

Evaluation et analyse multicritère des mesures de qualité des règles d'association

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

*GET ENST Bretagne / Département IASC – CNRS TAMCIC
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

***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. Le nombre particulièrement important de règles générées par les algorithmes utilisés en extraction de connaissances à partir de données (ECD) ne permet pas aux utilisateurs de faire eux-mêmes la sélection des règles pertinentes. Un des problèmes centraux de l'ECD est le développement de mesures évaluant l'intérêt des règles découvertes. Ainsi de nombreuses mesures ont été proposées, parmi lesquelles l'expert est censé choisir celle qui est la plus appropriée à ses besoins. Mais le nombre élevé de mesures est lui-même un frein à la capacité de choix d'un expert. Pour y remédier, de nombreuses publications dressent des panoramas partiels de ces mesures, s'attachant tantôt à leur qualité algorithmique, tantôt à la formalisation de leurs propriétés, etc. Le résultat n'est malheureusement pas à la hauteur des espérances. A la multiplicité des mesures s'ajoute maintenant la diversité des caractéristiques, lesquelles ne reflètent généralement pas les objectifs de l'expert. Enfin, l'hétérogénéité des valeurs prises par ces caractéristiques n'est pas la moindre des difficultés que rencontre l'utilisateur quand il entreprend de comparer les mesures. Dans ce contexte, et malgré les efforts entrepris, le choix, par un utilisateur métier, de mesures répondant à son objectif reste un défi majeur de la recherche en ECD. La recherche des meilleures règles parmi le vaste ensemble de règles produit, passe aussi par la recherche et l'utilisation des bonnes mesures. On se trouve donc dans une problématique d'aide multicritère à la décision (AMD). Nous fondons notre approche sur l'évaluation de 20 mesures à partir de 8 critères.

Mots-clés : mesures d'intérêt, règles d'association, aide à la décision.

1 Introduction

Un des principaux objectifs de l'Extraction de Connaissances à partir des Données (ECD) est la production de règles intéressantes du point de vue des utilisateurs. Ceux-ci ne sont généralement pas des experts en ECD, mais des experts métier, spécialistes du domaine sur lequel portent les données.

Il est aujourd'hui notoire que l'intérêt d'une règle est une information particulièrement difficile à évaluer objectivement. Idéalement, une règle intéressante doit être "valide, nouvelle et compréhensible" [Fayyad *et al.*, 1996], mais ces termes génériques, appliqués à des contextes précis, couvrent des réalités très différentes.

Dans [Silberschatz et Tuzhilin, 1995], l'intérêt d'une règle est ainsi analysé d'un double point de vue : la concordance avec les croyances de l'expert métier, et la capacité à la transformer en action. De tels modèles s'avèrent néanmoins extrêmement difficiles à mettre en œuvre.

Dans ce contexte, l'utilisation de mesures (ou encore indices) évaluant l'intérêt des règles est un des principaux défis de l'ECD. De tels indices, en permettant d'extraire les "meilleures règles" d'un ensemble trop volumineux pour que chaque règle soit examinée directement par un expert métier, jouent un rôle majeur dans les processus d'ECD. La sélection des "meilleures" règles passe par l'utilisation de mesures adaptées au contexte et les mesures de qualité doivent être étudiées et manipulées dans une perspective d'aide à la décision [Lenca *et al.*, 2003b]. Le choix de la mesure d'intérêt adaptée aux préférences d'un utilisateur métier reste néanmoins un problème difficile. De fait, cette difficulté est justifiée étant donné que les indices présentent des caractéristiques variées et parfois contradictoires [Tan *et al.*, 2002, Lenca *et al.*, 2003b]. La mesure *optimale* n'existe pas et la qualité d'une règle doit prendre en compte plusieurs critères. [Francisci *et al.*, 2003] proposent ainsi une approche basée sur un algorithme génétique afin de sélectionner des règles réalisant un compromis. A titre d'exemple, le support (ou taux de couverture), pourtant célèbre, présente un bilan contrasté : d'un côté, sa propriété d'antimonotonie en fait un filtre particulièrement intéressant pour les algorithmes de génération de règles [Agrawal *et al.*, 1993] [Pasquier *et al.*, 1999]; pour autant, il possède à peu près tous les défauts qu'un utilisateur peut craindre d'un tel filtre (variabilité de ses valeurs pour les règles logiques et à l'hypothèse d'indépendance, linéarité) [Picouet et Lenca, 2001].

Plusieurs travaux ont ainsi été consacrés à l'étude formelle de telles mesures et à la formalisation de caractéristiques significatives (en particulier [Tan et Kumar, 2000], [Hilderman et Hamilton, 2001] et [Lallich et Teytaud, 2004]). En particulier, l'importance de disposer de critères objectifs d'évaluation des indices de qualité a été soulignée dans [Piatetsky-Shapiro, 1991] et [Freitas, 1999]. La pertinence de ces critères du point de vue des utilisateurs métier reste néanmoins difficile à évaluer. Dans [Tan *et al.*, 2002], les auteurs réalisent une étude comparative d'une vingtaine de mesures en fonction de telles caractéristiques. Cette étude se distingue de la nôtre par le fait qu'ils ne prennent en compte que des mesures symétriques ou symétrisées, alors que nous nous focalisons sur les mesures dissymétriques (cette restriction est justifiée section 2). Ils proposent une approche originale pour que l'expert puisse sélectionner la mesure la plus appro-

priée. Curieusement, cette méthode¹ n'utilise pas l'étude comparative présentée en début d'article, ce qui confirme de notre point de vue l'inadaptation de ces critères au problème de choix d'une mesure de qualité correspondant aux besoins de l'expert.

Si notre travail se focalise également sur le problème du choix de la mesure de qualité par un expert métier, notre article propose toutefois une approche alternative aux travaux précédents. Nous avons déjà montré l'intérêt d'une évaluation des caractéristiques des mesures [Lallich et Teytaud, 2004], ainsi que celui d'une aide multicritère à la décision (AMD) [Lenca *et al.*, 2003b]. Ce papier est une version développée de [Lenca *et al.*, 2003a]. Nous y présentons entre autres des aspects méthodologiques et l'espace problème, non développés dans nos précédents travaux. Des résultats d'une étude expérimentale sur des données y sont également présentés. Ces résultats renforcent la nécessité d'une aide à la recommandation pour l'utilisation d'un ensemble de mesures adapté aux objectifs de l'utilisateur. En effet, même si cette étude est présentée sur peu de données, le seul fait que les résultats puissent varier fortement selon la mesure utilisée justifie une aide à la sélection des mesures de qualité.

Nous développons un processus de choix en deux étapes :

- dans un premier temps, nous proposons une évaluation d'un ensemble de mesures en fonction d'une liste de propriétés en partie différentes de celles proposées dans [Tan *et al.*, 2002] : en effet, certaines d'entre elles s'avèrent mal adaptées au contexte des règles d'association, tandis que d'autres s'avèrent inutiles en ne permettant pas de faire la distinction entre les mesures d'intérêt que nous étudions. Nous introduisons en revanche de nouvelles propriétés, telles que la facilité de fixer un seuil, qui nous semble déterminante pour faciliter l'utilisation d'une mesure par un expert métier.
- dans un deuxième temps, nous suggérons d'utiliser une méthode d'aide multicritère à la décision sur l'ensemble de propriétés précédemment établi. Ces méthodes ont fait preuve de leur utilité dans de nombreux et différents domaines. Nous croyons qu'elles peuvent s'avérer d'un secours particulièrement pertinent pour le problème spécifique du choix par un utilisateur métier d'une mesure (ou de plusieurs mesures) de qualité de règles. Notre approche est résolument anthropocentrée. Le choix d'une mesure est inséparable des buts et des préférences de l'expert métier. Nous nous focalisons particulièrement sur ces deux points afin de les traduire en critères d'évaluation des mesures.

L'article est organisé de la façon suivante. Dans la prochaine section, nous présentons la liste des 20 mesures que nous avons considérées. Nous évaluons ces mesures sur des jeux de données en section 3 afin de montrer la diversité des ensembles de règles obtenus. Les sections 4 et 5 posent le problème de la sélection des n meilleures règles en termes d'utilisation de mesures adaptées aux priorités de l'utilisateur et concluent à l'intérêt d'une approche AMD. La section 6 détaille les critères de sélection des mesures que nous avons retenus. Enfin les sections 7 et 8 décrivent la méthode d'aide à la décision utilisée ainsi qu'une évaluation des mesures étudiées sur les critères proposés.

¹A partir de l'ensemble des règles issues d'un algorithme de fouille de données, les auteurs proposent d'extraire un petit sous-ensemble de règles sur lesquelles les différentes mesures produisent des résultats variés ; les auteurs montrent expérimentalement que la diversité des résultats donnés par les différentes mesures permettent à l'utilisateur métier de sélectionner la mesure la mieux adaptée.

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 *items* qui constituent A sont dans une transaction, alors le plus souvent les items qui constituent B le sont aussi. Les règles d'association $A \rightarrow B$ se focalisent sur les coprésences en traitant les variables de façon non symétrique. Elles doivent donc être distinguées 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 (cf. figure 1). C'est ainsi que sont d'emblée exclues certaines mesures comme le χ^2 , le r^2 de Pearson, la J-mesure de Goodman et Smyth [Goodman et Smyth, 1988] ou encore la mesure de Pearl [Pearl, 1988].

Soit E un ensemble de données et $n = |E|$.

Pour $A \rightarrow B$, on note :

$n_a = |A|$, le nombre de données vérifiant A

$n_b = |B|$, le nombre de données vérifiant B

$n_{ab} = |A \cap B|$, le nombre d'exemples de la règle

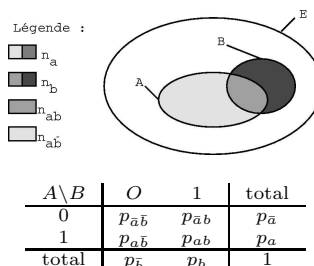
$n_{a\bar{b}} = |A \cap \bar{B}|$, le nombre de contre-exemples à

la règle, $n_{a\bar{b}} = n_a - n_{ab}$

$A \rightarrow B$ est évaluée à l'aide de mesures monotones décroissantes en fonction de $n_{a\bar{b}}$.

$A \rightarrow B$ est jugée intéressante selon la mesure μ lorsque $\mu(A \rightarrow B) \geq \alpha$, α devant être fixé par l'utilisateur.

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



On remarque que 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. Celles-ci ainsi que les références bibliographiques correspondantes sont rassemblées dans les tableaux 1 et 2.

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 (ensembles fréquents). 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

	Mesures	Références
SUP	support	[Agrawal <i>et al.</i> , 1993]
CONF	confiance	[Agrawal <i>et al.</i> , 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 <i>et al.</i> , 2000]
- INDIMP	indice d'implication	[Lerman <i>et al.</i> , 1981]
LIFT	Lift	[Brin <i>et al.</i> , 1997b]
SURP	surprise	[Azé et Kodratoff, 2002]
SEB	Sebag et Schoenauer	[Sebag et Schoenauer, 1988]
MC	multiplicateur de cote	[Lallich et Teytaud, 2004]
CONV	conviction	[Brin <i>et al.</i> , 1997a]
TEC	taux d'exemples et de contre-exemples	
IQC	indice de qualité de Cohen	[Cohen, 1960]
GI	gain informationnel	[Church et Hanks, 1990]
INTIMP	intensité d'implication	[Gras <i>et al.</i> , 1996]
IIE	intensité d'implication entropique	[Gras <i>et al.</i> , 2001]
IPD	indice probabiliste discriminant	[Lerman et Azé, 2003]
LAP	Laplace	[Good, 1965]

TAB. 1 – Liste des mesures éligibles

gain informationnel est une transformation monotone croissante du lift.

Certaines mesures privilégient les contre-exemples, comme la conviction $p_{\bar{b}|a}/p_{\bar{b}}$ 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 un risque relatif calculé en comparant la cote de B suivant celle de A et celle de \bar{B} , se distinguant de l'odd-ratio qui comparerait les cotes de B suivant A et \bar{A} . Enfin, 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 (ce qui est illustré par les valeurs égales à 1 dans le tableau 3) : par exemple, la confiance, la mesure de Sebag, le taux d'exemples et de contre-exemples ; le lift et le gain informationnel ; l'indice de Loevinger et la conviction.

Choix d'une mesure de qualité adaptée au contexte

	Déf. en fréquences absolues	Déf. en fréquences relatives
SUP	$\frac{n_a - n_{a\bar{b}}}{n}$	p_{ab}
CONF	$1 - \frac{n_{a\bar{b}}}{n_a}$	$p_{b/a}$
R	$\frac{nn_{ab} - n_a n_b}{\sqrt{nn_a n_b n_{a\bar{b}} \cdot n_{\bar{b}}}}$	$\frac{p_{ab} - p_a p_b}{\sqrt{p_a p_a p_b p_{\bar{b}}}}$
CONF CEN	$\frac{nn_{ab} - n_a n_b}{nn_a}$	$p_{b/a} - p_b$
PS	$\frac{1}{n} \left(\frac{nn_{a\bar{b}}}{n} - n_{a\bar{b}} \right)$	$n p_a (p_{b/a} - p_b) = n p_a p_b (\text{Lift} - 1)$
LOE	$1 - \frac{n_{a\bar{b}}}{n_a n_{\bar{b}}}$	$\frac{p_{b/a} - p_b}{p_{\bar{b}}} = \frac{1}{p_{\bar{b}}} \text{ConfCen} = 1 - \frac{1}{\text{Conv}}$
ZHANG	$\frac{nn_{ab} - n_a n_b}{\max\{n_{ab} n_{\bar{b}}, n_b n_{a\bar{b}}\}}$	$\frac{p_{ab} - p_a p_b}{\max\{p_{ab} p_{\bar{b}}, p_b (p_a - p_{ab})\}}$
- INDIMP	$\frac{nn_{a\bar{b}} - n_a n_{\bar{b}}}{\sqrt{nn_a n_{\bar{b}}}}$	$\sqrt{n} \frac{p_{a\bar{b}} - p_a p_{\bar{b}}}{\sqrt{p_a p_{\bar{b}}}}$
LIFT	$\frac{nn_{ab}}{n_a n_b}$	$\frac{p_{ab}}{p_a p_b}$
SURP	$\frac{n_{ab} - n_{a\bar{b}}}{n_b}$	$\frac{p_{ab} - p_{a\bar{b}}}{p_b} = 2 \frac{p_a}{p_b} (\text{Conf} - 0.5)$
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	$\frac{n_a n_{\bar{b}}}{n_a n_b}$	$\frac{p_a p_{\bar{b}}}{p_{ab}}$
TEC	$\frac{n_a - 2n_{a\bar{b}}}{n_a - n_{a\bar{b}}} = 1 - \frac{1}{\frac{n_a}{n_{a\bar{b}}} - 1}$	$1 - \frac{p_{a\bar{b}}}{p_{ab}} = 1 - \frac{1}{\text{Seb}}$
IQC	$2 \frac{nn_a - nn_{a\bar{b}} - n_a n_b}{nn_a + nn_b - 2n_a n_b}$	$2 \frac{p_{ab} - p_a p_b}{p_a + p_b - 2p_a p_b}$
GI	$\log \left(\frac{nn_{ab}}{n_a n_b} \right)$	$\log \frac{p_{ab}}{p_a p_b} = \log (\text{Lift})$
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 ²	$\left\{ \left[(1 - h_1 \left(\frac{n_{a\bar{b}}}{n} \right)^2) (1 - h_2 \left(\frac{n_{a\bar{b}}}{n} \right)^2) \right]^{1/4} \text{IntImp} \right\}^{1/2}$	$\left\{ \left[(1 - h_1(p_{ab}))^2 (1 - h_2(p_{ab}))^2 \right]^{1/4} \text{IntImp} \right\}^{1/2}$
IPD ³	$P \left[\mathcal{N}(0, 1) > \text{IndImp}^{CR/B} \right]$	
LAP	$\frac{n_{ab} + 1}{n_a + 2}$	$\frac{p_{b/a} + \frac{n}{p_a}}{1 + \frac{2n}{p_a}}$

TAB. 2 – Mesures éligibles

3 Approche expérimentale

Le choix d'une mesure de qualité, afin de filtrer ou d'ordonner les règles issues d'un algorithme de fouille de données, influence le résultat du processus d'ECD. L'approche AMD doit permettre de fournir des recommandations à l'utilisateur final confronté au choix d'une ou plusieurs mesures. Elle permet aussi de proposer une cartographie

²avec :

$$h_1(t) = -\left(1 - \frac{t}{p_a}\right) \log_2 \left(1 - \frac{t}{p_a}\right) - \frac{t}{p_a} \log_2 \left(\frac{t}{p_a}\right) \text{ pour } t \in [0, p_a/2], h_1(t) = 1 \text{ sinon ;}$$

$$h_2(t) = -\left(1 - \frac{t}{p_b}\right) \log_2 \left(1 - \frac{t}{p_b}\right) - \frac{t}{p_b} \log_2 \left(\frac{t}{p_b}\right) \text{ pour } t \in [0, p_b/2], h_2(t) = 1 \text{ sinon.}$$

³ $\text{IndImp}^{CR/B}$ correspond à INDIMP centré réduit (CR) pour une base de règles admissibles B.

des mesures en regroupant les mesures aux comportements similaires (et en distinguant celles aux comportements différents) en fonction du contexte (nous précisons cette notion section 4). Dans cette section, nous illustrons ceci au moyen de résultats expérimentaux sur les préordres engendrés par les mesures.

3.1 Comparaison de préordres

Les valeurs d'une mesure de qualité pour un ensemble de règles donné engendrent un préordre total sur cet ensemble. Il est ainsi possible de mesurer l'influence du choix de la mesure de qualité sur l'ordonnement des règles en comparant ces préordres. [Giakoumakis et Monjardet, 1987] recensent 16 coefficients d'accord entre préordres totaux dérivés du τ de Kendall ou du ρ de Spearman.

Étant donné deux mesures μ_1 et μ_2 , et un couple de règles (r_1, r_2) , r_1 et r_2 pouvant représenter la même règle, on distingue six situations :

- il y a accord strict lorsque $\mu_1(r_1) < \mu_1(r_2)$ et $\mu_2(r_1) < \mu_2(r_2)$.
- il y a co-accord lorsque $\mu_1(r_1) > \mu_1(r_2)$ et $\mu_2(r_1) > \mu_2(r_2)$.
- il y a accord large lorsque $\mu_1(r_1) = \mu_1(r_2)$ et $\mu_2(r_1) = \mu_2(r_2)$.
- il y a semi-accord lorsque pour l'une des mesures, la valeur prise pour r_1 est inférieure strictement à la valeur prise pour r_2 , et que les valeurs sont égales pour l'autre mesure.
- il y a semi-désaccord s'il y a semi-accord pour le couple (r_2, r_1) .
- il y a désaccord strict si $\mu_1(r_1) < \mu_1(r_2)$ et $\mu_2(r_1) > \mu_2(r_2)$ ou si $\mu_1(r_1) > \mu_1(r_2)$ et $\mu_2(r_1) < \mu_2(r_2)$.

Nous utilisons le coefficient τ_1 , qui, étant donné ces 6 situations distinctes, vaut $\frac{2(a-d)+2((c-n)/2-b)}{n(n-1)}$, où a est le nombre d'accords stricts, b le nombre de semi-accords, c le nombre d'accords larges, $2d$ le nombre de désaccords stricts, et n le nombre de règles.

Ce coefficient vérifie les propriétés suivantes :

- les valeurs prises par le coefficient sont comprises entre -1 et 1 .
- le coefficient vaut 1 si et seulement si les deux préordres sont égaux.
- le coefficient vaut -1 si et seulement si chaque couple de règles distinctes est soit en désaccord strict, soit semi-désaccord, soit semi-accord.

3.2 Protocole expérimental

La comparaison des préordres a été effectuée sur 2 bases de règles d'association, générées par l'implémentation de l'algorithme APRIORI de [Borgelt et Kruse, 2002], utilisée avec le paramètre **-m2** afin d'assurer la présence d'au moins un attribut dans la prémisse et d'un attribut dans la conclusion.

Des bases de cas, mises à disposition par l'UCI ⁴ et DELVE ⁵, ont été utilisées pour

⁴<http://www.ics.uci.edu/~mllearn/MLSummary.html>

⁵<http://www.cs.toronto.edu/~delve/data/datasets.html>

Choix d'une mesure de qualité adaptée au contexte

générer les bases de règles **cmc_apriori** et **titanic_apriori**, décrites ci-dessous :

cmc_apriori : seuil de support minimal de 10% et un seuil de confiance minimale de 80% (options par défaut). Elle comporte 444 règles.

titanic_apriori seuil de support minimal de 5% et un seuil de confiance minimale de 80%. Elle comporte 51 règles.

Étant donné un couple de mesures de qualité à comparer sur une base de n règles, l'évaluation de τ_1 s'effectue en $O(n^2)$. Le calcul de ce coefficient d'accord τ_1 a été implémenté au sein d'un outil d'étude de mesures de qualité sur des données, HERBS ([Vaillant, 2002], [Vaillant *et al.*, 2003]). Seules 18 mesures des 20 étudiées formellement dans cet article sont actuellement codées dans HERBS, ce qui explique qu'il n'y a que 18 mesures dans les tableaux 3, 4 et 5. Cet outil, écrit en PERL/TK, interfacé à un gestionnaire de bases de données (MYSQL⁶) ainsi qu'à un outil d'analyse statistique (R⁷), a pour objectif de faciliter la post-analyse de résultats issus d'un processus d'ÉCD, en particulier l'étude du comportement de mesures de qualité sur des données.

3.3 Résultats et analyse

La table 3 présente les résultats de la comparaison des préordres engendrés par les mesures de qualité sur la base de règles **cmc_apriori**. Les résultats pour les 2 bases de règles sont présentés de manière schématique dans les tableaux 4 et 5.

Afin d'effectuer ces représentations graphiques, les valeurs du coefficient d'accord ont été normalisées entre 0 et 1 par la transformation affine $(1 + \tau_1)/2$ (la valeur correspondante est proportionnelle à la surface de chaque carré).

Les lignes et les colonnes ont été réorganisées grâce au logiciel AMADO (Analyse graphique d'une MATrice de DONnées) [Chauchat et Risson, 1998] suivant les principes de sémiologie graphique de [Bertin, 1977] : un tableau est plus lisible et sa représentation graphique est plus explicite lorsque les lignes et les colonnes du tableau sont réécrites dans l'ordre qui fait le mieux apparaître une éventuelle structure diagonale ou par blocs du tableau. Le logiciel AMADO utilise les méthodes de codage optimal de variables qualitatives pour mettre en œuvre ces principes.

On retrouve les groupes de mesures qui se déduisent les unes des autres par transformation monotone croissante : (CONF, SEB, TEC, LAP), (LIFT, GI), (LOE, CONV) et (-INDIMP, IPD).

On voit également que SUP et SURP ordonnent les règles de manière très différente des autres mesures. Deux mesures sont proches de LOE et CONV, ce sont ZHANG et MC.

⁶<http://www.mysql.com/>

⁷<http://www.R-project.org/>

	IQC	CONF	CONFcen	CONV	GI	-INDIMP	IPD	LAP	LIFT	LOE	MC	PS	R	SEB	SUP	SURP	TEC	ZHANG	
IQC	1																		
CONF	0,284	1																	
CONFcen	0,780	0,408	1																
CONV	0,636	0,614	0,774	1															
GI	0,783	0,390	0,980	0,754	1														
-INDIMP	0,839	0,425	0,837	0,796	0,820	1													
IPD	0,839	0,425	0,837	0,796	0,820	1	1												
LAP	0,290	0,984	0,397	0,604	0,380	0,432	0,432	1											
LIFT	0,783	0,390	0,980	0,754	1	0,820	0,820	0,380	1										
LOE	0,636	0,614	0,774	1	0,754	0,796	0,796	0,604	0,754	1									
MC	0,663	0,587	0,805	0,969	0,785	0,823	0,823	0,577	0,785	0,969	1								
PS	0,965	0,312	0,805	0,669	0,803	0,873	0,873	0,318	0,803	0,669	0,696	1							
R	0,873	0,396	0,830	0,763	0,814	0,966	0,966	0,403	0,814	0,763	0,79	0,902	1						
SEB	0,284	1	0,408	0,614	0,390	0,425	0,425	0,984	0,390	0,614	0,587	0,312	0,396	1					
SUP	0,143	0,097	-0,026	0,028	-0,03	0,089	0,089	0,113	-0,03	0,028	0,021	0,134	0,101	0,097	1				
SURP	0,383	0,147	0,215	0,224	0,210	0,329	0,329	0,161	0,210	0,224	0,229	0,374	0,343	0,147	0,753	1			
TEC	0,284	1	0,408	0,614	0,390	0,425	0,425	0,984	0,390	0,614	0,587	0,312	0,396	1	0,097	0,147	1		
ZHANG	0,663	0,587	0,805	0,969	0,785	0,823	0,823	0,577	0,785	0,969	1	0,696	0,79	0,587	0,021	0,229	0,587	1	

TAB. 3 – Comparaison de préordres; base de règles **cmc_apriori**

	IQC	PS	GI	LIFT	CONFcen	R	-INDIMP	IPD	MC	ZHANG	CONV	LOE	SURP	CONF	SEB	TEC	LAP
IQC	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
PS	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
GI	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
LIFT	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
CONFcen	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
R	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
-INDIMP	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
IPD	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
MC	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
ZHANG	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
CONV	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
LOE	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
SURP	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
CONF	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
SEB	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
TEC	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
LAP	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

TAB. 4 – Comparaison de préordres; base de règles **cmc_apriori**

	CONF	SEB	TEC	LAP	CONV	LOE	MC	ZHANG	CONFcen	LIFT	-INDIMP	IPD	R	PS	IQC	SURP
CONF	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
SEB	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
TEC	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
LAP	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
CONV	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
LOE	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
MC	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
ZHANG	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
CONFcen	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
LIFT	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
-INDIMP	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
IPD	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
R	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
PS	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
IQC	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
SURP	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

TAB. 5 – Comparaison de préordres; base de règles **titanic_apriori**

4 Positionnement du problème

Nous avons montré que la recherche des meilleures règles parmi le vaste ensemble de règles produit lors d'un processus ECD, passe aussi par la recherche et l'utilisation des bonnes mesures. Ces mesures étant décrites par de nombreuses propriétés, nous montrons que l'on se trouve dans une problématique AMD. Ainsi posé, du point de vue de l'utilisateur, il s'agit de trouver la ou les bonnes mesures à utiliser en fonction du contexte. Ce contexte est défini par de nombreux paramètres. Parmi ces derniers, notons la nature des données (quel est leur type, sont-elles bruitées?), l'algorithme utilisé pour extraire les règles (quels sont les biais inhérents?) et les buts et préférences de l'utilisateur. Nous nous focalisons dans cet article sur ces deux derniers points.

Nous posons le problème à partir d'un sextuplet $\langle \mathcal{D}, \mathcal{R}, \mathcal{M}, \mathcal{A}, \mathcal{P}, \mathcal{F} \rangle$ où :

- \mathcal{D} est un ensemble de données; les données sont décrites par une liste d'attributs (de diverses natures);
- \mathcal{R} est un ensemble de règles de la forme $A \rightarrow B$ s'appliquant sur \mathcal{D} où A est l'antécédent (ou la condition d'application de la règle) et B le conséquent (ou la conclusion de la règle); A et B sont des formes logiques sur les attributs; dans cette étude nous nous intéressons au cas particulier des règles d'association telles que définies par [Agrawal *et al.*, 1993];
- \mathcal{M} est un ensemble de mesures (nous utiliserons indifféremment mesure ou indice pour parler des éléments de \mathcal{M}) de la qualité des règles de \mathcal{R} (cf. section 2 tableau 1);
- \mathcal{A} est un ensemble de propriétés permettant de décrire les caractéristiques des mesures de \mathcal{M} (cf. section 6);
- \mathcal{P} est un ensemble de préférences exprimées par l'utilisateur expert (du domaine \mathcal{D} sur lequel s'applique \mathcal{R}) sur \mathcal{A} en fonction de ses objectifs; la difficulté principale de la construction de \mathcal{P} est la formalisation d'objectifs utilisateurs. Ces derniers étant bien souvent exprimés en langage naturel, il s'agit d'en conserver la sémantique;
- \mathcal{F} est un ensemble de critères d'évaluation des mesures de \mathcal{M} ; \mathcal{F} est construit à partir des ensembles \mathcal{A} et \mathcal{P} . Pour simplifier, l'ensemble \mathcal{F} correspond à une évaluation des mesures de \mathcal{M} sur les propriétés de \mathcal{A} en fonction des préférences de \mathcal{P} (cette évaluation est réalisée sur au moins une échelle ordinale).

Les mesures de qualité étudiées dans cet article n'évaluent que les qualités des règles prises individuellement. Nous n'évaluons ni la qualité globale des ensembles de règles \mathcal{R} ni la qualité de sous-ensembles de \mathcal{R} . On notera la particularité IPD qui est une version contextuelle de INDIMP afin d'accroître le pouvoir discriminant de INDIMP. Elle reste cependant une mesure évaluant la qualité d'une règle et non d'un ensemble de règles.

On distinguera principalement deux intervenants : l'utilisateur expert des données (expert de \mathcal{D} et \mathcal{R}) qui cherche à sélectionner les "meilleures" règles de \mathcal{R} et l'analyste, spécialiste de l'aide à la décision et de l'extraction des connaissances à partir des données mais certainement pas spécialiste du domaine de l'expert, qui cherche à l'aider. Nous noterons E_r le premier et E_a le second. Le problème principal consiste alors à traduire les propriétés de \mathcal{A} en un ensemble \mathcal{F} de critères à partir des préférences \mathcal{P} afin de mettre en œuvre une méthode répondant à l'une des problématiques de référence présentées section 5. Notons que les ensembles \mathcal{D} , \mathcal{R} et \mathcal{P} relèvent prioritairement de

l'expertise de E_r tandis que les ensembles \mathcal{M} , \mathcal{A} et \mathcal{F} relèvent principalement de l'expertise de E_a . La résolution du problème traité implique une collaboration étroite et un dialogue permanent entre les différents acteurs : le spécialiste E_a a besoin de connaître les préférences \mathcal{P} et les objectifs de l'utilisateur expert E_r . Ces préférences pourront être modélisées et former une famille de critères \mathcal{F} pour une aide à la sélection des bonnes mesures.

5 Caractériser les mesures pour mieux les choisir

Des statistiques exploratoires à la théorie de l'information, nombreux sont les domaines qui ont contribué à la profusion de mesures de la qualité des règles. Chaque mesure proposée répond à un besoin particulier, tour à tour algorithmique, qualitatif ou subjectif. Par exemple [Piatetsky-Shapiro, 1991] impose de ne présenter que des connaissances statistiquement significatives, [Agrawal *et al.*, 1993] soulignent l'intérêt d'indices tels que le support pour servir d'opérateur de filtrage efficace dans un contexte d'exploration de treillis d'*items*, [Gras *et al.*, 2001] s'intéressent à l'invraisemblance du nombre de contre-exemples, [Azé et Kodratoff, 2002] à la résistance au bruit, etc. Du point de vue subjectif, la mesure proposée par [Silberschatz et Tuzhilin, 1996] compare les règles produites à une représentation de la connaissance de l'utilisateur, espérant ainsi coller au mieux à la notion d'intérêt de l'utilisateur, reste que ces modèles de connaissances sont très difficiles à produire eux-mêmes.

La diversité des mesures a entraîné une multiplication des tentatives de caractérisations formelles, mais la mesure parfaite n'existe pas. Malgré les nombreuses études comparatives, par exemple [Hilderman et Hamilton, 2001, Lallich et Teytaud, 2004], certes jamais exhaustives, le choix d'une bonne mesure reste un problème largement ouvert. Ce problème résulte de plusieurs éléments :

- une difficulté à comparer des mesures caractérisées par des propriétés très différentes les unes des autres; différentes en nature, et en valeur (symboliques ou numériques, ordonnées ou non).
- une difficulté à exprimer des critères utilisateur à l'aide de ces caractéristiques : par exemple comment retranscrire la propriété de résistance à des données bruitées en fonction des caractéristiques disponibles ?
- une difficulté à intégrer dans des études existantes des critères propres aux utilisateurs, par exemple la facilité de mise en œuvre, qui est très subjective.

Dans [Lenca *et al.*, 2003b] nous montrons qu'un processus d'aide multicritère à la décision peut être intéressant dans la prise en compte de ces différents aspects par l'utilisateur métier. Ce type d'approche distingue classiquement quatre problèmes de référence [Roy, 1985] (figure 2) :

- choisir un sous-ensemble de mesures : problématique de choix P_α (éclairer la décision par le choix d'un sous-ensemble aussi restreint que possible en vue par exemple d'un choix final d'une seule mesure, ce sous-ensemble contenant les "meilleures" mesures ou, à défaut, des mesures "satisfaisantes");
- déterminer toutes les bonnes mesures : problématique de tri P_β (éclairer la décision par un tri résultant d'une affectation de chaque mesure à une catégorie, les catégories étant définies *a priori* en fonction de normes ayant trait à la suite

Choix d'une mesure de qualité adaptée au contexte

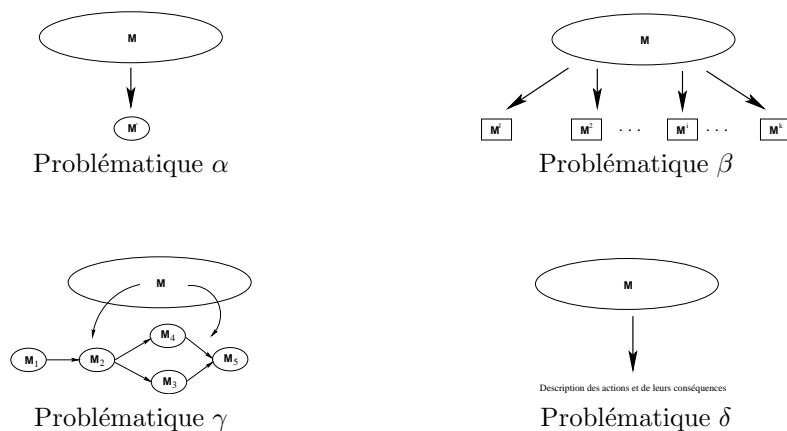


FIG. 2 – Problématiques de référence en aide à la décision

- à donner aux mesures qu'elles sont destinées à recevoir);
- classer les mesures de la meilleure à la moins bonne : problématique de rangement P_γ (éclairer la décision par un rangement obtenu en regroupant tout ou une partie (les plus satisfaisantes) des mesures en classes d'équivalence, ces classes étant ordonnées, de façon complète ou partielle, conformément aux préférences);
- décrire les mesures et/ou leurs conséquences de façon formalisée : problématique P_δ (éclairer la décision par une description, dans un langage approprié, des mesures et de leurs conséquences).

Une analyse peut correspondre soit à l'une des problématiques de référence soit à une combinaison de plusieurs de ces problématiques. Dans l'état actuel des choses, en cherchant à réaliser un état de l'art (description complète et rigoureuse des mesures existantes, des qualités manquantes, etc.) le groupe GAFOQUALITÉ de l'action spécifique STIC *Fouille de Bases de Données* est clairement dans la problématique δ . Dans [Lenca *et al.*, 2003b] nous avons abordé les problématiques α et γ sur quelques mesures et quelques critères. Nous proposons dans cet article une étude beaucoup plus complète d'un ensemble de 20 mesures sur 8 critères qu'il nous semble important de considérer.

Pour un décideur donné, une telle analyse procède généralement en quatre grandes étapes (avec d'éventuels retours en arrière) : établir la liste \mathcal{M} des mesures potentielles (section 2), établir la famille \mathcal{F} des critères (section 6) servant à évaluer les mesures (\mathcal{F} est construite à partir des attributs de \mathcal{A} et des préférences \mathcal{P} du décideur), évaluer les mesures de \mathcal{M} sur les critères de \mathcal{F} afin de bâtir une ou plusieurs matrices de décision, et enfin agréger les performances en fonction de la problématique choisie (section 7).

Les trois premières étapes vont conduire à une cartographie des indices. Elles sont généralement communes à toutes les méthodes et à toutes les problématiques. Il existe en revanche de très fortes différences dans la manière d'agréger les performances. Nous traitons dans cet article de la problématique de type γ .

6 Critères d'évaluation

Nous présentons ci-dessous huit propriétés intéressantes caractérisant (et définissant l'ensemble \mathcal{A}) le comportement des mesures éligibles étudiées dans cet article (ensemble \mathcal{M}). Nous proposons des valeurs possibles et des évaluations ordinales de ces propriétés afin de les transformer en critères g_i d'évaluation des mesures. Notre ensemble \mathcal{F} de critères est donc composé de huit critères ordinaux $\{g_1, \dots, g_8\}$ (le codage $\{0, 1, 2, \dots\}$ utilisé pour les critères n'est qu'une représentation numérique de l'ordre sur les différentes valeurs pouvant être prises par chaque critère; la valeur numérique est en elle-même dénuée de sens et par exemple, pour g_5 , la valeur 2 ne signifie pas le double de 1). Les résultats d'une évaluation des différentes mesures suivant ces critères sont rassemblés dans le tableau 7. Nous proposons parfois des évaluations de façon normative si nous pensons que cela relève plutôt de \mathbf{E}_a (voir le tableau 6). Il est cependant important de noter que nous ne proposons pas une norme, l'évaluation présentée au tableau 7 est de *bon sens*, mais elle reste critiquable et surtout modifiable en fonction des préférences de \mathbf{E}_r .

6.1 Critère 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 μ évalue de façon différente les règles $A \rightarrow B$ et $B \rightarrow A$, ce qui se traduit par $\mu(A \rightarrow B) \neq \mu(B \rightarrow A)$. Dans le cas précis des règles d'association que nous étudions, cette propriété relève de \mathbf{E}_a et on définit le critère g_1 , *mesure non symétrique*, de façon normative à l'aide de deux valeurs ordonnées : $g_1(\mu) = 0$ si la mesure μ est symétrique, et $g_1(\mu) = 1$ si elle est dissymétrique.

6.2 Critère g_2 : décroissance avec n_b

Pour n_{ab} , $n_{a\bar{b}}$ et $n_{\bar{a}b}$ fixés, une règle est d'autant plus intéressante que n_b est faible. On attend donc d'une mesure qu'elle soit une fonction décroissante de n_b . Le centrage ou/et la division de la confiance par p_b permet que ce critère soit réalisé.

Le critère g_2 , *décroissance avec n_b* , est de la responsabilité de \mathbf{E}_a et nous lui attribuons deux valeurs : $g_2(\mu) = 0$ si la mesure μ n'est pas décroissante avec n_b , et $g_2(\mu) = 1$ si c'est une fonction décroissante avec n_b .

6.3 Critère 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 reste identique, 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 logique que la mesure lui attribue une valeur fixe, 0 ou 1 le plus fréquemment.

Cette propriété est fortement souhaitée et relève de l'expertise de \mathbf{E}_a . Le critère g_3 prend deux valeurs : $g_3(\mu) = 0$ si la mesure n'a pas une valeur fixe en cas d'indépendance et $g_3(\mu) = 1$ sinon.

6.4 Critère 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. Les mesures étudiées étant des fonctions décroissantes de $n_{a\bar{b}}$, une valeur fixe (1 par exemple), ou à la rigueur infinie, est alors attendue. Bien que cette valeur fixe puisse poser problème (perte du pouvoir discriminant de la mesure), elle est fréquemment souhaitée et relève principalement de l'expertise de \mathbf{E}_a . Le critère g_4 prend deux valeurs : $g_4(\mu) = 0$ si la mesure n'a pas une valeur fixe en cas de règle logique et $g_4(\mu) = 1$ sinon.

Nous écartons le critère constitué par l'évaluation de l'incompatibilité, une troisième 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 $A \rightarrow \bar{B}$ qui est alors une règle logique et non pas la règle $A \rightarrow B$.

6.5 Critère g_5 : non-linéarité en fonction de $p_{a\bar{b}}$ autour de 0^+

Différents auteurs ([Gras *et al.*, 2004]) souhaitent une décroissance lente au début avec l'apparition des premiers contre-exemples, 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. 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.

Cette propriété relève principalement de la tolérance de \mathbf{E}_r aux contre-exemples et contrairement aux précédents critères, il est plus difficile de définir ici un système normatif de valeurs. Reprenant les arguments de [Gras *et al.*, 2004], on peut noter $g_5(\mu) = 0$ si μ est convexe pour $n_{a\bar{b}}$ voisin de 0^+ , $g_5(\mu) = 1$ si elle est linéaire et $g_5(\mu) = 2$ si elle est concave. Une décroissance convexe qui accroît la sensibilité aux faux positifs peut être préférée dans certains cas.

6.6 Critère g_6 : prise en compte du nombre de cas n

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 la prise en compte de n dans de telles mesures est que celles-ci fournissent une grande masse de règles dont l'évaluation est proche du maximum et pour certaines perdent leur pouvoir discriminant, alors même que toutes 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. Le fait de satisfaire au critère 6 s'accompagne d'une perte de capacité discriminante.

On note 0 si la mesure est invariante en cas de dilatation des données et 1 si elle est croissante en cas de dilatation des données. Remarquons qu'une notation inverse peut aussi être justifiée en fonction des préférences de \mathbf{E}_r .

6.7 Critère g_7 : facilité à fixer un seuil d'acceptation de la règle

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 et Teytaud, 2004], 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 l'intensité d'implication [Gras, 1979] ou la mesure IPD [Lerman et 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. Cette propriété relève principalement de E_a .

6.8 Critère 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. On note 0 s'il y a impossibilité de donner le sens concret de la mesure, 1 si la mesure se ramène à des quantités usuelles et 2 si le sens concret de la mesure s'écrit en une phrase.

Le tableau 6 résume les différents critères utilisés dans notre étude.

Critère	Sémantique	Nombre de modalités	Responsabilité
g_1	non symétrie	2	E_a
g_2	décroissance avec n_b	2	E_a
g_3	situation à l'indépendance	2	E_a
g_4	situation à la règle logique	2	E_a
g_5	non-linéarité autour de 0^+	3	E_r
g_6	prise en compte de n	2	E_r
g_7	facilité à fixer un seuil	2	E_a et E_r
g_8	intelligibilité	3	E_r

TAB. 6 – Famille \mathcal{F} de critères

Les mesures de la section 2 sont évaluées à partir des critères ainsi définis et nous obtenons une matrice de décision présentée au tableau 7.

	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
IIE	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. 7 – Matrice de décision

7 La méthode PROMETHEE-GAIA

Dans cette section nous présentons PROMETHEE-GAIA, une méthode d'aide à la décision multicritère [Brans et Mareschal, 1994, Brans et Mareschal, 2002]. Son but est de traiter des problèmes de nature multicritère en vue d'un rangement (problématique γ). PROMETHEE se base sur des comparaisons de paires d'alternatives (μ_1, μ_2) de $\mathcal{M} \times \mathcal{M}$ pour les k critères de \mathcal{F} (elle ne nécessite pas des critères commensurables) afin de déterminer un indice de préférence agrégé $\pi(\mu_1, \mu_2)$. Cet indice sert ensuite à calculer des flots de surclassement en vue de la création d'un ordre partiel sur les mesures.

7.1 Comparaison paire par paire

Etant donné que cette méthode se base sur des comparaisons par paire d'actions, il est nécessaire de représenter et de formaliser le degré de préférence d'une alternative sur une autre. Ainsi, pour une paire d'alternatives (μ_1, μ_2) de $\mathcal{M} \times \mathcal{M}$, pour un critère particulier, une petite (respectivement grande) différence d'évaluation sera représentée par un degré de préférence faible (respectivement élevé). Les auteurs de la méthode suggèrent de représenter ces préférences par des réels de l'intervalle $[0, 1]$. Le concept de fonction de préférence P_j sert à calculer ce degré de préférence à partir des différences d'évaluations entre les alternatives pour un critère g_j donné :

$$P_j(\mu_1, \mu_2) = P_j(g_j(\mu_1) - g_j(\mu_2)), \quad \text{où } 0 \leq P_j(\mu_1, \mu_2) \leq 1.$$

Dans le cas d'un critère à maximiser, la fonction de préférence a l'allure de la fonction de préférence typique (figure 3). Il s'agit d'une fonction croissante et continue.

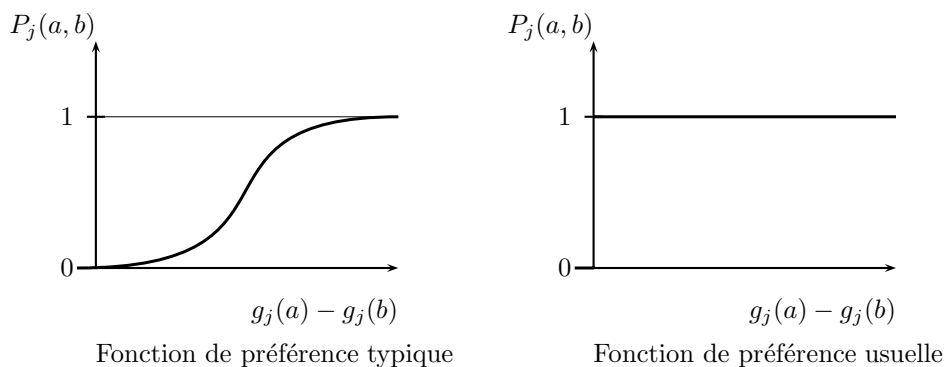


FIG. 3 – Fonctions de préférence

L'approche PROMETHEE propose six types de fonctions de préférence qui nécessitent que l'utilisateur fixe entre 0 et 3 paramètres (seuils de préférence et d'indifférence). Suivant les auteurs de cette méthode, ces six fonctions sont suffisantes pour l'analyse de la plupart des applications.

Vu la nature purement ordinale des critères retenus, nous considérons qu'une différence d'évaluation non nulle de a sur b doit être reflétée dans la préférence de a sur b . Nous choisissons par conséquent d'utiliser la fonction de préférence usuelle, qui vaut 1 entre $]0, +\infty[$ et 0 partout ailleurs (figure 3).

7.2 Indice de préférence agrégé

L'étape suivante consiste à agréger les degrés de préférence relatifs à chaque critère en un indice de préférence agrégé, $\pi(\mu_1, \mu_2)$, qui exprime le degré avec lequel μ_1 est préféré à μ_2 sur l'ensemble des critères :

$$\pi(\mu_1, \mu_2) = \sum_{j=1}^k P_j(\mu_1, \mu_2) \cdot w_j$$

où w_j est le poids associé au critère g_j sous la contrainte $\sum_{j=1}^k w_j = 1$. On voit que $\pi(a, a) = 0$ et $0 \leq \pi(a, b) \leq 1, \forall a, b \in \mathcal{M}$. Il est clair que pour une paire donnée (μ_1, μ_2) d'alternatives, un indice de préférence proche de 0 (respectivement 1) implique une préférence globale faible (respectivement forte) de μ_1 par rapport à μ_2 .

7.3 Flots de surclassement

Le rangement des n alternatives de \mathcal{M} est construit en comparant chacune des alternatives aux $n - 1$ autres. Ceci se fait par l'intermédiaire de deux flots de surclassement

Choix d'une mesure de qualité adaptée au contexte

définis comme suit :

- le flot de surclassement positif $\phi^+(\mu) = \frac{1}{n-1} \sum_{x \in \mathcal{M}} \pi(\mu, x)$
- le flot de surclassement négatif $\phi^-(\mu) = \frac{1}{n-1} \sum_{x \in \mathcal{M}} \pi(x, \mu)$

Le flot de surclassement positif exprime la force de l'alternative en question alors que le flot de surclassement négatif donne une indication sur sa faiblesse.

7.4 Les rangements

Deux rangements différents peuvent être obtenus par la méthode PROMETHEE : un préordre partiel et un préordre total.

Le préordre partiel est l'intersection des rangements qui peuvent être déduits des flots de surclassement positifs et négatifs. Il est construit de la façon suivante :

$$\left\{ \begin{array}{l} aP_p b \iff \left\{ \begin{array}{l} \phi^+(a) > \phi^+(b) \text{ et } \phi^-(a) < \phi^-(b) \\ M\phi^+(a) = \phi^+(b) \text{ et } \phi^-(a) < \phi^-(b) \\ \phi^+(a) > \phi^+(b) \text{ et } \phi^-(a) = \phi^-(b) \end{array} \right. \\ aI_p b \iff \phi^+(a) = \phi^+(b) \text{ et } \phi^-(a) = \phi^-(b) \\ aR_p b \text{ sinon} \end{array} \right.$$

où P , I et R représentent respectivement la préférence, l'indifférence et l'incomparabilité.

Le préordre total est construit à partir du flot de surclassement net $\phi(a)$, avec $\phi(a) = \phi^+(a) - \phi^-(a)$:

$$\left\{ \begin{array}{l} aP_t b \iff \phi(a) > \phi(b) \\ aI_t b \iff \phi(a) = \phi(b) \end{array} \right.$$

7.5 Le plan GAIA

Après la détermination des rangements, la méthode PROMETHEE permet à l'utilisateur de visualiser les alternatives et les critères dans un plan commun appelé le plan GAIA. Il permet d'avoir une vue synthétique et claire du caractère conflictuel de certains critères, et de l'impact des poids sur la décision finale. Il s'agit d'une projection similaire à l'analyse en composantes principales des données. Cependant l'analyse du plan GAIA a l'avantage de permettre de visualiser le comportement du rangement final par rapport à différentes configurations de poids.

La construction du plan GAIA se base sur l'analyse de flots nets particuliers, relatifs à chacun des critères, en décomposant le flot de surclassement net.

Ainsi, on a :

$$\phi(\mu) = \phi^+(\mu) - \phi^-(\mu) = \frac{1}{n-1} \sum_{j=1}^k \sum_{x \in \mathcal{M}} [P_j(\mu, x) - P_j(x, \mu)] w_j$$

Par conséquent :

$$\phi(\mu) = \sum_{j=1}^k \phi_j(\mu) w_j \tag{1}$$

où $\phi_j(\mu) = \frac{1}{n-1} \sum_{x \in \mathcal{M}} [P_j(\mu, x) - P_j(x, \mu)]$ est le flot net unicritère. Chaque alternative μ peut ainsi être caractérisée par ses k flots unicritères $\alpha(\mu) = (\phi_1(\mu), \dots, \phi_k(\mu))$ et être représentée dans un espace de dimension k dont les axes correspondent aux différents critères.

Le nuage de points correspondant est projeté dans un plan de dimension 2 de manière à présenter l'information de façon plus synthétique. Remarquons d'une part que dans l'espace de dimension k , le nuage de points représentant les alternatives est centré en l'origine, puisque $\sum_{\mu \in \mathcal{M}} \phi_j(\mu) = 0$. Cette projection est effectuée selon les deux premiers axes factoriels de l'analyse en composantes principales. D'autre part, dans le plan final, on représente aussi les projections des vecteurs unités des axes de coordonnées de l'espace original. Ces projections représentent alors les critères.

D'après la décomposition (1), le flot net d'une alternative est le produit scalaire entre le vecteur représentant son flot net unicritère et le vecteur des poids w . Ceci signifie en d'autres termes que le flot net de μ est aussi la projection de $\alpha(\mu)$ sur le vecteur w , dans l'espace de dimension k . Il en découle que la projection des $\alpha(x), \forall x \in \mathcal{M}$ sur w fournit le rangement total. Le vecteur w est donc clairement un *axe de décision*. Sa projection dans le plan GAIA est appelée l'axe de décision PROMETHEE π .

Notons quelques points importants pour l'évaluation et l'analyse du plan GAIA :

- plus l'axe d'un critère est long dans le plan GAIA, plus ce critère sera discriminant pour les alternatives.
- des critères représentant des préférences similaires (respectivement opposées) sur l'ensemble d'alternatives sont représentés par des axes de direction similaire (respectivement opposée).
- des critères indépendants sont représentés par des axes orthogonaux.
- des alternatives qui ont de “bonnes” valeurs pour un critère particulier sont représentées par des points proches de l'axe de ce critère.
- des alternatives similaires sont voisines dans le plan GAIA.
- si l'axe π est long, il a un pouvoir de décision fort, et le décideur est invité à choisir des alternatives dans la direction et le sens de l'axe.
- si l'axe π est court, il a un pouvoir de décision faible. Dans ce cas, le vecteur w est quasi perpendiculaire au plan GAIA. Ceci signifie que pour cette configuration de poids, les critères sont en conflit, et un bon compromis proche de l'origine peut être choisi.

8 Application au rangement d'indices de qualité de règles d'association

Cette section traite de l'analyse des indices décrits précédemment par la méthode d'aide à la décision multicritère PROMETHEE. L'expert E_r peut avoir des préférences très variées et nous examinons deux scénarios principaux de *bon sens* :

- **Sc1** : l'expert tolère l'apparition d'un certain nombre de contre-exemples à une règle. Dans ce cas, la remise en question de la règle est retardée. L'allure de la courbe représentant la valeur de l'indice en fonction du nombre de contre-exemples doit idéalement être concave (du moins dans le voisinage du maximum).

Choix d'une mesure de qualité adaptée au contexte

L'ordre sur les valeurs du critère g_6 (non-linéarité en fonction du nombre de contre-exemples) est donc concave \succ linéaire \succ convexe.

- **Sc2** : l'expert refuse l'apparition de trop de contre-exemples à une règle. Dans ce cas, la remise en question de la règle doit se faire plus rapidement. L'allure de la courbe est par conséquent idéalement convexe (dans le voisinage du maximum du moins) et l'ordre sur les valeurs du critère g_6 est convexe \succ linéaire \succ concave.

Nous analysons dans une première étape les rangements obtenus avec des poids égaux pour les critères. Les rangements totaux sont donnés dans le tableau 8.

Rang :	1	2	3	4	5	6	7
Sc1 :	INTIMP	IIE	LOE	MC	CONF CEN	CONV	-INDIMP, IPD
Sc2 :	MC	CONV	LOE	CONF CEN	INTIMP	-INDIMP, IPD	
Rang :	8	9	10	11	12	13	14
Sc1 :		ZHANG	PS	TEC	CONF	GI	R
Sc2 :	IIE	ZHANG	PS	R, LIFT		SEB	CONF
Rang :	15	16	17	18	19	20	
Sc1 :	LIFT	SURP	SEB	IQC	SUP	LAP	
Sc2 :	IQC	SURP	GI	TEC	LAP	SUP	

TAB. 8 – Rangements totaux pour les scénarios **Sc1** et **Sc2**.

La première observation intéressante à tirer du tableau 8 est que la différenciation entre les deux scénarios est importante. En effet, nous voyons que pour **Sc1**, les deux mesures de tête sont INTIMP et IIE, qui sont toutes les deux concaves. De même pour **Sc2**, les deux mesures de tête sont MC et CONV, qui sont toutes les deux convexes. Cette première constatation nous semble importante. En effet, malgré un système de poids neutre signifiant que l'expert n'exprime pas de préférences quant à l'importance des différentes caractéristiques de la mesure souhaitée, le choix de la (ou des meilleures) mesure(s) tient compte de l'allure de la courbe, et par conséquent de la sensibilité de l'expert face à l'apparition de contre-exemples. Dans cette même optique, nous observons qu'en augmentant relativement le poids de g_6 (non-linéarité) (33%) par rapport aux autres critères, les premières positions du rangement pour **Sc1** (respectivement **Sc2**) sont occupées par INTIMP, IIE, ZHANG, TEC, LOE et GI (respectivement par MC, CONV, SEB, et LOE) qui sont majoritairement concaves (respectivement convexes). Cette analyse supplémentaire montre également que LOE, qui est linéaire, est une mesure très intéressante, car elle reste bien rangée pour les deux scénarios. C'est un bon compromis.

Une analyse de stabilité sur ces deux systèmes de poids par le biais des intervalles de stabilité nous indique que de légères modifications des poids perturbent légèrement les rangements totaux (les perturbations sont locales). Ainsi un large éventail de systèmes de poids fournit des rangements similaires à ceux du tableau 8 pour les premières positions. Ceci est par ailleurs confirmé par la valeur des flots nets ϕ des 5 premiers éléments de chacun des rangements, comme résumé dans le tableau 9.

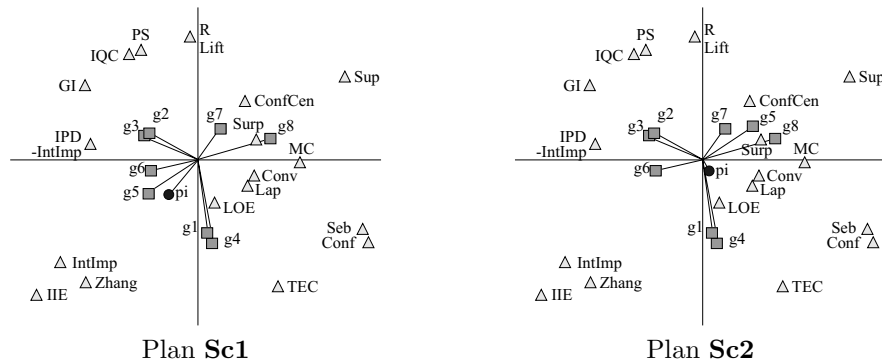
Ce tableau indique que les $\phi(a)$, $a \in \mathcal{M}$ sont répartis de manière assez homogène entre leur minimum et leur maximum pour les deux scénarios, en particulier pour les quelques premières mesures des rangements. Ceci signifie donc que des bouleversements majeurs sur ces premières positions n'ont lieu que pour des variations de poids très significatives. Nous en déduisons que pour un expert qui n'a pas de préférences

Sc1	INTIMP	IIE	LOE	MC	CONFCEM	CONV	...	LAP
ϕ	.32	.18	.18	.16	.13	.08	...	-.32
Sc2	MC	CONV	LOE	CONFCEM	INTIMP	-INDIMP	...	LAP
ϕ	.38	.30	.20	.15	.12	.10	...	-.30

TAB. 9 – Flots nets pour **Sc1** et **Sc2**.

particulières sur l'importance des différents critères, ou qui considère que les critères sont équi-importants, les rangements du tableau 8 sont un bon compromis.

Une analyse du plan GAIA nous fournit des informations supplémentaires. La figure 4 montre les plans GAIA pour **Sc1** et **Sc2**.

FIG. 4 – Plans GAIA pour **Sc1** et **Sc2**

Notons tout d'abord que le pourcentage de variance cumulé pour les deux facteurs représentés par le plan GAIA est de 58.8%. Les informations issues du plan GAIA doivent donc être considérées comme approximatives, et des conclusions trop hâtives risquent d'être erronées. Nous observons tout d'abord que les mesures, représentées par des triangles, sont réparties de manière homogène dans le plan. En outre, les critères, représentés par des axes reliés à des carrés, couvrent également bien toutes les régions du plan. Nous en déduisons que la caractérisation des mesures par les critères est discriminante, et peu redondante.

Pour **Sc1**, nous observons que plusieurs couples de critères sont indépendants : (g_4, g_5) , (g_4, g_6) , (g_1, g_5) , (g_1, g_6) , (g_4, g_8) et (g_1, g_8) . En outre, nous observons des groupes de critères qui expriment des préférences conflictuelles. Ainsi, l'ensemble $\{g_1, g_4\}$ et l'ensemble $\{g_2, g_3\}$ sont en conflit. Il en est de même, pour les deux ensembles de critères $\{g_5, g_6\}$ et $\{g_7, g_8\}$. Ceci donne donc une indication sur le comportement des mesures, ainsi que sur la structure générale du problème. Ainsi, par exemple, pour l'ensemble \mathcal{M} des mesures étudiées, une mesure qui est bonne pour le critère g_5 (concave) aura tendance à être mauvaise pour le critère g_8 (non intelligible).

Choix d'une mesure de qualité adaptée au contexte

Pour **Sc2**, des observations similaires peuvent être faites. Le changement majeur concerne le critère g_5 qui représente clairement des préférences similaires aux critères g_7 et g_8 , et est désormais en conflit avec g_6 .

En ce qui concerne l'axe de décision π , on observe que dans **Sc1** il est relativement long et pointe clairement dans une direction opposée aux critères g_7 et g_8 . Ceci veut dire que les mesures très bonnes en facilité de fixation du seuil et en intelligibilité (et plutôt mauvaises sur le reste des critères) ne se retrouveront en début de classement qu'à condition d'imposer un poids relatif très élevé aux critères g_7 et g_8 . Or nous estimons que l'importance d'un critère tel que g_3 (hypothèse d'indépendance) ne peut être négligeable vis-à-vis d'un critère comme g_8 . De plus, ces deux critères sont quasiment en conflit. Ainsi, si l'expert est conscient de l'impact de son choix de poids sur le résultat, on peut supposer de manière raisonnable qu'une mesure telle que SUP, exclusivement bonne sur g_7 et g_8 ne se retrouvera jamais en début de rangement.

Pour **Sc2** en revanche, l'axe de décision est court. Ceci signifie que le vecteur w est quasi perpendiculaire au plan GAIA dans l'espace de dimension k . Comme indiqué en section 7.5, les bons compromis se situent proches de l'origine. Ceci explique aussi le rangement du tableau 8.

Les positions des mesures dans le plan GAIA (pour **Sc1** et **Sc2**) montrent que plusieurs alternatives ont des comportements similaires face à des variations de poids. Ceci est par ailleurs confirmé par des profils similaires dans la matrice de décision. Ainsi, des mesures telles que PS et IQC, ou SEB et CONF, ou LIFT et R, ou -INDIMP et IPD sont proches dans le plan GAIA et ont en effet des profils similaires. Ces couples de mesures auront donc tendance à se retrouver à des positions proches dans les rangements totaux. Une remarque importante est cependant à faire concernant l'interprétation du plan GAIA. Etant donné qu'il ne représente qu'une partie de l'information contenue dans le nuage original de l'espace de dimension k , toute observation faite dans le plan GAIA devra être vérifiée sur les données ou par d'autres techniques. Ainsi, les mesures CONV et LAP sont proches dans le plan, mais leurs profils sont fort différents. Par conséquent, leur comportement vis-à-vis de la variation des poids des critères sera également différent.

Cette étude assez détaillée sert à démontrer l'utilité d'une analyse menée avec un outil d'aide à la décision tel que la méthode PROMETHEE. A partir des observations faites ci-avant, on peut donc envisager deux stratégies d'aide à la décision.

La première stratégie consiste à s'assurer tout d'abord que l'expert a bien compris le sens de chacun des critères et leur influence sur le résultat final. Ensuite, il s'agit de le questionner sur l'importance relative des poids de chacun des critères, tout en vérifiant la cohérence de ses dires. Ainsi, certains critères tels que g_3 , g_4 et g_7 devront nécessairement avoir des poids élevés, afin de garantir une certaine cohérence. En effet, une mesure n'ayant pas de valeur fixe pour l'indépendance et pour le règle logique, avec de surcroît un seuil difficile à fixer est fort peu utile pour une recherche efficace de règles. Suivant les préférences de l'expert, l'importance de critères tels que g_1 et g_8 pourra varier. De cette manière, un premier système de poids pourra être utilisé comme point départ de l'analyse. Le premier rangement proposé ne doit cependant pas être accepté d'office. Ainsi, une analyse des poids fournis par l'expert sera nécessaire afin de valider la stabilité des premières positions du rangement. Si une petite variation

du système de poids induit des perturbations importantes dans la tête du rangement, un nouveau questionnement de l’expert est peut-être nécessaire. Lorsqu’un rangement stable est obtenu, le plan GAIA, la valeur des flots nets et l’outil de comparaison des profils des mesures permettent d’analyser plus finement la tête du rangement. La valeur des flots nets permet d’avoir une idée de la “distance” séparant deux alternatives dans le rangement total. Ainsi, deux mesures proches d’après les flots peuvent d’une certaine manière aussi être considérées comme semblables, et le choix final pourra se faire entre l’une ou l’autre.

La seconde stratégie consiste dans une première étape en une exploration du plan GAIA. En effet, cette façon de procéder aide l’expert à comprendre la structure du problème et à repérer des indices qui sont similaires (voisins) ou complètement différents (éloignés). D’autre part, la représentation des critères dans le même plan que les alternatives permet de visualiser l’influence d’un changement de poids sur le rangement final. Cette façon de procéder s’applique plutôt à l’expert qui a une connaissance *a priori* sur certaines mesures. En repérant ces mesures dans le plan GAIA, il pourra plus aisément déterminer leurs caractéristiques. De cette façon, il pourra déterminer ses préférences concernant l’importance des différents critères, en s’inspirant des mesures qu’il connaît. En partant de ce système de poids, la première stratégie peut être appliquée. Une vérification *a posteriori* peut être intéressante afin de déterminer les positions des mesures connues dans le rangement final.

Afin de montrer l’utilité et l’applicabilité de cette méthode à ce cas particulier de sélection d’une mesure, nous terminons cette section par une simulation d’un comportement expert \mathcal{E} . Nous supposons que \mathcal{E} désire une mesure facile à utiliser. Il souhaite donc qu’elle soit lisible (et donc facilement compréhensible) et que les bornes soient fixes. Le système de poids qu’il suggère est le suivant : g_1 (10%), g_2 (5%), g_3 (15%), g_4 (15%), g_5 (10%), g_6 (5%), g_7 (15%) et g_8 (25%).

Les premières positions du rangement total obtenu sont données dans le tableau 10.

1	2	4	5	...
MC (.30)	CONF CEN, LOE (.22)	CONF (.17)	INTIMP (.16)	...

TAB. 10 – Rangement et flot net pour les préférences de \mathcal{E} .

On y voit clairement que MC est de loin la meilleure mesure pour ce système de poids. Il s’agit d’une mesure facilement interprétable, ce qui est en accord avec l’un des objectifs principaux de \mathcal{E} . Une analyse de la variation de ce rangement par rapport à de légères modifications des poids indique une stabilité au sein des premières positions de ce rangement. De plus, la mesure MC reste constamment la première avec un flot net significativement plus élevé que le second indice. D’autre part, la mesure MC vérifie bien les préférences de \mathcal{E} . En effet, cette mesure est performante pour les critères g_3 , g_4 , g_7 et g_8 qui sont les plus importants d’après les préférences de \mathcal{E} . En outre, MC est aussi performante pour les critères g_1 et g_2 . Ses faiblesses sont dans l’allure convexe de la courbe, ainsi que dans son insensibilité vis-à-vis de la dilatation des données. Or justement ces deux critères sont considérés comme peu importants par \mathcal{E} .

9 Conclusion

Dans cet article nous abordons le problème de la sélection des règles issues d'un processus ECD du point de vue de la sélection des mesures de qualité adaptées au contexte. Ce problème relève clairement d'une approche d'aide multicritère à la décision afin de faire des recommandations répondant aux objectifs ainsi qu'aux préférences de l'utilisateur.

Nous avons étudié 20 mesures de qualité couramment utilisées pour l'évaluation des règles d'association en fonction de 8 propriétés pertinentes d'un point de vue utilisateur. Nous avons montré comment utiliser une méthode d'aide multicritère à la décision pour aider l'expert dans son choix de la mesure de qualité la mieux adaptée à ses propres besoins.

En particulier, l'utilisation de PROMETHEE a confirmé le fait que les préférences des experts ont un impact direct sur l'ordre de sélection des mesures, et que certaines mesures ont des comportements similaires. Cette observation est également faite expérimentalement sur des données avec le simulateur HERBS. En outre, la méthode PROMETHEE nous a permis d'analyser l'impact des préférences de l'utilisateur en termes de poids des critères (le plan GAIA permet de visualiser facilement des classes de critères).

Notre ensemble de critères couvre une partie importante des préférences couramment exprimées par les utilisateurs. Il n'est cependant pas exhaustif et peut être étendu en fonction des besoins. De nouveaux critères sont susceptibles de nous permettre de distinguer des mesures qui sont actuellement indissociables sur la base des préférences exprimées. Une propriété importante, *qualité algorithmique de la mesure*, n'est par exemple pas prise en compte dans notre étude car seul le support possède cette propriété. Ainsi, bien que très mal classé, SUP reste incontournable pour des algorithmes de la famille APRIORI. Il doit cependant rester cantonné à un rôle de premier filtre.

Nous pensons également que des critères provenant de campagnes d'expérimentation et plus difficiles à évaluer formellement doivent avoir leur place dans la grille d'évaluation, permettant ainsi de déboucher sur une cartographie fiable et stable des mesures de qualité. C'est le cas du pouvoir discriminant d'une mesure ou bien de la résistance au bruit.

Remerciements

Benoît Vaillant remercie la Communauté Urbaine de Brest pour le soutien financier de son travail de doctorat.

Patrick Meyer remercie la Région Wallonne pour le soutien financier de ses recherches.

Les auteurs tiennent à remercier les membres du groupe GAFOQUALITÉ pour les discussions fructueuses à propos des propriétés d'évaluation des mesures ainsi que les rapporteurs pour leurs remarques avisées.

Références

- [Agrawal *et al.*, 1993] R. Agrawal, T. Imielinski, et A.N. Swami. Mining association rules between sets of items in large databases. In P. Buneman et S. Jajodia, editors, *Proceedings of the 1993 ACM SIGMOD Int. Conference on Management of Data*, pages 207–216, Washington, D.C., 1993.
- [Azé et Kodratoff, 2002] J. Azé et Y. Kodratoff. Evaluation de la résistance au bruit de quelques mesures d'extraction de règles d'association. *Extraction des connaissances et apprentissage (EGC 2002)*, 1(4) :143–154, 2002.
- [Bertin, 1977] J. Bertin. *La graphique et le traitement graphique de l'information*. Flammarion, 1977.
- [Borgelt et Kruse, 2002] C. Borgelt et R. Kruse. Induction of association rules : APRIORI implementation. In *Proceedings of the 15th Conference on Computational Statistics*, Heidelberg, Germany, 2002. Physika Verlag.
- [Brans et Mareschal, 1994] J.P. Brans et B. Mareschal. The PROMETHEE-GAIA decision support system for multicriteria investigations. *Investigation Operativa*, 4(2) :102–117, 1994.
- [Brans et Mareschal, 2002] J.P. Brans et B. Mareschal. PROMETHEE-GAIA – Une méthode d'aide à la décision en présence de critères multiples. Ellipses, 2002.
- [Brin *et al.*, 1997a] S. Brin, R. Motwani, J. D. Ullman, et S. Tsur. Dynamic itemset counting and implication rules for market basket data. In Joan Peckham, editor, *SIGMOD 1997, Proceedings ACM SIGMOD Int. Conference on Management of Data, Tucson, Arizona, USA*, pages 255–264. ACM Press, 05 1997.
- [Brin *et al.*, 1997b] Sergey Brin, Rajeev Motwani, et Craig Silverstein. Beyond market baskets : generalizing association rules to correlations. In *ACM SIGMOD/PODS '97 Joint Conference*, pages 265–276, 1997.
- [Chauchat et Risson, 1998] J.-H. Chauchat et A. Risson. *Visualization of Categorical Data*, chapter 3, pages 37–45. Blasius J. & Greenacre M. ed., 1998. New York : Academic Press.
- [Church et Hanks, 1990] K. W. Church et P. Hanks. Word association norms, mutual information an lexicography. *Computational Linguistics*, 16(1) :22–29, march 1990.
- [Cohen, 1960] J. Cohen. A coefficient of agreement for nominal scale. *Educational and Psychological Measurement*, 20 :37–46, 1960.
- [Fayyad *et al.*, 1996] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, et R. Uthurusamy, editors. *Advances in Knowledge Discovery and Data Mining*. AAAI/MIT Press, 1996.
- [Francisci *et al.*, 2003] D. Francisci, L. Brisson, et M. Collard. Extraction de règles selon des critères multiples : l'art du compromis. Technical Report ISRN I3S/RR-2003-11-FR, Université de Nice, Mai 2003.
- [Freitas, 1999] A. Freitas. On rule interestingness measures. *Knowledge-Based Systems journal*, pages 309–315, 1999.

- [Giakoumakis et Monjardet, 1987] V. Giakoumakis et B. Monjardet. Coefficients d'accord entre deux préordres totaux. *Statistique et Analyse des Données*, 12(1 et 2) :46–99, 1987.
- [Good, 1965] I. J. Good. The estimation of probabilities : An essay on modern bayesian methods. The MIT Press, Cambridge, MA, 1965.
- [Goodman et Smyth, 1988] R. M. Goodman et P. Smyth. Information-theoretic rule induction. In *Proceedings of the 1988 European Conference on Artificial Intelligence*. Pitman, 1988.
- [Gras *et al.*, 1996] R. Gras, S. Ag. Almouloud, M. Bailleuil, A. Larher, M. Polo, H. Ratsimba-Rajohn, et A. Totohasina. *L'implication Statistique, Nouvelle Méthode Exploratoire de Données. Application à la Didactique, Travaux et Thèses*. La Pensée Sauvage, 1996.
- [Gras *et al.*, 2001] R. Gras, P. Kuntz, R. Couturier, et F. Guillet. Une version entropique de l'intensité d'implication pour les corpus volumineux. *Extraction des connaissances et apprentissage (EGC 2001)*, 1(1-2) :69–80, 2001.
- [Gras *et al.*, 2004] R. Gras, R. Couturier, M. Bernadet, J. Blanchard, H. Briand, F. Guillet, P. Kuntz, R. Lehn, et P. Peter. Quelques critères pour une mesure de qualité de règles d'association - un exemple : l'intensité d'implication. *Revue des Nouvelles Technologies de l'Information*, 2004. Ce numéro.
- [Gras, 1979] R. Gras. *Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques*. PhD thesis, Université de Rennes I, 1979.
- [Hilderman et Hamilton, 2001] J. R. Hilderman et H. J. Hamilton. Evaluation of interestingness measures for ranking discovered knowledge. *Lecture Notes in Computer Science*, 2035 :247–259, 2001.
- [Lallich et Teytaud, 2004] S. Lallich et O. Teytaud. Évaluation et validation de l'intérêt des règles d'association. *Revue des Nouvelles Technologies de l'Information*, 2004. Ce numéro.
- [Lenca *et al.*, 2003a] P. Lenca, P. Meyer, P. Picouet, B. Vaillant, et S. Lallich. Critères d'évaluation des mesures de qualité en ECD. *Revue des Nouvelles Technologies de l'Information (Entreposage et Fouille de données)*, (1) :123–134, 2003.
- [Lenca *et al.*, 2003b] P. Lenca, P. Meyer, B. Vaillant, et P. Picouet. Aide multicritère à la décision pour évaluer les indices de qualité des connaissances – modélisation des préférences de l'utilisateur. *RSTI-RIA (EGC 2003)*, 1(17) :271–282, 2003.
- [Lerman *et al.*, 1981] I.C. Lerman, R. Gras, et H. Rostam. Elaboration d'un indice d'implication pour les données binaires, i et ii. *Mathématiques et Sciences Humaines*, (74, 75) :5–35, 5–47, 1981.
- [Lerman et Azé, 2003] I.C. Lerman et J. Azé. Une mesure probabiliste contextuelle discriminante de qualité des règles d'association. *RSTI-RIA (EGC 2003)*, 1(17) :247–262, 2003.
- [Loevinger, 1947] J. Loevinger. A systemic approach to the construction and evaluation of tests of ability. *Psychological monographs*, 61(4), 1947.

- [Pasquier *et al.*, 1999] N. Pasquier, Y. Bastide, R. Taouil, et L. Lakhal. Discovering frequent closed itemsets for association rules. In Catriel Beeri et Peter Buneman, editors, *Database Theory - ICDT '99, 7th Int. Conference, Jerusalem, Israel, January 10-12, 1999, Proceedings*, volume 1540 of *Lecture Notes in Computer Science*, pages 398–416. Springer, 1999.
- [Pearl, 1988] J. Pearl. *Probabilistic reasoning in intelligent systems*. Morgan Kaufmann, 1988.
- [Pearson, 1896] K. Pearson. Mathematical contributions to the theory of evolution. iii. regression, heredity and panmixia. *Philosophical Transactions of the Royal Society, A*, 1896.
- [Piatetsky-Shapiro, 1991] G. Piatetsky-Shapiro. Discovery, analysis and presentation of strong rules. In G. Piatetsky-Shapiro et W.J. Frawley, editors, *Knowledge Discovery in Databases*, pages 229–248. AAAI/MIT Press, 1991.
- [Picouet et Lenca, 2001] P. Picouet et P. Lenca. *Bases de données et internet*, chapter Extraction de connaissances à partir des données, pages 395–420. Hermes Science, 2001.
- [Roy, 1985] B. Roy. *Méthodologie multicritère d'aide à la décision*. Ed. Economica, collection Gestion, 1985.
- [Sebag et Schoenauer, 1988] M. Sebag et M. Schoenauer. Generation of rules with certainty and confidence factors from incomplete and incoherent learning bases. In J. Boose, B. Gaines, et M. Linster, editors, *Proc. of the European Knowledge Acquisition Workshop (EKAW'88)*, pages 28–1 – 28–20. Gesellschaft für Mathematik und Datenverarbeitung mbH, 1988.
- [Silberschatz et Tuzhilin, 1995] A. Silberschatz et A. Tuzhilin. On subjective measures of interestingness in knowledge discovery. In *Knowledge Discovery and Data Mining*, pages 275–281, 1995.
- [Silberschatz et Tuzhilin, 1996] A. Silberschatz et A. Tuzhilin. What makes patterns interesting in knowledge discovery systems. *IEEE Transactions On Knowledge And Data Engineering*, 8(6) :970–974, 1996.
- [Tan *et al.*, 2002] P.-N. Tan, V. Kumar, et J. Srivastava. Selecting the right interestingness measure for association patterns. In *Proceedings of the Eighth ACM SIGKDD Int. Conference on Knowledge Discovery and Data Mining*, pages 32–41, 2002.
- [Tan et Kumar, 2000] P. Tan et V. Kumar. Interestingness measures for association patterns : A perspective. Technical Report TR00-036, University of Minnesota, Department of Computer Science, 2000.
- [Terano *et al.*, 2000] T. Terano, H. Liu, et A. L. P. Chen, editors. *Association Rules*, volume 1805 of *Lecture Notes in Computer Science*. Springer, April 2000.
- [Vaillant *et al.*, 2003] B. Vaillant, P. Picouet, et P. Lenca. An extensible platform for rule quality measure benchmarking. In R. Bisdorff, editor, *Human Centered Processes (HCP'2003)*, pages 187–191, Luxembourg, Mai 2003.
- [Vaillant, 2002] B. Vaillant. Evaluation de connaissances : le problème du choix d'une mesure de qualité en extraction de connaissances à partir des données. Master's thesis, Ecole Nationale Supérieure des Télécommunications de Bretagne, 2002.

Summary

The huge number of rules resulting from procedures of knowledge discovery in databases (KDD) does not allow expert users to easily select the interesting or relevant ones by hand. A major concern of KDD is the development of quality measures evaluating the interestingness of the discovered rules. Many measures have been proposed and the expert is supposed to select the appropriate one, according to his needs. Unfortunately, due to the large number of suggested measures, the selection task is itself a hard problem. To help solving this problem, many articles are presenting partial panoramas of these measures. Some are limiting themselves to their algorithmic performances, others are attempting to formalise their quality, and so on. The results are not convincing at all. Next to the large number of measures, now the expert is also confronted to the variety of their characteristics which don't always reflect his objectives. Finally, the heterogeneousness of these characteristics also adds to the selection problem of the expert when it comes to comparing the different measures. In this context, the choice by an expert of adequate measures remains a major challenge in KDD. The search for the best rules is directly linked to the selection and the use of adapted measures. We are therefore confronted to a multiple criteria decision aiding problematic. We address this problem by evaluating 20 measures on 8 criteria.

Keywords : interestingness measures, association rules, decision aid.