

Critères d'évaluation des mesures de qualité des règles d'association

Philippe Lenca*, Patrick Meyer**, Philippe Picouet*
Benoît Vaillant*, Stéphane Lallich***

*Département IASC - GET ENST Bretagne
Technopôle de Brest Iroise - CS 83818, 29238 Brest Cedex, France
{*prenom.nom*}@enst-bretagne.fr,
<http://www-iasc.enst-bretagne.fr/~nom/>

**Département de Mathématique - Université de Liège
Bât. B37 Grande Traverse, 12, B-4000 Liège, Belgique
patrick.meyer@internet.lu,
<http://patrickmeyer.tripod.com>

***Laboratoire ERIC - Université Lumière - Lyon 2
5 avenue Pierre Mendès-France, 69676 Bron Cedex, France
lallich@univ-lyon2.fr,
<http://eric.univ-lyon2.fr/~lallich/>

Résumé. Les algorithmes de fouille de données, en particulier dans le cadre de l'apprentissage non supervisé, génèrent un grand nombre de règles. Il est donc concrètement impossible de procéder à une validation de ces règles en les présentant à un expert du domaine. Afin d'assister ce traitement, de nombreuses mesures de qualité de règles ont été proposées pour sélectionner et ordonner automatiquement les règles. Se pose alors le problème du choix de la mesure adaptée aux données et aux besoins de l'expert, celui-ci n'étant pas - *a priori* - expert en fouille de données. Dans le contexte des règles d'association, nous proposons une caractérisation des mesures en fonction de propriétés ayant une sémantique intuitive nous permettant de mettre en œuvre un processus d'aide à la décision afin d'assister l'expert dans son choix d'une mesure adaptée à ses besoins et ainsi l'aider à sélectionner les *meilleures* règles.

1 Introduction

L'un des objectifs essentiels de l'extraction de connaissances à partir de données (ECD) est de produire des règles intéressantes du point de vue de l'utilisateur (expert métier). Il est aujourd'hui admis que l'intérêt d'une règle est difficile à évaluer de façon objective, tant son appréciation dépend des experts capables de l'analyser. Idéalement, une règle doit être valide, nouvelle, compréhensible [Fayyad et al., 1996] mais ces termes, très génériques, couvrent des réalités très variées lorsqu'une règle est examinée dans un contexte précis : maximiser le nombre de données concernées, minimiser le nombre de contre-exemples, etc.

Les mesures d'évaluation de la qualité des règles jouent alors un rôle essentiel dans les processus d'ECD afin de préfiltrer les règles qui seront effectivement présentées à

l'expert. L'exemple le plus connu, dans le domaine de la recherche de règles d'association, est celui du support (ou taux de couverture). Cette mesure, grâce à sa propriété d'antimonotonie, est utilisée de manière très efficace comme indice de filtrage dans de nombreux algorithmes [Agrawal et al., 1993]. En revanche, elle possède à peu près tous les défauts qu'un expert peut craindre d'une mesure de qualité des règles [Picouet and Lenca, 2001]. Certains travaux ([Hilderman and Hamilton, 2001], [Lallich and Teytaud, 2002]) ont mis en évidence des caractéristiques très diverses des mesures. D'autres encore ([Tan et al., 2002], [Vaillant et al., 2003b]) montrent que l'ensemble des n meilleures règles peut varier grandement selon la mesure utilisée. Le tableau 1 contient des résultats expérimentaux obtenus avec le simulateur HERBS ([Vaillant et al., 2003a]) utilisant les résultats de [Giakoumakis and Monjardet, 1987] pour comparer les préordres (à l'aide d'une grandeur entre -1 et 1 ; plus la valeur est grande, plus les préordres concordent) obtenus à partir de huit mesures de qualité¹. Les lignes et les colonnes ont été réorganisées dans l'ordre qui fait le mieux apparaître une éventuelle structure diagonale ou par blocs du tableau grâce au logiciel AMADO (Analyse graphique d'une MAtrice de DONnées) [Chauchat and Risson, 1998]. On observe des similitudes mais aussi d'importantes différences de classement.

	CONF	SEB	TEC	LOE	ZHANG	LAP	PS	SUP
CONF	1	1	1	0,992	0,992	0,136	-0,079	-0,136
SEB	1	1	1	0,992	0,992	0,136	-0,079	-0,136
TEC	1	1	1	0,992	0,992	0,136	-0,079	-0,136
LOE	0,992	0,992	0,992	1	0,999	0,125	-0,072	-0,133
ZHANG	0,992	0,992	0,992	0,999	1	0,125	-0,072	-0,133
LAP	0,136	0,136	0,136	0,125	0,125	1	0,774	0,733
PS	-0,079	-0,079	-0,079	-0,072	-0,072	0,774	1	0,882
SUP	-0,136	-0,136	-0,136	-0,133	-0,133	0,733	0,882	1

TAB. 1 – Comparaison des préordres de 8 mesures

Face à ce constat, notre objectif est double : dans le contexte particulier de la recherche de règles d'association, il consiste dans un premier temps à dresser un tableau comparatif de plusieurs mesures de qualité en regard d'un ensemble de propriétés faisant sens du point de vue d'un expert ; dans un deuxième temps, nous appliquons un processus d'aide multicritère à la décision (AMD) pour faciliter le choix d'une ou de quelques mesures. Nous avons montré l'intérêt d'une telle approche sur un ensemble restreint de mesures et de critères dans [Lenca et al., 2003].

La nécessité de disposer de critères objectifs d'évaluation pour les mesures a déjà été soulignée dans [Piatetsky-Shapiro, 1991] et dans [Freitas, 1999]. Les propriétés qui y sont proposées ne sont néanmoins pas toutes reprises dans notre travail, soit par manque de pertinence dans le cadre des règles d'association, soit parce qu'elles ne permettraient pas de différencier les mesures. Nous introduisons en revanche d'autres propriétés, telles que la facilité à fixer le seuil, qui nous semblent essentielles pour le travail d'un expert. Celui-ci est en effet le plus concerné par le choix de la mesure. Dans [Silberschatz and Tuzhilin, 1995], la notion d'intérêt de la mesure est rapportée à sa concordance avec les croyances de l'expert et sa capacité à être transformée en action. Dans [Hilderman and Hamilton, 2001], les auteurs dressent une matrice d'évaluation de mesures qui doit permettre à l'expert de faire le bon choix. Bien que se situant dans

1. {Base de cas = car} [Blake and Merz, 1998], {Base de règles = car_apriori}, contenant 386 règles calculées avec l'algorithme APRIORI de [Borgelt and Kruse, 2002] avec un support de 2% et une confiance de 80%. Les mesures sont définies dans les tableaux 2 et 3.

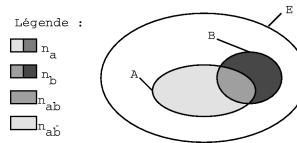
un contexte différent (résumés codés vs. règles d’association), notre travail suit une approche comparable mais nous prolongeons néanmoins cette démarche en justifiant une méthode d’AMD plus à même d’aiguiller l’expert.

L’article est organisé de la façon suivante. Dans la section 2, nous dressons une liste de vingt mesures éligibles (contexte des règles d’association). En section 3, nous décrivons une liste de huit critères pertinents pour évaluer les mesures du point de vue d’objectifs utilisateur. Nous évaluons alors chaque mesure sur chacun des critères. La matrice de décision ainsi bâtie nous permet de mettre en œuvre la méthode PROMETHEE afin de ranger les mesures étudiées (section 4). Nous concluons en section 5.

2 Mesures éligibles

Nous nous limitons volontairement aux mesures relatives à l’intérêt des règles d’association telles que celles-ci ont été définies par [Agrawal et al., 1993] : dans une base de données transactionnelle, $A \rightarrow B$ signifie que si les articles qui constituent A sont “dans le panier d’une ménagère”, alors le plus souvent les articles qui constituent B le sont aussi. On doit donc distinguer les règles d’association $A \rightarrow B$, qui se focalisent sur les coprésences en traitant les variables de façon non symétrique, de l’implication $A \Rightarrow B$ et de l’équivalence $A \Leftrightarrow B$. Une bonne façon de préciser le critère d’éligibilité des mesures de l’intérêt des règles d’association consiste à ne retenir que les mesures qui sont décroissantes avec $n_{a\bar{b}}$ (resp. croissantes avec n_{ab}), les effectifs marginaux étant fixés. C’est ainsi que sont d’emblée exclues certaines mesures comme le χ^2 , le r^2 de Pearson, la J-mesure [Goodman and Smyth, 1988] ou encore la mesure de Pearl [Pearl, 1988].

Soit $n = |E|$, le nombre total d’enregistrements
 Pour $A \rightarrow B$, on note :
 $n_a = |A|$, le nombre d’enregistrements vérifiant A.
 $n_b = |B|$, le nombre d’enregistrements vérifiant B.
 $n_{ab} = |A \cap B|$, le nombre d’exemples de la règle.
 $n_{a\bar{b}} = |A \cap \bar{B}|$, le nombre de contre-exemples à la règle.
 $A \rightarrow B$ est évaluée à l’aide de mesures généralement monotones décroissantes en fonction de $n_{a\bar{b}} = n_a - n_{ab}$. $A \rightarrow B$ est jugée intéressante selon la mesure μ lorsque $\mu(A \rightarrow B) \geq \alpha$, α devant être fixé par l’utilisateur.



$A \setminus B$	0	1	total
0	$p_{a\bar{b}}$	p_{ab}	p_a
1	$p_{\bar{a}\bar{b}}$	$p_{\bar{a}b}$	$p_{\bar{a}}$
total	$p_{\bar{b}}$	p_b	1

Pour $X \subset E$, on remplace n_X/n par p_X lorsque l’on considère les fréquences relatives plutôt que les fréquences absolues.

Si l’on fixe les caractéristiques marginales du tableau (n , n_a et n_b ou p_a et p_b), il suffit de connaître une cellule du tableau pour reconstruire les autres.

FIG. 1 – Notations.

Nous avons retenu les principales mesures d’intérêt des règles d’association rencontrées dans la littérature (tableaux 2 et 3). On citera d’abord le support p_{ab} et la confiance $p_{b|a}$ qui sont à la base des algorithmes de recherche des règles fondés sur l’extraction des itemsets plus fréquents qu’un certain seuil [Agrawal et al., 1993].

Un grand nombre de mesures sont des transformées affines de la confiance à marges fixées ayant pour objet de comparer celle-ci à p_b . Cette comparaison se fait le plus souvent en centrant la confiance sur p_b avec différents coefficients d’échelle (confiance

centrée, mesure de Piatetsky-Shapiro, mesure de Loevinger, mesure de Zhang, coefficient de corrélation, indice d'implication, surprise). Elle peut aussi se faire en divisant la confiance par p_b (lift).

D'autres mesures sont des transformées monotones croissantes de la confiance, ainsi la mesure de Sebag-Schoenauer, le taux d'exemples et de contre-exemples, alors que le gain informationnel est une transformation monotone croissante du lift. Certaines mesures privilégient les contre-exemples, par exemple la conviction ou l'indice d'implication déjà cité. Ce dernier est à la base de différents indices probabilistes comme l'intensité d'implication et l'indice probabiliste discriminant. L'intensité d'implication entropique module l'intensité d'implication par une entropie qui remédie au faible pouvoir discriminant de l'intensité d'implication. Enfin, le multiplicateur de cote est une sorte d'odd-ratio calculé non pas en comparant les cotes de B suivant A et \bar{A} , mais en comparant la cote de B suivant A et celle de B, alors que la mesure de Laplace est une variante de la confiance qui prend en compte le nombre de cas n .

On peut considérer que deux mesures liées par une transformation monotone croissante sont équivalentes dans la mesure où elles classent les règles de la même façon. Par exemple : la confiance, le taux d'exemples et de contre-exemples, et la mesure de Sebag (ce qui est illustré par les valeurs égales à 1 dans le tableau 1) ; le lift et le gain informationnel ; l'indice de Loevinger et la conviction.

	Mesures	Références
SUP	support	[Agrawal et al., 1993]
CONF	confiance	[Agrawal et al., 1993]
R	coefficient de corrélation de Pearson	[Pearson, 1896]
CONF CEN	confiance centrée	
PS	Piatetsky-Shapiro	[Piatetsky-Shapiro, 1991]
LOE	Loevinger	[Loevinger, 1947]
ZHANG	Zhang	[Terano et al., 2000]
- INDIMP	indice d'implication	[Lerman et al., 1981]
LIFT	Lift	[IBM, 1996]
SURP	surprise	[Azé and Kodratoff, 2001]
SEB	Sebag et Schoenauer	[Sebag and Schoenauer, 1988]
MC	multiplicateur des cotes	[Lallich and Teytaud, 2002]
CONV	conviction	[Brin et al., 1997]
TEC	taux d'exemples et de contre-exemples	
IQC	indice de qualité de Cohen	[Cohen, 1960]
GI	gain informationnel	[Church and Hanks, 1990]
INTIMP	intensité d'implication	[Gras et al., 1996]
IIE	intensité d'implication entropique	[Gras et al., 2001]
IPD	indice probabiliste discriminant	[Lerman and Azé, 2003]
LAP	Laplace	[Good, 1965]

TAB. 2 – Liste des mesures éligibles

3 Critères d'évaluation

Nous proposons ci-dessous huit critères possibles pour évaluer, sur des échelles ordinales, les mesures éligibles. Les résultats d'une évaluation des mesures sur ces critères sont rassemblés dans le tableau 4. Il est important de noter que nous ne proposons pas une norme. L'évaluation présentée est de *bon sens*, mais elle reste arbitraire et modifiable en fonction des préférences de l'utilisateur. Les auteurs tiennent à remercier les

	Déf. en fréquences absolues	Déf. en fréquences relatives
SUP	$\frac{n_a - n_{a\bar{b}}}{n_a n_{a\bar{b}}}$	p_{ab}
CONF	$1 - \frac{n_{a\bar{b}}}{n_a}$	$p_{b/a}$
R	$\frac{n n_a b - n_a n_b}{\sqrt{n n_a n_b n_{a\bar{b}}}}$	$\frac{p_{ab} - p_a p_b}{\sqrt{p_a p_{a\bar{b}} p_b p_{\bar{b}}}}$
CONF CEN	$\frac{n n_a b - n_a n_b}{n_a}$	$p_{b/a} - p_b$
PS	$\frac{1}{n} \left(\frac{n_a n_{\bar{b}}}{n} - n_{a\bar{b}} \right)$	$n p_a (p_{b/a} - p_b) = n p_a p_b (\text{Lift} - 1)$
LOE	$1 - \frac{n n_a \bar{b}}{n_a n_{\bar{b}}}$	$\frac{p_{b/a} - p_b}{p_{\bar{b}}} = \frac{1}{p_{\bar{b}}} \text{ConfCen} = 1 - \frac{1}{\text{Conv}}$
ZHANG	$\frac{n n_a b - n_a n_b}{\max\{n_a b n_{\bar{b}}, n_b n_{a\bar{b}}\}}$	$\frac{p_{ab} - p_a p_b}{\max\{p_{ab} p_{\bar{b}}, p_b (p_a - p_{ab})\}}$
INDIMP	$\frac{n n_a \bar{b} - n_a n_{\bar{b}}}{\sqrt{n n_a n_{\bar{b}}}}$	$\frac{p_{a\bar{b}} - p_a p_{\bar{b}}}{\sqrt{n p_a p_{\bar{b}}}}$
LIFT	$\frac{n n_a b}{n_a n_b}$	$\frac{p_{ab}}{p_a p_b}$
SURP	$\frac{n a b - n_a \bar{b}}{n n_b}$	$\frac{p_{ab} - p_{a\bar{b}}}{p_b} = 2 \frac{p_a}{p_b} (\text{Conf} - 0.5)$
SEB	$\frac{n_a - n_{a\bar{b}}}{n_{a\bar{b}}}$	$\frac{p_{ab}}{p_{a\bar{b}}} = \frac{\text{Conf}}{1 - \text{Conf}}$
MC	$\frac{(n_a - n_{a\bar{b}}) n_{\bar{b}}}{n_b n_{a\bar{b}}}$	$\frac{p_{\bar{b}/a}}{p_{\bar{b}}} = \frac{p_{ab}}{p_b} \frac{p_{\bar{b}}}{p_{a\bar{b}}} = \text{Lift} \cdot \text{Conv}$
CONV	$\frac{n_a n_{\bar{b}}}{n n_{a\bar{b}}}$	$\frac{p_a p_{\bar{b}}}{p_{a\bar{b}}}$
TEC	$\frac{n_a - 2 n_{a\bar{b}}}{n_a - n_{a\bar{b}}} = 1 - \frac{1}{\frac{n_a}{n_{a\bar{b}}} - 1}$	$1 - \frac{p_{a\bar{b}}}{p_{ab}} = 1 - \frac{1}{\text{SEB}}$
IQC	$\frac{2 n n_a - n n_{a\bar{b}} - n_a n_b}{2 n n_a + n n_{a\bar{b}} - 2 n_a n_b}$	$\frac{2 p_{ab} - p_a p_b}{p_a + p_b - 2 p_a p_b}$
GI	$\frac{n n_a b}{\log(\frac{n_a n_b}{n})}$	$\log \frac{p_{ab}}{p_a p_b} = \log(\text{Lift})$
INTIMP	$P \left[\text{Poisson} \left(\frac{n_a n_{\bar{b}}}{n} \right) \geq n_{a\bar{b}} \right]$	$P \left[\text{Poisson} (n p_a p_{\bar{b}}) \geq n p_{a\bar{b}} \right]$
IIE	$\left\{ \left[(1 - h_1 \left(\frac{n_{a\bar{b}}}{n} \right)^2) (1 - h_2 \left(\frac{n_a n_{\bar{b}}}{n} \right)^2) \right]^{1/4} \text{IntImp} \right\}^{1/2}$	$\left\{ \left[(1 - h_1 (p_{ab})^2) (1 - h_2 (p_{a\bar{b}})^2) \right]^{1/4} \text{IntImp} \right\}^{1/2}$
IPD ^a	$P \left[\mathcal{N}(0, 1) > \text{IndImp}^{CR/B} \right]$	
LAP	$\frac{n_{a\bar{b}} + 1}{n_a + 2}$	$\frac{p_{b/a} + \frac{n}{p_a}}{1 + \frac{2n}{p_a}}$

TAB. 3 – Mesures éligibles

^a $\text{IndImp}^{CR/B}$ correspond à INDIMP centré réduit (CR) pour une base de règles admissibles B .

membres du groupe GAFOQUALITÉ (AS STIC Fouille de Données) pour les discussions fructueuses à propos des propriétés d'évaluation des mesures.

g_1 : traitement non symétrique de A et B. L'antécédent et le conséquent d'une règle n'ayant pas le même rôle, il est souhaitable qu'une mesure évalue de façon différente les règles $A \rightarrow B$ et $B \rightarrow A$. On note 0 si la mesure est symétrique, 1 sinon.

g_2 : décroissance avec n_b . Pour n_{ab} , $n_{a\bar{b}}$ et n fixés, une règle est d'autant plus intéressante que n_b est faible et on attend qu'une mesure soit une fonction décroissante de n_b . Le centrage ou/et la division de la confiance par p_b permet(tent) que ce critère soit réalisé. On note 0 si la mesure n'est pas une fonction décroissante de n_b , et 1 sinon.

g_3 : évaluation des situations de références : indépendance. Une bonne mesure doit donner des valeurs fixes aux situations de référence. La première situation de référence est l'indépendance. On dit que A et B sont indépendants lorsque la fréquence de réalisation de B est la même, que A soit vrai ou non. Dans ce cas, la règle $A \rightarrow B$ n'apporte aucune information, eut-elle une très forte confiance. Il est souhaité que la

mesure lui attribue une valeur fixe, 0 le plus fréquemment. On note 0 si la mesure n'a pas une valeur fixe en cas d'indépendance et 1 sinon.

g_4 : évaluation des situations de références : règle logique. A l'opposé, la seconde situation de référence est la règle logique lorsqu'il n'y a pas de contre-exemples. Une valeur constante ou à la rigueur $+\infty$ est alors attendue. On note 0 si la mesure n'a pas une valeur fixe en cas de règle logique et 1 sinon.

Nous écartons le critère constitué par l'évaluation de l'incompatibilité, une 3^e situation de référence définie par $A \cap B = \emptyset$, qui exprime que B ne peut pas se réaliser si A est réalisé. En effet, cette situation intéresse la règle $A \rightarrow \bar{B}$ et non $A \rightarrow B$.

g_5 : non linéarité en fonction de $p_{a\bar{b}}$ autour de 0^+ . Différents auteurs souhaitent une décroissance lente au début, plus rapide ensuite, plutôt qu'une décroissance linéaire comme dans le cas de la confiance et de ses transformées affines, ou qu'une décroissance convexe ([Gras et al., 2002]). Il faut remarquer que dans la mesure où la confiance est une transformée affine du nombre d'exemples ou de contre-exemples à marges fixées, toutes les mesures qui sont des transformées affines de la confiance ont cette même caractéristique. On note 0 si la mesure est convexe pour $n_{a\bar{b}}$ voisin de 0, 1 si elle est linéaire et 2 si elle est concave. Une décroissance convexe qui accroît la sensibilité aux faux positifs peut être préférée dans certains cas.

g_6 : prise en compte du nombre de cas. Intuitivement, pour des proportions identiques dans le tableau $A \times B$ (cf. tableau de la fig. 1), la règle est d'autant plus fiable que le nombre de transactions n est élevé. On peut donc légitimement souhaiter que la mesure soit croissante lorsque l'on dilate les données en augmentant n , toutes proportions fixées. L'inconvénient de telles mesures est qu'elles fournissent une grande masse de règles dont l'évaluation est proche de 1 et pour certaines perdent leur pouvoir discriminant, alors même que les règles concurrentes présentent la même valeur de n ! C'est ainsi qu'une mesure purement descriptive pour laquelle n intervient dans le calcul du seuil à partir duquel on retient la règle peut être préférée. On choisit ici de noter 0 si la mesure est invariante en cas de dilatation des données et 1 si elle est croissante. Le choix inverse peut tout aussi bien être retenu par l'utilisateur.

g_7 : facilité à fixer un seuil. Face à la multitude de règles évaluées, il est important de pouvoir facilement fixer le seuil à partir duquel on considère que les règles ont un réel intérêt, sans avoir à les classer. Une façon classique de procéder consiste à se référer à la probabilité critique de la valeur observée de la mesure sous l'hypothèse d'indépendance (ou p-value). Celle-ci ne doit pas être interprétée comme un risque statistique compte tenu de la multitude de tests effectués ([Lallich and Teytaud, 2002]), mais comme un paramètre de contrôle. La fixation d'un seuil est immédiate lorsque la mesure est elle-même définie comme une telle probabilité critique, ainsi INTIMP ([Gras et al., 1996]) ou IPD ([Lerman and Azé, 2003]). Sinon, on peut le plus souvent définir un tel seuil en partant de l'un ou l'autre des trois types de modélisation proposés par [Lerman et al., 1981] pour établir la loi de n_{ab} sous l'hypothèse d'absence de lien (H_0). Les seules mesures qui se prêtent mal à un tel calcul sont la mesure de Zhang en raison de son dénominateur et l'intensité d'implication entropique en raison de la pondération par l'entropie de l'expérience. On note 0 si la mesure se prête mal au calcul d'un seuil et 1 si elle se prête bien à un tel calcul.

g_8 : intelligibilité. Dans certaines situations, notamment pour communiquer et

expliquer les résultats, l'intelligibilité de la mesure a de l'importance. On considérera que la mesure est intelligible si son sens concret peut s'écrire par une phrase simple. On note 0 si il y a impossibilité de donner le sens concret de la mesure, 1 si la mesure se ramène à des quantités usuelles et 2 sinon.

Nous envisageons d'intégrer d'autres critères, plus particulièrement le caractère plus ou moins discriminant de la mesure et l'existence d'une propriété d'antimonotonie. Le caractère plus ou moins discriminant de la mesure est important car souvent le fait de satisfaire au critère 6 (prise en compte du nombre de cas) s'accompagne d'une perte de capacité discriminante. L'existence d'une propriété d'antimonotonie, comme pour le support, facilite l'extraction des règles lorsque l'on explore le treillis des itemsets.

	g_1	g_2	g_3	g_4	g_5	g_6	g_7	g_8
SUP	0	0	0	0	1	0	1	2
CONF	1	0	0	1	1	0	1	2
R	0	1	1	0	1	0	1	1
CONF CEN	1	1	1	0	1	0	1	2
PS	0	1	1	0	1	1	1	1
LOE	1	1	1	1	1	0	1	1
ZHANG	1	1	1	1	2	0	0	0
-INDIMP	1	1	1	0	1	1	1	0
LIFT	0	1	1	0	1	0	1	1
SURP	1	1	0	0	1	0	1	1
SEB	1	0	0	1	0	0	1	1
MC	1	1	1	1	0	0	1	2
CONV	1	1	1	1	0	0	1	1
TEC	1	0	0	1	2	0	1	1
IQC	0	1	1	0	1	0	1	0
GI	0	1	1	0	2	0	1	0
INTIMP	1	1	1	1	2	1	1	0
HE	1	1	1	1	2	1	0	0
IPD	1	1	1	0	1	1	1	0
LAP	1	0	0	0	1	0	1	0

TAB. 4 – Matrice de décision

4 Evaluation des mesures

Nous évaluons à l'aide de la méthode PROMETHEE-Gaia [Brans and Mareschal, 1994, Brans and Mareschal, 2002] les mesures présentées en section 2 à partir des critères de la section 3. L'information nécessaire à l'analyse est résumée dans la matrice du tableau 4. PROMETHEE vise à construire un préordre total sur les mesures et permet de visualiser dans un plan commun (Gaia) les mesures et les critères de façon similaire à l'analyse en composantes principales : le plan renseigne sur le pouvoir de discrimination des critères, leur interdépendance et leur influence sur le rangement. Il permet d'avoir une vue synthétique et claire du caractère conflictuel de certains critères et de l'impact des poids des critères sur la décision finale. De plus, il permet de repérer les mesures similaires par simple proximité visuelle.

4.1 Examen de deux scénarios-expert

Nous avons retenus deux scénarios (on rappelle que nous ne proposons pas une norme et que les scénarios dépendent de l'utilisateur et de ses objectifs). Ceux-ci s'intéressent au critère g_5 (non-linéarité par rapport à $n_{a\bar{b}}$ au voisinage de 0^+) :

SC1 : l'expert tolère l'apparition de certains contre-exemples à une règle. Dans ce cas, la mise à jour de la règle est retardée. L'allure de la courbe représentant la mesure

est donc idéalement concave; l'ordre sur g_5 est alors (concave \succ linéaire \succ convexe).

SC2: l'expert refuse l'apparition de trop de contre-exemples à une règle. Dans ce cas, la mise à jour de la règle doit se faire rapidement. L'allure de la courbe représentant la mesure est donc idéalement convexe; l'ordre sur les valeurs de g_5 est alors (convexe \succ linéaire \succ concave). La colonne g_5 du tableau 4 est modifiée en conséquence.

Nous analysons le problème dans l'hypothèse où les poids des critères sont égaux. Le tableau 5 résume les rangements totaux pour les deux scénarios.

Rang	SC1	SC2
1	IIE	MC
2	INTIMP	CONV
3	LOE	LOE
4	MC	CONF CEN
5	CONF CEN	IIE
6	CONV	-INDIMP
7	-INDIMP	IPD
8	IPD	INTIMP
9	ZHANG	ZHANG
10	PS	PS
11	TEC	R
12	CONF	LIFT
13	GI	SEB
14	R	CONF
15	LIFT	IQC
16	SURP	SURP
17	SEB	GI
18	IQC	TEC
19	SUP	LAP
20	LAP	SUP

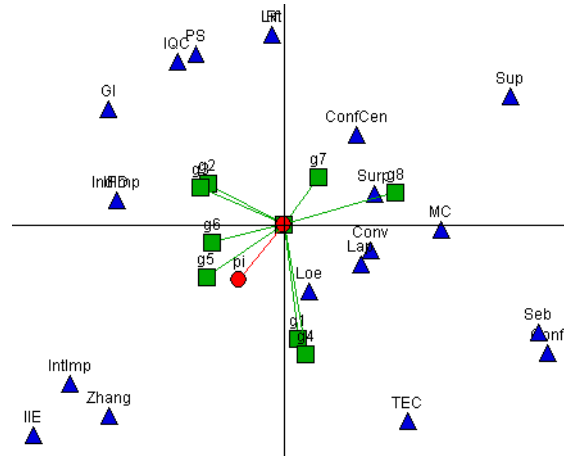
TAB. 5 – Rangements obtenus avec PROMETHEE

4.2 Analyse du scénario SC1

L'analyse des intervalles de stabilité des poids indique que l'on peut légèrement faire varier les poids des critères sans modifier le rangement total (*i.e.* a du sens). De plus, il est possible de modifier le poids des critères 1, 2, 4, 5, 6, 7 de manière significative sans trop perturber le début du rangement total.

Le plan Gaia (figure 2) nous permet de faire les observations suivantes :

- les critères 3 et 4 (décroissance avec n_b et hypothèse d'indépendance) sont représentés par des axes ayant approximativement la même direction : ces deux critères expriment des préférences similaires sur l'ensemble des mesures. La même observation peut être faite pour les critères 1 et 5 (asymétrie et règle logique).
- les axes représentant les critères sont répartis de manière homogène dans le plan : toute mesure peut se retrouver dans les premières positions du rangement, à condition de choisir un système de poids adéquat (par exemple, SUP est susceptible de rejoindre les premières positions en augmentant de manière significative le poids des critères 8 et 9 – facilité à placer le seuil et intelligibilité).
- le plan permet aussi de visualiser les mesures au comportement similaire. Elles sont représentées par des points proches dans le plan. Pour notre système de poids, on constate par exemple une forte similarité entre IIE, ZHANG et INTIMP.

FIG. 2 – *Plan Gaia pour SC1*

5 Conclusion

Nous avons examiné l'intérêt d'une approche multicritère d'aide à la décision pour aider l'expert dans le choix d'une mesure de qualité pour les règles d'association. Nous avons étudié vingt mesures sur huit critères pertinents du point de vue de l'utilisateur.

En dehors de l'intérêt de disposer d'une grille de lecture identique pour un nombre important de mesures, l'application de la méthode PROMETHEE a permis de confirmer l'influence des préférences de l'expert sur le classement des mesures étudiées, mais aussi de constater des comportements similaires. Enfin, la méthode permet de mieux analyser les préférences du décideur (on peut alors aisément déterminer quels critères sont importants).

L'ensemble des critères reste bien évidemment à élargir (ceci permettra peut-être de différencier certaines mesures actuellement confondues dans nos espaces de représentation). En particulier, des critères importants mais difficiles à évaluer formellement doivent être étudiés concrètement en situation sur des données. C'est le cas, par exemple du pouvoir discriminant d'une mesure.

Summary

Datamining algorithms, especially those used for unsupervised learning, generate a large quantity of rules. It is hence impossible for an expert of the field being mined to sustain these rules. To help in carrying out the task, many rule quality measures have been developed in order to filter and sort automatically a set of rules with respect to given goals. Then arises the problem of selecting the appropriate quality measure. As the expert is not supposed to be a datamining expert, we propose a characterisation of quality measures, based on meaningful classical properties. This set of properties

can be used in a multicriteria decision assistance process and hence help the expert selecting a measure fitting his needs.

Références

- [Agrawal et al., 1993] Agrawal, R., Imielinski, T., and Swami, A. N. (1993). Mining association rules between sets of items in large databases. In Buneman, P. and Jajodia, S., editors, *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pages 207–216, Washington, D.C.
- [Azé and Kodratoff, 2001] Azé, J. and Kodratoff, Y. (2001). Evaluation de la résistance au bruit de quelques mesures d'extraction de règles d'association. *Extraction des connaissances et apprentissage (EGC02)*, 1(4):143–154.
- [Blake and Merz, 1998] Blake, C. and Merz, C. (1998). Uci repository of machine learning databases. <http://www.ics.uci.edu/~mlearn/MLRepository.html>.
- [Borgelt and Kruse, 2002] Borgelt, C. and Kruse, R. (2002). Induction of association rules: Apriori implementation. In *Proceedings of the 15th Conference on Computational Statistics (Compstat 2002, Berlin, Germany)*, Heidelberg, Germany. Physika Verlag.
- [Brans and Mareschal, 1994] Brans, J. and Mareschal, B. (1994). The promethee-gaia decision support system for multicriteria investigations. *Investigation Operativa*, 4(2):102–117.
- [Brans and Mareschal, 2002] Brans, J. and Mareschal, B. (2002). *PROMETHEE-GAIA - Une méthode d'aide à la décision en présence de critères multiples*. Ellipses.
- [Brin et al., 1997] Brin, S., Motwani, R., Ullman, J. D., and Tsur, S. (1997). Dynamic itemset counting and implication rules for market basket data. In Peckham, J., editor, *SIGMOD 1997, Proceedings ACM SIGMOD International Conference on Management of Data, Tucson, Arizona, USA*, pages 255–264. ACM Press.
- [Chauchat and Risson, 1998] Chauchat, J.-H. and Risson, A. (1998). *Visualization of Categorical Data*, chapter 3, pages 37–45. Blasius J. & Greenacre M. ed. New York : Academic Press.
- [Church and Hanks, 1990] Church, K. W. and Hanks, P. (1990). Word association norms, mutual information an lexicography. *Computational Linguistics*, 16(1):22–29.
- [Cohen, 1960] Cohen, J. (1960). A coefficient of agreement for nominal scale. *Educational and Psychological Measurement*, 20:37–46.
- [Fayyad et al., 1996] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R., editors (1996). *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press.

- [Freitas, 1999] Freitas, A. (1999). On rule interestingness measures. *Knowledge-Based Systems journal*, pages 309–315.
- [Giakoumakis and Monjardet, 1987] Giakoumakis, V. and Monjardet, B. (1987). Coefficients d'accord entre deux préordres totaux. *Statistique et Analyse des Données*, 12(1 et 2):46–99.
- [Good, 1965] Good, I. J. (1965). The estimation of probabilities: An essay on modern bayesian methods. The MIT Press, Cambridge, MA.
- [Goodman and Smyth, 1988] Goodman, R. M. and Smyth, P. (1988). Information-theoretic rule induction. In *Proceedings of the 1988 European Conference on Artificial Intelligence*. Pitman.
- [Gras et al., 1996] Gras, R., Ag. Almouloud, S., Bailleuil, M., Larher, A., Polo, M., Ratsimba-Rajohn, H., and Totohasina, A. (1996). *L'implication Statistique, Nouvelle Méthode Exploratoire de Données. Application à la Didactique, Travaux et Thèses*. La Pensée Sauvage.
- [Gras et al., 2002] Gras, R., Couturier, R., Bernadet, M., Blanchard, J., Briand, H., Guillet, F., Kuntz, P., Lehn, R., and Peter, P. (2002). Quelques critères pour une mesure de qualités de règles d'association - un exemple: l'intensité d'implication. Rapport de recherche pour le groupe de travail GAFOQUALITÉ de l'as STIC fouille de bases de données, Ecole Polytechnique de l'Université de Nantes.
- [Gras et al., 2001] Gras, R., Kuntz, P., Couturier, R., and Guillet, F. (2001). Une version entropique de l'intensité d'implication pour les corpus volumineux. *Extraction des connaissances et apprentissage (EGC01)*, 1(1-2):69–80.
- [Hilderman and Hamilton, 2001] Hilderman, R. J. and Hamilton, H. J. (2001). Evaluation of interestingness measures for ranking discovered knowledge. *Lecture Notes in Computer Science*, 2035:247–259.
- [IBM, 1996] IBM (1996). *IBM Intelligent Miner User's Guide, Version 1 Release 1, SH12-6213-00*. International Business Machines.
- [Lallich and Teytaud, 2002] Lallich, S. and Teytaud, O. (2002). Evaluation et validation de l'intérêt des règles d'association. Rapport de recherche pour le groupe de travail GAFOQUALITÉ de l'action spécifique STIC fouille de bases de données, E.R.I.C., Université Lyon 2.
- [Lenca et al., 2003] Lenca, P., Meyer, P., Vaillant, B., and Picouet, P. (2003). Aide multicritère à la décision pour évaluer les indices de qualité des connaissances - modélisation des préférences de l'utilisateur. *RSTI-RIA (EGC03)*, 1(17):271–282.
- [Lerman and Azé, 2003] Lerman, I. and Azé, J. (2003). Une mesure probabiliste contextuelle discriminante de qualité des règles d'association. *RSTI-RIA (EGC03)*, 1(17):247–262.

- [Lerman et al., 1981] Lerman, I., Gras, R., and Rostam, H. (1981). Elaboration d'un indice d'implication pour les données binaires, i et ii. *Mathématiques et Sciences Humaines*, (74, 75):5–35, 5–47.
- [Loevinger, 1947] Loevinger, J. (1947). A systemic approach to the construction and evaluation of tests of ability. *Psychological monographs*, 61(4).
- [Pearl, 1988] Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. Morgan Kaufmann.
- [Pearson, 1896] Pearson, K. (1896). Mathematical contributions to the theory of evolution. iii. regression, heredity and panmixia. *Philosophical Transactions of the Royal Society*, A.
- [Piatetsky-Shapiro, 1991] Piatetsky-Shapiro, G. (1991). Discovery, analysis and presentation of strong rules. In Piatetsky-Shapiro, G. and Frawley, W., editors, *Knowledge Discovery in Databases*, pages 229–248. AAAI/MIT Press.
- [Picouet and Lenca, 2001] Picouet, P. and Lenca, P. (2001). *Bases de données et internet*, chapter Extraction de connaissances à partir des données, pages 395–420. Hermes science.
- [Sebag and Schoenauer, 1988] Sebag, M. and Schoenauer, M. (1988). Generation of rules with certainty and confidence factors from incomplete and incoherent learning bases. In Boose, J., Gaines, B., and Linster, M., editors, *Proc. of the European Knowledge Acquisition Workshop (EKAW'88)*, pages 28–1 – 28–20. Gesellschaft für Mathematik und Datenverarbeitung mbH.
- [Silberschatz and Tuzhilin, 1995] Silberschatz, A. and Tuzhilin, A. (1995). On subjective measures of interestingness in knowledge discovery. In *Knowledge Discovery and Data Mining*, pages 275–281.
- [Tan et al., 2002] Tan, P.-N., Kumar, V., and Srivastava, J. (2002). Selecting the right interestingness measure for association patterns. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 32–41.
- [Terano et al., 2000] Terano, T., Liu, H., and Chen, A. L. P., editors (2000). *Association Rules*, volume 1805 of *Lecture Notes in Computer Science*. Springer.
- [Vaillant et al., 2003a] Vaillant, B., Picouet, P., and Lenca (2003a). An extensible platform for rule quality measure benchmarking. In *Human Centered Processes (HCP'2003)*, Luxembourg. To be published.
- [Vaillant et al., 2003b] Vaillant, B., Picouet, P., and Lenca (2003b). Une plateforme extensible pour l'évaluation de mesures de qualité de règles. *RSTI-RIA (EGC03)*, 1(17):556.