combinons les qualités apportées par celles-ci. Remarquons toutefois que nous n'avons pas cherché à définir des propriétés caractéristiques d'une qualité "idéale" d'une mesure mais que nous avons privilégié la mise en évidence des propriétés minimales que doit posséder une mesure pour appartenir à \mathcal{M} . Comme toutes les mesures de \mathcal{M} ont un comportement similaire, cela signifie que la question du choix d'une mesure de \mathcal{M} devient secondaire pour l'utilisateur.

2.2 Définitions et notations

Une base de données notée \mathcal{D} est décrite par un triplet $(\mathcal{A}, R, \mathcal{O})$ où \mathcal{A} est l'ensemble des *attributs*, \mathcal{O} est l'ensemble des *objets* et R est une relation binaire définie sur $(\mathcal{A}, \mathcal{O})$ par $\forall a \in \mathcal{A}, \forall o \in \mathcal{O} \quad aRo \Leftrightarrow (a, o) = 1$ dans \mathcal{D} . Un *motif* est un ensemble d'attributs de \mathcal{A} . La *fréquence* d'un motif X correspond au nombre d'objets de \mathcal{D} contenant X ; elle est notée $\mathcal{F}(X)$. Soit $\mathcal{C} = \{c_1, \dots, c_n\}$ un ensemble de valeurs de classe. À chaque objet de \mathcal{D} est associée une valeur de classe c_i . Pour toute classe c_i , \mathcal{D}_i représente la sous-base constituée des objets étiquetés par c_i . Le tableau 1 donne un exemple de base de données booléennes dont les objets sont étiquetés par deux valeurs de classe c_1 et c_2 . Dans cet exemple, les classes sont équilibrées mais il s'agit d'un cas particulier.

¹A contrario, certaines couvertures structurelles (Toivonen *et al.*, 1995) s'entendent en ignorant les confiances des règles extraites, ce qui peut conduire à l'élimination de règles ayant une forte confiance.

\mathcal{D}	Attributs								Classes		
	Objets	A	B	C	D	E	F	G	H	c_1	c_2
o_1	1	0	1	0	1	0	1	0	1	0	\mathcal{D}_1
o_2	0	1	1	0	1	0	1	1	1	0	
o_3	1	0	1	0	1	0	0	1	1	0	
o_4	1	0	1	0	1	0	0	1	1	0	
o_5	0	1	1	0	1	1	0	0	0	1	\mathcal{D}_2
o_6	1	0	0	1	0	1	0	1	0	1	
o_7	0	1	1	0	1	1	0	1	0	1	
o_8	1	0	1	0	0	1	0	1	0	1	

TAB. 1 – Un exemple \mathcal{D} de base de données booléennes.

Règles de classe

La définition suivante précise la notion de règle de classe (rappelons que le nombre de classes est quelconque).

Définition 1 (Règle de classe)

Une règle de la forme $r : X \rightarrow c_i$ où X est un motif d'attributs et c_i une valeur de classe est appelée règle de classe. X est la prémisse de r et c_i sa conclusion.

Par exemple, $r_1 : F \rightarrow c_2$ et $r_2 : EH \rightarrow c_1$ sont deux règles de classe issues de la base de données décrite par le tableau 1.

La fréquence d'une règle r est notée $\mathcal{F}(Xc_i)$ et la fréquence de sa prémisse $\mathcal{F}(X)$.

Définition 2 (Nombre d'exceptions)

La quantité $\mathcal{F}(X) - \mathcal{F}(Xc_i)$ est appelée nombre d'exceptions de la règle r .

La règle r_2 admet 1 exception (sur l'objet o_7) et la fréquence de sa prémisse est 4. Remarquons que l'expression $\mathcal{F}(X) - \mathcal{F}(Xc_i)$ dénombre en fait les objets de \mathcal{D} pour lesquels la prémisse de r ne conclut pas sur c_i . Par la suite, nous verrons que la fréquence de la prémisse d'une règle et son nombre d'exceptions sont deux paramètres fondamentaux pour l'évaluation de la qualité d'une règle de classe.

2.3 Qualité d'une mesure

De manière générale, une *mesure d'intérêt* est une fonction qui à une règle $X \rightarrow Y$ associe une valeur. Nous considérons ici que la valeur de la mesure croît quand la qualité de la règle augmente, ce qui est la convention la plus courante. Nous précisons maintenant le cadre de G. Piatetsky-Shapiro (Piatetsky-Shapiro, 1991) destiné à évaluer la qualité d'une mesure. Nous exprimons les trois propriétés de ce cadre dans le cas des règles de classe puisque c'est l'objet de cet article. Ces propriétés ont été initialement énoncées sous forme probabiliste, nous les avons ici traduites en terme de fréquence puisque c'est ainsi que nous les utiliserons par la suite.

Définition 3 (Propriétés de Piatetsky-Shapiro)

Soient $r : X \rightarrow c_i$ une règle de classe et M une mesure d'intérêt.

- P1 : $M(r) = 0$ si X et c_i apparaissent indépendamment l'un de l'autre dans \mathcal{D} ;
- P2 : Lorsque $\mathcal{F}(X)$ et $|\mathcal{D}_i|$ sont fixés, $M(r)$ croît strictement quand $\mathcal{F}(Xc_i)$ croît ;
- P3 : Lorsque $\mathcal{F}(Xc_i)$ et $\mathcal{F}(X)$ (resp. $|\mathcal{D}_i|$) sont fixés, $M(r)$ décroît strictement quand $|\mathcal{D}_i|$ (resp. $\mathcal{F}(X)$) croît.

P1 signifie que la valeur d'une mesure est nulle lorsque la valeur de classe est indépendante de la prémisse de la règle. P2 assure que M croît strictement avec la fréquence de la règle, ce qui traduit la prise en compte de la "représentativité" d'une règle. Enfin, P3 exprime la "force" entre la prémisse et la conclusion de la règle par rapport à la proportion d'objets supportant la règle.

Remarquons que beaucoup de mesures usuelles satisfont P2 (on trouvera une définition de la plupart des mesures dans (Tan *et al.*, 2002)) : le support, la confiance, l'intérêt, la conviction, etc... Il existe cependant quelques exceptions telles que la J-mesure, la mesure de Goodman-Kruskal et l'index de Gini. Dans la section suivante, nous positionnons notre cadre par rapport à celui de G. Piatetsky-Shapiro.

3 Ensemble des mesures optimisables

Nous définissons dans cette section l'ensemble des mesures optimisables. L'idée clé est d'exprimer les mesures en fonction de deux paramètres : la fréquence de la prémisse d'une règle et son nombre d'exceptions. Dans la suite de l'article, γ est la fréquence minimale pour la prémisse d'une règle et δ est un entier positif qui va être utilisé pour borner le nombre d'exceptions d'une règle. Dans cette section, nous modélisons une mesure par une fonction continue et nous lui associons une fonction qui traduit l'influence de δ dans son comportement. Cette formalisation permet d'obtenir les résultats de la section 4.

3.1 δ -dépendance

Pour un jeu de données fixé, la définition 4 indique qu'une mesure peut être considérée comme une fonction continue à deux variables : la fréquence de la règle et celle de sa prémisse.

Définition 4

Soient $r : X \rightarrow c_i$ une règle de classe et M une mesure d'intérêt. $\Psi_M(x, y)$ est la fonction continue associée à $M(r)$ où $x = \mathcal{F}(X)$ et $y = \mathcal{F}(Xc_i)$.

Par exemple, pour le Taux de Croissance, $\Psi_{TC}(x, y) = \frac{y}{x-y} \times \frac{|\mathcal{D}|-|\mathcal{D}_i|}{|\mathcal{D}_i|}$ et pour le Gain Informationnel $\Psi_{GI}(x, y) = \log\left(\frac{y}{x} \times \frac{|\mathcal{D}|}{|\mathcal{D}_i|}\right)$.

Les propriétés P2 et P3 du cadre de Piatetsky-Shapiro (cf. section 2.3) peuvent s'interpréter à partir de Ψ_M : P2 signifie que "la fonction Ψ_M est strictement croissante suivant y " et P3 que "la fonction Ψ_M est strictement décroissante suivant x ".

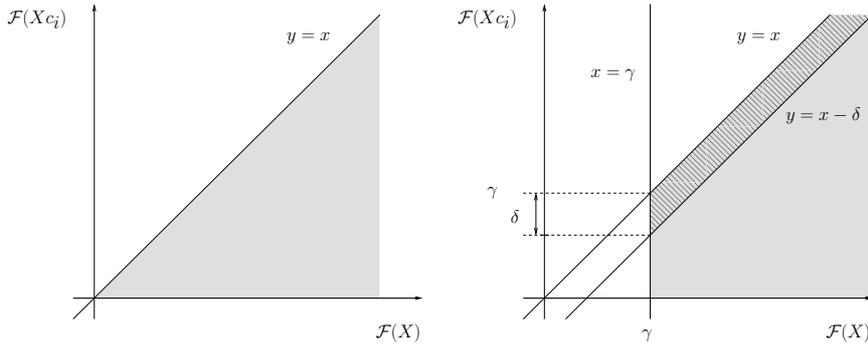


FIG. 1 – Domaines des variables $x = \mathcal{F}(X)$ et $y = \mathcal{F}(Xc_i)$.

La figure 1 illustre graphiquement le lien entre les deux variables $x = \mathcal{F}(X)$ et $y = \mathcal{F}(Xc_i)$ de Ψ_M pour une règle de classe $r : X \rightarrow c_i$. Cette figure indique la fréquence $\mathcal{F}(Xc_i)$ en fonction de celle de la prémisse (i.e., $\mathcal{F}(X)$). Sur la figure de gauche, le domaine en gris symbolise les valeurs possibles pour x et y : x et y sont positives puisque ce sont des fréquences et on a $x \geq y$ car $\mathcal{F}(X) \geq \mathcal{F}(Xc_i)$. Sur la figure de droite, la zone grisée est réduite : cela montre la restriction imposée aux variables en fixant un seuil minimal γ pour la fréquence de la prémisse de r . De plus, la zone hachurée, de largeur δ , illustre le lien existant entre x et y lorsque le nombre d'exceptions de r est borné par δ (borner ce nombre d'exceptions signifie que $\mathcal{F}(Xc_i)$ reste proche de $\mathcal{F}(X)$). Les tailles respectives de ces deux zones illustrent bien la diminution du nombre de règles due à l'introduction de cette *dépendance* entre variables.

La définition de la fonction $\Psi_{M,\delta}$ traduit justement cette dépendance entre les variables x et y dans l'expression même de la fonction associée à la mesure M . Cette dépendance est liée à δ d'où le nom de fonction δ -dépendante.

Définition 5 (Fonction δ -dépendante)

Soit M une mesure d'intérêt. $\Psi_{M,\delta}$ est la fonction à une variable obtenue par le changement de variable $y = x - \delta$ dans Ψ_M i.e., $\Psi_{M,\delta}(x) = \Psi_M(x, x - \delta)$. $\Psi_{M,\delta}$ est appelée fonction δ -dépendante associée à M .

En continuant l'exemple sur le Taux de Croissance et le Gain Informationnel, on obtient les fonctions δ -dépendantes suivantes :

$$\Psi_{TC,\delta}(x) = \frac{x - \delta}{\delta} \times \frac{|\mathcal{D}| - |\mathcal{D}_i|}{|\mathcal{D}_i|} \text{ et } \Psi_{GI,\delta}(x) = \log \left(\frac{x - \delta}{x} \times \frac{|\mathcal{D}|}{|\mathcal{D}_i|} \right)$$

3.2 Croissance liée et mesures optimisables

La fonction $\Psi_{M,\delta}$ formalise le lien entre la fréquence de la prémisse et la fréquence de la règle. Nous pensons que ce lien est une caractéristique importante pour une mesure. C'est pourquoi nous proposons une nouvelle propriété, nommée *propriété de croissance liée* et notée P4, qui capture ce lien. Celui-ci est d'autant meilleur que l'écart entre la

fréquence de la prémisse et celle de la règle est faible, relativement à la fréquence de la prémisse. Aussi, la propriété P4 impose qu'une mesure M soit croissante suivant la variable x (i.e., fréquence de la prémisse) de $\Psi_{M,\delta}$.

Définition 6 (P4 : Propriété de croissance liée)

Une mesure M est dite à croissance liée si $\Psi_{M,\delta}$ est croissante.

P4 ne donne pas des conditions de variation directement sur les variables intervenant dans le calcul de la mesure M mais sur une fonction associée à M après changement de variable. Nous pensons que P4 permet d'exprimer une caractéristique originale et importante d'une mesure : l'évolution de M relativement à l'évolution conjointe des deux variables fréquence de la prémisse et nombre d'exceptions. Cette caractéristique n'apparaît pas dans les propriétés du cadre de Piatetsky-Shapiro (cf. section 2.3). Une autre différence avec P2 et P3 est que P4 n'impose pas une croissance stricte de $\Psi_{M,\delta}$ (la conviction est ainsi une mesure optimisable (cf. définition 7)).

Si Ψ_M est croissante à la fois en x et en y , alors M vérifie nécessairement P4. Cependant, la réciproque est fautive. Par exemple, la confiance vérifie P4 (i.e., $\Psi_{Conf,\delta}(x) = 1 - \frac{\delta}{x}$ est croissante) mais $\Psi_{Conf}(x, y) = \frac{y}{x}$ est décroissante en x .

Le tableau 2 indique un échantillon de mesures satisfaisant P4. La plupart de ces mesures proviennent de (Tan *et al.*, 2002; Vaillant *et al.*, 2005).

Mesure	Définition pour une règle de classe
Support	$\frac{\mathcal{F}(Xc_i)}{ \mathcal{D} }$
Confidence	$\frac{\mathcal{F}(Xc_i)}{\mathcal{F}(X)}$
Sensitivity	$\frac{\mathcal{F}(X)}{\mathcal{F}(Xc_i)}$
Specificity	$1 - \frac{ \mathcal{D}_i }{\mathcal{F}(X) - \mathcal{F}(Xc_i)}$
Success Rate	$\frac{ \mathcal{D} - \mathcal{D}_i - \mathcal{F}(X) + 2\mathcal{F}(Xc_i)}{ \mathcal{D} - \mathcal{D}_i - \mathcal{F}(X) + 2\mathcal{F}(Xc_i)}$
Lift	$\frac{ \mathcal{D} \times \mathcal{F}(Xc_i)}{ \mathcal{D}_i \times \mathcal{F}(X)}$
Piatetsky-Shapiro	$\frac{\mathcal{F}(Xc_i) - \mathcal{D}_i \times \mathcal{F}(X)}{ \mathcal{D} - \mathcal{D}_i ^2}$
Laplace (k=2)	$\frac{\mathcal{F}(Xc_i) + 1}{\mathcal{F}(X) + 2}$
Odds ratio	$\frac{\mathcal{F}(Xc_i)}{\mathcal{F}(X) - \mathcal{F}(Xc_i)} \times \frac{ \mathcal{D}_i - \mathcal{F}(Xc_i)}{ \mathcal{D}_i - \mathcal{D}_i }$
Taux de croissance	$\frac{\mathcal{F}(Xc_i)}{\mathcal{F}(X) - \mathcal{F}(Xc_i)} \times \frac{ \mathcal{D}_i }{ \mathcal{D}_i }$
Sebag & Schoenauer	$\frac{\mathcal{F}(X) - \mathcal{F}(Xc_i)}{\mathcal{F}(Xc_i)}$
Jaccard	$\frac{ \mathcal{D}_i + \mathcal{F}(X) - \mathcal{F}(Xc_i)}{\mathcal{F}(X)}$
Conviction	$\frac{ \mathcal{D} - \mathcal{D}_i }{ \mathcal{D} } \times \frac{\mathcal{F}(X) - \mathcal{F}(Xc_i)}{\mathcal{F}(X)}$
ϕ -coefficient	$\frac{ \mathcal{D} \times \mathcal{F}(Xc_i) - \mathcal{D}_i \times \mathcal{F}(X)}{\sqrt{\mathcal{F}(X) \times \mathcal{D}_i \times (\mathcal{D} - \mathcal{F}(X)) \times (\mathcal{D} - \mathcal{D}_i)}}$
Added Value	$\frac{\mathcal{F}(Xc_i) - \mathcal{D}_i }{\mathcal{F}(X) - \mathcal{D} }$
Certainty Factor	$\frac{\mathcal{F}(Xc_i) \times (\mathcal{D} - \mathcal{F}(X))}{\mathcal{F}(Xc_i) \times \mathcal{D} - \mathcal{F}(X) \times \mathcal{D}_i }$
Gain Informationnel	$\log \left(\frac{\mathcal{F}(Xc_i)}{\mathcal{F}(X)} \times \frac{ \mathcal{D} }{ \mathcal{D}_i } \right)$

TAB. 2 – Exemples de mesures d'intérêt vérifiant P4

Nous donnons maintenant la définition d'une mesure optimisable :

Définition 7 (\mathcal{M} : Ensemble des mesures optimisables)

L'ensemble des mesures optimisables \mathcal{M} est défini comme l'ensemble des mesures satisfaisant P2 et P4.

La définition 7 n'impose pas qu'une mesure optimisable vérifie aussi P1 et P3 même si cela n'est pas exclu. L'objectif est de définir l'ensemble le plus général possible de manière à ce que beaucoup de mesures usuelles appartiennent à \mathcal{M} . Par exemple, toutes les mesures présentées dans le tableau 2 sont dans \mathcal{M} . De part sa définition en intension, \mathcal{M} est un ensemble infini. Nous verrons dans la section 4.1 comment combiner P2 et P4 permet de minorer simplement chaque mesure optimisable.

4 Optimisation simultanée des mesures de \mathcal{M}

4.1 Minoration des mesures de \mathcal{M}

Nous prouvons maintenant que toute mesure de \mathcal{M} admet un minorant dépendant de γ et δ et nous l'exprimons en fonction de ces paramètres. Rappelons que le nombre total d'objets (i.e., $|\mathcal{D}|$) et la taille des sous-bases (i.e., $|\mathcal{D}_i|$) sont des constantes.

Théorème 1 (Minoration)

Soit $r : X \rightarrow c_i$ une règle de classe. Si $\mathcal{F}(X) \geq \gamma$ et si le nombre d'exceptions de r est borné par δ , alors $M(r)$ est supérieure à $\Psi_{M,\delta}(\gamma)$ pour toute mesure M de \mathcal{M} .

Preuve 1

Comme r admet moins de δ exceptions, on a $\mathcal{F}(Xc_i) \geq \mathcal{F}(X) - \delta$. D'après P2, $\Psi_M(x, y)$ croît avec y donc $\Psi_M(x, y) \geq \Psi_M(x, x - \delta) = \Psi_{M,\delta}(x)$. x est minoré par γ (seuil de fréquence de la prémisse de la règle) et puisque M satisfait P4 (croissance liée), alors $\Psi_{M,\delta}(x)$ est minoré par $\Psi_{M,\delta}(\gamma)$.

Ce théorème signifie que toute règle dont la prémisse est fréquente et dont le nombre d'exceptions est inférieur à δ a une qualité supérieure ou égale à $\Psi_{M,\delta}(\gamma)$. Comme ce résultat est valable pour toute mesure M , une conséquence immédiate de ce théorème est qu'une telle règle a une qualité suivant chaque mesure supérieure au minorant de la mesure. On dispose ainsi d'un réservoir de règles de bonne qualité par rapport à l'ensemble de mesures \mathcal{M} .

Les propriétés P2 et P4 sont nécessaires à l'établissement de la preuve 1. Sans P2, on ne peut pas minorer $\Psi_M(x, y)$ en utilisant le fait que le nombre d'exceptions d'une règle est borné et sans P4, il est impossible d'exploiter le seuil minimal de fréquence γ pour obtenir un minorant. Inversement, si $\Psi_{M,\delta}$ était décroissante, on obtiendrait alors un majorant pour chaque mesure de \mathcal{M} , ce qui n'aurait que peu d'intérêt.

Si l'on considère des mesures telle que $\Psi_M(x, y)$ est décroissante en y (contrairement aux mesures satisfaisant P2), il est possible de les minorer par $\Psi_M(x, x)$ puisqu'on a toujours $x \geq y$. Cependant, peu de mesures vérifient cette propriété (c'est pourquoi

nous n'envisageons pas ce cas) et il faudrait alors définir une nouvelle propriété pour assurer la croissance de la fonction $\Psi_M(x, x)$.

Outre les caractéristiques fixes du jeu de données, le minorant d'une mesure ne dépend que de δ et γ . On peut donc le calculer et ainsi quantifier précisément la qualité minimale des règles vérifiant le théorème 1. La deuxième colonne du tableau 3 indique ces minorants théoriques en fonction de γ et δ pour les dix-sept mesures de \mathcal{M} données au tableau 2. En considérant que $\gamma = 4$ et $\delta = 1$, les colonnes 3, 4 et 5 du tableau 3 indiquent les valeurs numériques des minorants et des mesures pour les règles r_1 et r_2 (cf. section 2.2) à partir de l'exemple du tableau 1. Comme ces deux règles vérifient les conditions du théorème 1, leur qualité est bien sûr supérieure ou égale au minorant de chaque mesure.

Notons que la valeur du minorant d'une mesure peut dépendre de la classe (pour deux règles concluant sur deux classes différentes mais avec même fréquence et même nombre d'exceptions, on s'attend à des qualités différentes suivant le nombre d'objets de la classe). Si on souhaite obtenir un minorant quelle que soit la classe, une solution immédiate est de prendre le minimum de l'ensemble des minorants. Évidemment, si les valeurs de classes sont équilibrées (comme pour les données du tableau 1), le minorant est le même pour toutes les classes.

Mesure	Minorant théorique $\Psi_{M,\delta}(\gamma)$	Exemple tableau 1		
		$\Psi_{M,1}(4)$	r_1	r_2
Support	$\frac{\gamma - \delta}{ \mathcal{D} }$	0.375	0.5	0.375
Confidence	$1 - \frac{\delta}{\gamma}$	0.75	1	0.75
Sensitivity	$\frac{\gamma - \delta}{ \mathcal{D}_i }$	0.75	1	0.75
Specificity	$1 - \frac{\delta}{ \mathcal{D} - \mathcal{D}_i }$	0.75	1	0.75
Success Rate	$1 + \frac{\gamma - 2\delta - \mathcal{D}_i }{ \mathcal{D} }$	0.75	1	0.75
Lift	$(1 - \frac{\delta}{\gamma}) \times \frac{ \mathcal{D} }{ \mathcal{D}_i }$	1.5	2	1.5
Piatetsky-Shapiro	$[\gamma \times (1 - \frac{ \mathcal{D}_i }{ \mathcal{D} }) - \delta] \times \frac{1}{ \mathcal{D} }$	0.125	0.25	0.125
Laplace (k=2)	$\frac{\gamma - \delta + 1}{\gamma + 2}$	$\frac{2}{3}$	$\frac{5}{6}$	$\frac{2}{3}$
Odds ratio	$[\frac{\gamma - \delta}{ \mathcal{D}_i - \gamma + \delta}] \times [\frac{ \mathcal{D} - \mathcal{D}_i - \delta}{ \mathcal{D} }]$	9	∞	9
Taux de croissance	$\frac{\gamma - \delta}{\delta} \times \frac{ \mathcal{D} - \mathcal{D}_i }{ \mathcal{D}_i }$	3	∞	3
Sebag & Schoenauer	$\frac{\gamma - \delta}{\delta}$	3	∞	3
Jaccard	$\frac{\gamma - \delta}{ \mathcal{D}_i + \delta}$	$\frac{3}{5}$	1	$\frac{3}{5}$
Conviction	$\frac{ \mathcal{D} - \mathcal{D}_i }{ \mathcal{D} } \times \frac{\gamma}{\delta}$	2	∞	2
ϕ -coefficient	$\frac{\gamma \times (\mathcal{D} - \mathcal{D}_i) - \delta \times \mathcal{D} }{\sqrt{\gamma \times (\mathcal{D} - \gamma) \times \mathcal{D}_i \times (\mathcal{D} - \mathcal{D}_i)}}$	$\frac{3}{4}$	1	$\frac{3}{4}$
Added Value	$\frac{\gamma - \delta - \mathcal{D}_i }{\gamma}$	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{1}{4}$
Certainty Factor	$\frac{\gamma \times (\mathcal{D} - \mathcal{D}_i) - \delta \times \mathcal{D} }{\gamma \times (\mathcal{D} - \mathcal{D}_i)}$	$\frac{3}{4}$	1	$\frac{3}{4}$
Gain Informationnel	$\log\left(\frac{\gamma - \delta}{\gamma} \times \frac{ \mathcal{D} }{ \mathcal{D}_i }\right)$	$\log\left(\frac{3}{2}\right)$	$\log(2)$	$\log\left(\frac{3}{2}\right)$

TAB. 3 – Minorants pour dix-sept mesures optimisables.

4.2 Valeur optimale des minorants

Intuitivement, plus γ augmente et plus δ diminue, plus on s'attend à ce qu'une règle ait une qualité élevée. La propriété 1 formalise cette intuition.

Propriété 1

$\Psi_{M,\delta}(\gamma)$ est une fonction croissante en γ et décroissante en δ .

Preuve 2

D'après P4, $\Psi_{M,\delta}(x)$ est croissante donc $\Psi_{M,\delta}(\gamma)$ est croissante en γ .

D'après P2, $\Psi_M(x, y)$ est croissante en y donc si $\delta_1 \geq \delta_2$ alors $\Psi_M(x, x - \delta_2) \geq \Psi_M(x, x - \delta_1)$ d'où $\Psi_{M,\delta_2}(\gamma) \geq \Psi_{M,\delta_1}(\gamma)$.

La valeur du minorant augmente quand γ croît et δ diminue. Le corollaire 1 précise qu'on peut ainsi faire tendre le minorant vers la valeur maximale de la mesure ce qui conduit à optimiser celle-ci.

Corollaire 1

$\Psi_{M,\delta}(\gamma)$ tend vers la valeur maximale de M quand $\gamma \rightarrow |\mathcal{D}|$ et quand $\delta \rightarrow 0$.

De façon pratique, ce corollaire signifie qu'il est possible de fixer γ et δ afin d'obtenir un minorant d'aussi bonne qualité que souhaitée. Par exemple, quand δ est égal à 0, la Confiance et la Spécificité valent 1. Si de plus $\gamma \rightarrow |\mathcal{D}|$, le Support tend vers 1 et la Sensibilité vers $\frac{|\mathcal{D}_1|}{|\mathcal{D}|}$ (i.e. sa valeur optimale).

La figure 2 précise les variations du minorant en fonction de γ et δ pour les mesures \mathcal{M} du tableau 2. Pour ces simulations, le nombre d'objets de \mathcal{D} est 100 et il y a deux classes équilibrées, le minorant obtenu est donc le même pour les deux classes. Ces courbes montrent les vitesses de croissance des mesures en fonction de γ et de décroissance en fonction de δ , elles permettent aussi de comparer les mesures entre elles et par rapport à leur valeur optimale.

Les courbes de gauche donnent l'évolution du minorant quand δ est fixé à 3 et quand γ varie entre 20 et 50. On remarque des comportements globalement linéaires en γ sur les trois premières courbes. Par contre, l'Odds ratio accuse une croissance rapide par rapport à γ à partir d'un seuil de 45 environ d'où la nécessité de fixer un seuil de fréquence suffisamment élevé pour obtenir des règles de classe ayant une bonne valeur pour cette mesure. Pour les courbes de droite, γ est fixé à 20 et δ varie de 0 à 20 (dans la pratique, on choisit δ et γ tel que $\delta < \gamma$ afin que la règle ait un sens, le nombre d'exceptions ne peut pas être supérieur au nombre d'objets vérifiant la prémisse). On distingue deux types de comportements : linéaire en δ pour les deux premières courbes, inversement proportionnel à δ pour les deux dernières. On peut déduire de ce deuxième cas que seul un très petit nombre d'exceptions permet de garantir une bonne qualité des règles de classe relativement au Taux de Croissance, à la Conviction et à l'Odds Ratio (décroissance très rapide en fonction de δ).

Optimisation de règles d'association

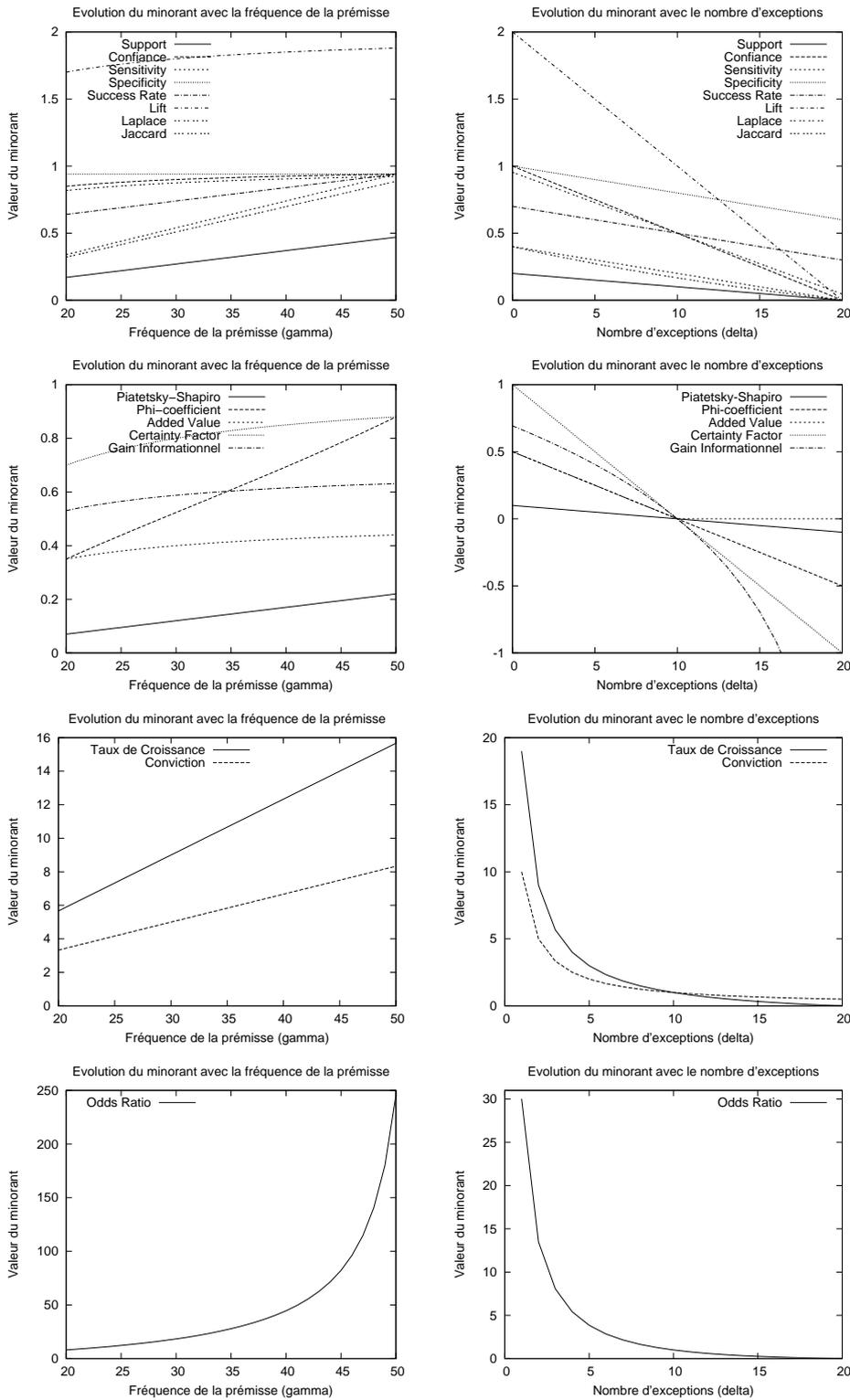


FIG. 2 – Variation des minorants en fonction de γ (gauche) et de δ (droite).

4.3 Exhaustivité de l'approche

Le théorème 2 indique que toute règle de classe r qui optimise l'ensemble des mesures \mathcal{M} (i.e., $\forall M, M(r) \geq \Psi_{M,\delta}(\gamma)$) est une règle dont la prémisse est fréquente et admettant au plus δ exceptions.

Théorème 2 (Exhaustivité)

Soit $r : X \rightarrow c_i$ une règle de classe. Si pour toute mesure M de \mathcal{M} , $M(r) \geq \Psi_{M,\delta}(\gamma)$ alors $\mathcal{F}(X) \geq \gamma$ et r admet au plus δ exceptions dans \mathcal{D} .

Preuve 3

Démontrons le théorème 2 en raisonnant par l'absurde. Notons Spe la mesure de Spécificité et Sup le Support.

Supposons que r admet δ' exceptions avec $\delta' > \delta$. Alors on a $-\frac{\delta'}{|\mathcal{D}|-|\mathcal{D}_i|} < -\frac{\delta}{|\mathcal{D}|-|\mathcal{D}_i|}$ et donc $Spe(r) < \Psi_{Spe,\delta}(\gamma)$ ce qui contredit notre hypothèse. Donc r admet au plus δ exceptions.

Supposons que $\mathcal{F}(X) = \gamma' < \gamma$. Alors $Sup(r) < \frac{\gamma'-\delta}{|\mathcal{D}|} < \Psi_{Sup,\delta}(\gamma)$. Ceci contredit également notre hypothèse de départ. Donc la fréquence de la prémisse de r est supérieure à γ .

Ce théorème est la réciproque du théorème 1. La combinaison de ces deux théorèmes signifie que l'ensemble des règles de classe à prémisses fréquentes et ayant un nombre d'exceptions borné est équivalent à l'ensemble des règles de classe ayant pour chaque mesure de \mathcal{M} une valeur supérieure au minorant de la mesure. Extraire toutes les règles de classe dont la fréquence de la prémisse est supérieure à γ et avec au plus δ exceptions assure l'extraction complète de toutes les règles de classe optimisant simultanément toutes les mesures de \mathcal{M} . Nous nous appuyons sur ce résultat à la section 5 pour extraire toutes les règles de classe informatives optimisant les mesures de \mathcal{M} .

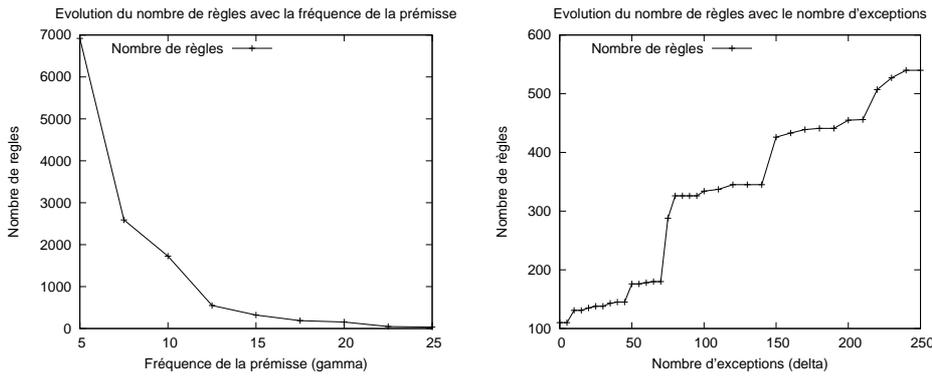
5 Implémentation et mise en œuvre

Nous présentons brièvement un prototype permettant l'extraction complète de toutes les règles de classe informatives optimisant les mesures \mathcal{M} ainsi qu'un exemple d'application sur le classique benchmark MUSHROOM de l'UCI².

5.1 Implémentation

Les règles informatives de classe présentent un double intérêt. D'une part, comme nous l'avons vu à la section 2.1, tout en étant moins nombreuses, elles ne font pas perdre d'information. D'autre part, leur obtention est efficace car elle s'appuie sur des motifs particuliers, appelés motifs "libres" ou "clés". Ces motifs sont caractérisés par une propriété donnant lieu à une contrainte anti-monotone par rapport à la spécialisation des attributs qu'il est particulièrement efficace de pousser avec un algorithme par niveaux (Boulicaut *et al.*, 2000; Bastide *et al.*, 2000). Une règle informative est construite

²<http://www.ics.uci.edu/~mlearn/MLSummary.html>

FIG. 3 – Évolution du nombre de règles en fonction de γ et δ .

à partir d'un motif libre X qui forme sa prémisse et de la fermeture de X (i.e., les attributs a tels que $\mathcal{F}(Xa) = \mathcal{F}(X)$). Par exemple, pour les données de la table 1, avec $\gamma = 4$ et $\delta = 1$, AC est un motif libre et $r_3 = AC \rightarrow c_1$ est une règle de classe informative. $ACE \rightarrow c_1$ n'est pas une règle de classe informative, elle est "couverte" par r_3 .

Étant donné γ et δ , le prototype CLARMINER (CLAR pour Class Rules) extrait la collection complète des règles de classe informatives dont la fréquence de la prémisse est supérieure ou égale à γ et admettant au plus δ exceptions. Nous avons vu (cf. section 4.3), qu'une telle extraction fournit toutes les règles de classe informatives optimisant toutes les mesures \mathcal{M} . CLARMINER repose sur le principe des algorithmes par niveaux. Il extrait tous les motifs libres et calcule pour chacun sa presque-fermeture (i.e., pour un motif libre X , les attributs a tels que $\mathcal{F}(X) - \mathcal{F}(Xa) \leq \delta$).

5.2 Extraction des règles de classe informatives

À partir du jeu de données MUSHROOM (8124 objets, 120 attributs et deux classes), nous quantifions le nombre de règles extraites ainsi que leur qualité. Les expérimentations sont menées en utilisant un processeur Pentium IV à 2.20 GHz avec le système d'exploitation Linux et 3Go de mémoire vive. Pour toutes les expériences, les temps d'extraction sont inférieurs à la seconde. La figure 3 (courbe de gauche) donne le nombre de règles de classe informatives lorsque δ est fixé à 50 et γ varie de 406 à 2031 (5% à 25% pour la fréquence relative); la courbe de droite lorsque γ est fixé à 1500 et δ varie entre 5 et 250. La décroissance en γ est assez rapide et fixer ce paramètre à une valeur "raisonnable" (10%) suffit à faire diminuer significativement le nombre de règles extraites. En faisant varier δ , on remarque un phénomène de paliers pour le nombre de règles obtenues. On dispose d'une certaine latitude dans le choix d'une borne pour le nombre d'exceptions puisque il existe des plages importantes sur lesquelles on obtient presque le même ensemble de règles.

Donnons quelques précisions sur la qualité des règles extraites sur l'expérience effectuée $\gamma = 812$ et $\delta = 40$ (cela signifie que les règles extraites ont une fréquence

minimale de leurs prémisses de l'ordre de 10% et admettent au plus 5% d'exceptions). On obtient 1598 règles. La longueur des prémisses des règles est inférieure ou égale à 7. Le tableau 4 donne pour chaque mesure la valeur du minorant, la valeur moyenne sur l'ensemble des règles extraites et l'écart minorant-moyenne exprimé en pourcentage du minorant. Pour cette expérience, on constate que les valeurs moyennes des mesures de \mathcal{M} pour ces règles sont nettement supérieures aux minorants théoriques donnés au tableau 3. Par exemple, on obtient une moyenne de 0.252 pour la Sensibilité (le minorant théorique est 0.184) et de 74, 65 pour la mesure de Sebag & Schoenauer (le minorant théorique est 19, 3). La moyenne de cette dernière mesure représente une augmentation de 286.81% par rapport au minorant annoncé. On remarque que pour d'autres mesures, la différence entre le minorant et la valeur moyenne est moins spectaculaire cependant, il faut noter que les minorants de ces mesures sont en général assez proches de la valeur optimale. Par exemple, en ce qui concerne la Confiance, le minorant vaut 0.951 pour une valeur optimale de 1 et l'écart à la valeur moyenne est de seulement 4.3%.

Mesure	Support	Confidence	Sensitivity	Specificity	Success Rate	Lift	PS	Laplace	Odds Ratio
Minorant	0.095	0.951	0.184	0.990	0.572	1.835	0.043	0.950	21.771
Moyenne	0.128	0.992	0.252	0.998	0.620	1.958	0.063	0.991	29.603
Ecart	34.74	4.3	37.33	0.81	8.39	6.70	45.63	4.3	35.97

Mesure	TC	S & S	Jaccard	Conviction	ϕ -coefficient	AV	Certainty Factor	GI
Minorant	17.961	19.30	0.185	9.785	0.289	0.433	0.898	0.607
Moyenne	71.178	74.654	0.252	36.901	0.72	0.485	0.984	0.671
Ecart	296.29	286.81	36.22	277.12	28.72	12.01	9.58	10.54

TAB. 4 – Minorant, valeur moyenne des mesures et écart (%) par rapport au minorant.

6 Conclusion

Dans cet article, nous avons défini un cadre formel (l'ensemble de mesures optimisables \mathcal{M}) qui permet de mieux comprendre le comportement de nombreuses mesures d'intérêt sur les règles de classe. Nous avons montré que ces mesures varient de façon similaire en fonction de deux paramètres (la fréquence de la prémisse d'une règle et son nombre d'exceptions). La question du choix d'une mesure de \mathcal{M} par l'utilisateur est ainsi relativisée. Nous avons montré que toute mesure de \mathcal{M} admet un minorant exprimé en fonction de la fréquence minimale et d'une borne du nombre d'exceptions d'une règle. Nous avons identifié la collection complète de toutes les règles de classe optimisant simultanément les mesures de \mathcal{M} et nous avons proposé un prototype qui extrait la collection complète des règles de classe informatives optimisant toutes les mesures \mathcal{M} .

Ce travail présente plusieurs perspectives. D'une part, il s'agit d'étendre ces résultats à d'autres types de règles. Cette généralisation n'est pas immédiate et nécessitera certainement des étapes intermédiaires car, pour une règle quelconque, le nombre d'objets supportant sa conclusion n'est pas connu alors que ce point est fortement utilisé dans le cas des règles de classe. Une autre voie est l'étude approfondie de la qualité réelle des règles obtenues par rapport aux valeurs des minorants et l'aide à l'utilisateur pour le choix des paramètres γ et δ (permettre à l'utilisateur, par exemple via une interface interactive, de visualiser l'effet du choix des paramètres sur les règles sélectionnées). Enfin, une voie particulièrement prometteuse est l'optimisation multi-critères. Il s'agit

de déterminer automatiquement les couples de paramètres (γ, δ) permettant de réaliser des extractions de règles vérifiant des qualités sur différentes mesures supérieures à des seuils fixés par l'utilisateur pour chaque mesure et ainsi combiner les différentes sémantiques apportées par ces mesures.

Remerciements. Ce travail a été partiellement financé par l'ACI "masse de données" (ministère de la recherche), projet BINGO (MD 46, 2004-2007).

Références

- AGRAWAL R., IMIELINSKI T. & SWAMI A. N. (1993). Mining association rules between sets of items in large databases. In P. BUNEMAN & S. JAJODIA, Eds., *SIGMOD'93 Conference*, p. 207–216 : ACM Press.
- AGRAWAL R. & SRIKANT R. (1994). Fast algorithms for mining association rules in large databases. In J. B. BOCCA, M. JARKE & C. ZANIOLO, Eds., *VLDB'94, Proceedings of 20th International Conference on Very Large Data Bases, September 12-15, 1994, Santiago de Chile, Chile*, p. 487–499 : Morgan Kaufmann.
- BARALIS E. & GARZA P. (2003). Majority classification by means of association rules. In *proceedings of the Seventh European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'03)*, volume 2838 of *Lecture notes in artificial intelligence*, p. 35–46, Dubvronik, Croatia : Springer.
- BASTIDE Y., PASQUIER N., TAOUIL R., STUMME G. & LAKHAL L. (2000). Mining minimal non-redundant association rules using frequent closed itemsets. In J. W. LLOYD, V. DAHL, U. FURBACH, M. KERBER, K.-K. LAU, C. PALAMIDESSI, L. M. PEREIRA, Y. SAGIV & P. J. STUCKEY, Eds., *Computational Logic*, volume 1861 of *Lecture Notes in Computer Science*, p. 972–986 : Springer.
- BASTIDE Y., TAOUIL R., PASQUIER N., STUMME G., & LAKHAL L. (2002). Pascal : un algorithme d'extraction des motifs fréquents. *Techniques et Sciences Informatiques*, **21**(1), 65–95.
- BAYARDO J. R. J. & AGRAWAL R. (1999). Mining the most interesting rules. In *KDD'99*, p. 145–154.
- BOULICAUT J.-F., BYKOWSKI A. & RIGOTTI C. (2000). Approximation of frequency queries by means of free-sets. In D. A. ZIGHED, H. J. KOMOROWSKI & J. M. ZYTKOW, Eds., *PKDD'00*, volume 1910 of *Lecture Notes in Computer Science*, p. 75–85 : Springer.
- BRIN S., MOTWANI R. & SILVERSTEIN C. (1997). Beyond market baskets : Generalizing association rules to correlations. In J. PECKHAM, Ed., *SIGMOD 1997, Proceedings ACM SIGMOD International Conference on Management of Data, May 13-15, 1997, Tucson, Arizona, USA*, p. 265–276 : ACM Press.
- CRÉMILLEUX B. & BOULICAUT J.-F. (2002). Simplest rules characterizing classes generated by delta-free sets. In *22nd Int. Conf. on Knowledge Based Systems and Applied Artificial Intelligence (ES'02)*, p. 33–46, Cambridge, UK.
- DONG G. & LI J. (1999). Efficient mining of emerging patterns : discovering trends and differences. In *proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining (ACM SIGKDD'99)*, p. 43–52, San Diego, CA : ACM Press.

- FRANCISCI D. & COLLARD M. (2003). Multi-criteria evaluation of interesting dependencies according to a data mining approach. In *Congress on Evolutionary Computation*, p. 1568–1574, Canberra, Australia : IEEE Press,.
- GUILLAUME S. (2000). *Traitement des données, volumineuses, mesures et algorithmes d'extraction de règles d'association et règles ordinales*. PhD thesis, Université de Nantes, France.
- HILDERMAN R. J. & HAMILTON H. J. (2003). Measuring the interestingness of discovered knowledge : A principled approach. *Intell. Data Anal.*, **7**(4), 347–382.
- KLEMETTINEN M., MANNILA H., RONKAINEN P., TOIVONEN H. & VERKAMO A. I. (1994). Finding interesting rules from large sets of discovered association rules. In *proceedings of the 3rd international Conference on Information and Knowledge Management (CIKM'94)*, p. 401–407 : ACM Press.
- LALLICH S. & TEYTAUD O. (2004). Evaluation et validation de l'intérêt des règles d'association. p. 193–218.
- LENT B., SWAMI A. & WIDOM J. (1997). Clustering association rules. In *Proceedings of Thirteenth International Conference on Data Engineering (ICDE'97)*, p. 220–231, Birmingham, UK.
- LIU B., HSU W. & MA Y. (1998). Integrating classification and association rules mining. In *proceedings of Fourth International Conference on Knowledge Discovery & Data Mining (KDD'98)*, p. 80–86, New York : AAAI Press.
- MCGARRY K. (2005). A survey of interestingness measures for knowledge discovery. *The Knowledge Engineering Review*, **20**(1), 39–61.
- NG R. T., LAKSHMANAN V. S., HAN J. & PANG A. (1998). Exploratory mining and pruning optimizations of constrained associations rules. In *proceedings of ACM SIGMOD'98*, p. 13–24 : ACM Press.
- PIATETSKY-SHAPIO G. (1991). Discovery, analysis, and presentation of strong rules. In *Knowledge Discovery in Databases*, p. 229–248. AAAI/MIT Press.
- PLASSE M., NIANG N., SAPORTA G. & LEBLOND L. (2006). Une comparaison de certains indices de pertinence des règles d'association. In G. RITSCHARD & C. DJERABA, Eds., *EGC*, volume RNTI-E-6 of *Revue des Nouvelles Technologies de l'Information*, p. 561–568 : Cépaduès-Éditions.
- SEBAG M. & SCHOENAUER M. (1988). Generation of rules with certainty and confidence factors from incomplete and incoherent learning bases. In M. L. J. BOOSE, B. GAINES, Ed., *European Knowledge Acquisition Workshop, EKAW'88*, p. 28–1–28–20.
- TAN P.-N., KUMAR V. & SRIVASTAVA J. (2002). Selecting the right interestingness measure for association patterns. In *KDD*, p. 32–41 : ACM.
- TOIVONEN H., KLEMETTINEN M., RONKAINEN P., HÄTÖNEN K. & MANNILA H. (1995). Pruning and grouping of discovered association rules. In *proceedings of the workshop on Statistics, Machine Learning and Knowledge Discovery in Databases co-located with ECML'95*, p. 47–52, Heraklion, Crete, Greece.
- VAILLANT B., LENCA P. & LALLICH S. (2004). Etude expérimentale de mesures de qualité de règles d'associations. In *4èmes journées Francophones sur l'Extraction et la Gestion de Connaissances (EGC 2004)*, p. 341–352.
- VAILLANT B., MEYER P., PRUDHOMME E., LALLICH S., LENCA P. & BIGARET S. (2005). Mesurer l'intérêt des règles d'association. In *EGC*, p. 421–426.
- ZAKI M. J. (2000). Generating non-redundant association rules. In *KDD'00*, p. 34–43.