Mesurer l'intérêt des règles d'association

Benoît Vaillant, Patrick Meyer, Elie Prudhomme, Stéphane Lallich, Philippe Lenca, Sébastien Bigaret

Mesurer l'intérêt des règles d'association

Benoît Vaillant*, Patrick Meyer***, Elie Prudhomme**,
Stéphane Lallich**, Philippe Lenca*, Sébastien Bigaret*

*GET ENST Bretagne / Département LUSSI – CNRS UMR 2872
Technopôle de Brest Iroise - CS 83818, 29238 Brest Cedex, France
{*prenom.nom*}@enst-bretagne.fr
**Laboratoire ERIC - Université Lumière - Lyon 2
5 avenue Pierre Mendès-France, 69676 Bron Cedex, France
lallich@univ-lyon2.fr
***Service de Mathématiques Appliquées, Faculté de Droit,
d'Economie et de Finance, Université du Luxembourg,
162a, avenue de la Faïencerie, L-1511 Luxembourg
patrick.meyer@uni.lu

Résumé. A l'occasion de l'action spécifique GaFoDonnées (2002), le laboratoire LUSSI, ENST Bretagne et le Laboratoire ERIC, Université Lyon 2, ont engagé une collaboration sur le thème de l'intérêt des règles d'association. Cet article présente les travaux ainsi réalisés. Une vingtaine de mesures ont été retenues, sur la base d'un critère d'éligibilité. Différentes propriétés sont d'abord proposées qui fondent une étude formelle des mesures. Cette étude formelle se double d'une étude de comportement, grâce à Herbs, une plate-forme développée pour expérimenter les mesures sur des bases de règles. Il est alors possible de confronter la typologie formelle des règles et la typologie expérimentale associée à leur comportement sur différentes bases. Une fois transformées en critères, ces propriétés fondent une méthode d'assistance au choix de l'utilisateur. Le problème de la validation est enfin abordé, où l'on présente une méthode de contrôle du risque multiple adaptée au problème.

1 Introduction

Nous nous intéressons aux mesures relatives à l'intérêt des règles d'association $\mathtt{A} \rightarrow \mathtt{B}$ telles que définies dans (Agrawal *et al.*, 1993) : dans une base de données transactionnelles, $\mathtt{A} \rightarrow \mathtt{B}$ signifie que si les articles qui constituent $\mathtt{A}$ sont dans *le panier d'une ménagère*, alors le plus souvent les articles qui constituent $\mathtt{B}$ le sont aussi. Les algorithmes de type Apriori (fondé sur le support et la confiance) ont tendance à produire un grand nombre de règles pas toujours intéressantes du point de vue de l'utilisateur. Les mesures d'intérêt jouent alors un rôle essentiel en permettant de pré-filtrer les règles extraites. Après nous être intéressés séparément à ce problème (Teytaud et Lallich, 2001), (Vaillant, 2002), le groupe de travail GafoQualité [1] nous a donné l'occasion de développer en commun nos recherches. On trouvera

[1] Groupe de travail sur les Mesures de Qualité, animé par Fabrice Guillet, de l'Action Spécifique STIC Fouille de Bases de Données (GaFoDonnées), animée par Rosine Cicchetti et Michèle Sebag

dans (Briand *et al.*, 2004) différents articles issus des travaux menés dans GafoQualité (qualité des données, des règles d'association, des arbres de décision, etc.). Cet article présente une synthèse de nos travaux sur la qualité des règles d'association.

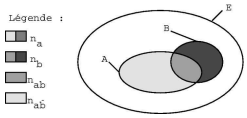
Nous nous plaçons en phase de *post-analyse*. Ainsi n'abordons nous ni les problèmes liés à la qualité des données, étudiés notamment par (Berti-Equille, 2004), ni ceux posés par l'extraction des règles (Pasquier, 2000). Les données et les règles issues du processus d'extraction sont des *entrées*.

Différentes voies ont été explorées. Ainsi, nous définissons des mesures et proposons des propriétés souhaitables section 2. La section 3 concerne le développement de la plateforme expérimentale Herbs. La section 4 est relative au développement d'une aide à la sélection de bonnes mesures. Les deux typologies des mesures, l'une fondée sur une approche expérimentale, l'autre sur une approche formelle sont mises en regard section 5. Enfin, la section 6 s'intéresse à la validation des règles.

2 Mesures et propriétés

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

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



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

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.

Notre premier travail a été de recenser les mesures de l'intérêt des règles d'association et de mettre en évidence leurs propriétés, tant dans une perspective d'analyse formelle (Teytaud et Lallich, 2001), (Lallich, 2002), que d'expérimentation (Vaillant, 2002). Ces propriétés ont ensuite été écrites de façon opérationnelle pour servir de base à nos travaux d'aide à la décision (Lenca *et al.*, 2002, Lenca *et al.*, 2004).

Les règles d'association $\mathtt{A} \rightarrow \mathtt{B}$ se focalisent sur les coprésences en traitant les itemsets de façon non symétrique, une mesure doit impérativement distinguer $A \rightarrow B$ de $A \rightarrow \overline{B}$ (Lallich et Teytaud, 2004) et une règle $\mathtt{A} \rightarrow \mathtt{B}$ doit être distinguée de l'implication $A \Rightarrow B$ et de l'équivalence $A \Leftrightarrow B$. Ainsi qu'à la différence de (Tan *et al.*, 2002), nous nous sommes limités aux mesures qui sont décroissantes avec $n_{a\overline{b}}$ (resp. croissantes avec n_{ab}), les effectifs marginaux étant fixés, excluant d'emblée les mesures comme le χ^2, le r^2 de Pearson ou la mesure de Pearl.

Nous avons retenu 20 mesures, listées tableau 1. A côté du support p_{ab} et de la confiance $p_{b|a}$, une première catégorie de mesures rassemble des transformées affines de

la confiance qui ont pour but de la comparer à p_b. Cette comparaison se fait le plus souvent en centrant la confiance sur p_b avec différents coefficients d'échelle (confiance centrée, coefficient de corrélation, indice d'implication, mesures de Piatetsky-Shapiro, Loevinger, Zhang). Elle peut aussi se faire en divisant la confiance par p_b (lift ou taux de liaison). 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. Certaines mesures privilégient les contre-exemples, ainsi la conviction $\frac{p_{\overline{b}|a}}{p_{\overline{b}}}$ et l'indice d'implication. Ce dernier est à la base de différents indices probabilistes comme l'intensité d'implication, l'intensité d'implication entropique et l'indice probabiliste discriminant. En outre, nous avons analysé les classes d'équivalences issues de la relation d'équivalence "classer comme" définie sur les couples de mesure, bon nombre de mesures étant des transformées monotones croissantes les unes des autres (*e.g* la mesure de Sebag et la confiance, ou la mesure de Loevinger et la conviction).

Pour analyser formellement ces mesures, puis les évaluer dans une perspective d'aide à la décision, nous proposons 8 propriétés (en gras, ci-dessous). 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 $B \rightarrow A$ (**dissymétrie**). Pour une même proportion d'exemples p_{ab}, une règle est d'autant plus intéressante que p_b est faible (**décroissance avec** p_b). Les comparaisons sont plus faciles lorsque les mesures ont une **valeur fixe en cas de règle logique**, ainsi qu'en cas d'**indépendance.** Certains auteurs, tel (Gras *et al.*, 2004), privilégient des mesures concaves lors de l'apparition des premiers contre-exemples, d'autres peuvent préférer une décroissance convexe plus brutale, ou simplement linéaire (**courbure à l'origine**). La mesure doit-elle prendre en compte le nombre total de transactions n (mesure statistique) ou non (descriptive) ? Les règles statistiques intuitivement plus fondées, ont l'inconvénient de perdre leur pouvoir discriminant dès que n est grand (**prise en compte de** n). 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 (**fixation du seuil**). On peut 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, mais comme un paramètre de contrôle, sauf à contrôler effectivement le risque multiple avec le critère UAFWER (voir section 6).

3 HERBS : une plate-forme d'expérimentation

Dans le cadre de la recherche de règles d'association, il est classique d'effectuer une opération de filtrage au moyen de mesures de qualité. Des logiciels permettant l'extraction de telles règles proposent ainsi d'utiliser le lift (IBM, 1996), en plus du support et de la confiance. D'autres proposent le coefficient de corrélation linéaire. Plus récemment, l'outil FELIX (Lehn, 2000) intègre l'intensité d'implication et sa version entropique. Mais à notre connaissance, les outils disponibles ne proposent qu'un sous-ensemble réduit de mesures, et qui plus est leur intégration est à but fonctionnel et non afin d'en étudier les comportements.

Nous avons développé HERBS (Vaillant, 2002), qui intègre 20 mesures de qualité

Mesure	Abréviation et référence	Définition
support	SUP (Agrawal *et al.*, 1993)	$\frac{n_a - n_{a\bar{b}}}{n}$
confiance	CONF (Agrawal *et al.*, 1993)	$1 - \frac{n_{a\bar{b}}}{n_a}$
coefficient de corrélation linéaire	R (Pearson, 1896)	$\frac{nn_{ab} - n_a n_b}{\sqrt{nn_a n_b n_{\bar{a}} . n_{\bar{b}}}}$
confiance centrée	CONFCEN	$\frac{nn_{ab} - n_a n_b}{nn_a}$
conviction	CONV (Brin *et al.*, 1997b)	$\frac{n_a n_{\bar{b}}}{nn_{a\bar{b}}}$
Piatetsky-Shapiro	PS (Piatetsky-Shapiro, 1991)	$\frac{1}{n}(\frac{n_a n_{\bar{b}}}{n} - n_{a\bar{b}})$
Loevinger	LOE (Loevinger, 1947)	$1 - \frac{nn_{a\bar{b}}}{n_a n_{\bar{b}}}$
gain informationnel	GI (Church et Hanks, 1990)	$\log(\frac{nn_{ab}}{n_a n_b})$
Sebag-Schoenauer	SEB (Sebag et Schoenauer, 1988)	$\frac{n_a - n_{a\bar{b}}}{n_{ab}}$
lift	LIFT (Brin *et al.*, 1997a)	$\frac{nn_{ab}}{n_a n_b}$
Laplace	LAP (Good, 1965)	$\frac{n_{ab}+1}{n_a+2}$
moindre contradiction	MOCO (Azé et Kodratoff, 2002)	$\frac{n_{ab} - n_{a\bar{b}}}{n_b}$
multiplicateur de cotes	MC (Lallich et Teytaud, 2004)	$\frac{(n_a - n_{a\bar{b}})n_{\bar{b}}}{n_b n_{a\bar{b}}}$
taux d'exemples et de contre-exemples	TEC	$\frac{n_a - 2n_{a\bar{b}}}{n_a - n_{a\bar{b}}}$
indice de qualité de Cohen	IQC (Cohen, 1960)	$2\frac{nn_a - nn_{a\bar{b}} - n_a n_b}{nn_a + nn_b - 2n_a n_b}$
Zhang	ZHANG (Terano *et al.*, 2000)	$\frac{nn_{ab} - n_a n_b}{\max\{n_{ab}n_{\bar{b}}, n_b n_{a\bar{b}}\}}$
indice d'implication	-INDIMP (Lerman *et al.*, 1981)	$\frac{nn_{a\bar{b}} - n_a n_{\bar{b}}}{\sqrt{nn_a n_{\bar{b}}}}$
intensité d'implication	INTIMP (Gras *et al.*, 1996)	$P\left[poisson\left(\frac{n_a n_{\bar{b}}}{n}\right) \geq n_{a\bar{b}}\right]$
intensité d'implication entropique	IIE (Gras *et al.*, 2001)	$\left\{[(1 - h_1(\frac{n_{a\bar{b}}}{n})^2) \times (1 - h_2(\frac{n_{a\bar{b}}}{n})^2)]^{1/4}{}_{\text{INTIMP}}\right\}^{1/2}$
indice probabiliste discriminant	IPD (Lerman et Azé, 2003)	$P\left[\mathcal{N}(0,1) >_{\text{INDIMP}}{}^{CR/\mathcal{B}}\right]$

- $h_1(t) = -(1 - \frac{n \cdot t}{n_a}) \log_2(1 - \frac{n \cdot t}{n_a}) - \frac{n \cdot t}{n_a} \log_2(\frac{n \cdot t}{n_a})$ si $t \in [0, n_a/2\ n[$; $h_1(t) = 1$ sinon
- $h_2(t) = -(1 - \frac{n \cdot t}{n_{\bar{b}}}) \log_2(1 - \frac{n \cdot t}{n_{\bar{b}}}) - \frac{n \cdot t}{n_{\bar{b}}} \log_2(\frac{n \cdot t}{n_{\bar{b}}})$ si $t \in [0, n_{\bar{b}}/2\ n[$; $h_2(t) = 1$ sinon
- *poisson* correspond à la loi de distribution de Poisson
- $\mathcal{N}(0,1)$ correspond à la fonction de distribution de la loi normale centrée réduite
- INDIMP $^{CR/\mathcal{B}}$ correspond à INDIMP, centré réduit (CR) pour une base de règle $\mathcal{B}$

TAB. 1 – Mesures étudiées

dont nous étudions par ailleurs les propriétés formelles. Il est possible d'importer des règles aux formats de sortie C4.5 (`http://www.rulequest.com/Personal`) et APRIORI (`http://fuzzy.cs.uni-magdeburg.de/~borgelt/doc/apriori`). Afin d'évaluer ces règles, il est possible d'importer des données au format csv. D'autres formats d'échanges sont envisagés, par exemple avec WEKA (`http://www.cs.waikato.ac.nz/~ml/weka`) et TANAGRA (`http://chirouble.univ-lyon2.fr/~ricco/tanagra`).

A partir de couples de bases de cas et de règles compatibles (*i.e.* portant sur les mêmes attributs), HERBS permet d'effectuer divers types d'analyse :

Etude d'une mesure Plusieurs traitements nous semblent intéressants afin de caractériser le comportement d'une mesure donnée :

- Evaluation des objets que la mesure va avoir à traiter au moyen de quelques grandeurs : nombre de cas et de règles, taux de couverture, indice de recouvrement, et nombre de règles "particulières" (logiques, sans exemples, et ne passant pas l'hypothèse d'indépendance).

– Sélection de l'ensemble des N meilleures règles selon la mesure donnée.
– Tracé de la distribution des valeurs prises par la mesure.

Comparaison de mesures Afin de comparer le comportement expérimental de mesures, trois voies ont été développées :
– L'extraction de l'ensemble des règles classées k fois parmi les N meilleures par p mesures.
– La comparaison des préordres induits par deux mesures.
– Le tracé des distributions croisées des valeurs de deux mesures.

Les résultats présentés dans la figure 2 sont extraits de (Vaillant *et al.*, 2004). Plusieurs jeux de données disponibles depuis le site de l'UCI (`ftp.ics.uci.edu/`) ont été utilisés afin de générer des règles. Pour une base de règles donnée, chaque mesure induit un préordre sur l'ensemble de règles. Afin d'étudier les similarités entre mesures, nous avons calculé un coefficient d'accord entre préordres (*cf.* figure 2 et (Lenca *et al.*, 2004) pour les détails de calcul).

Après une transformation linéaire de la valeur du coefficient d'accord afin d'obtenir des valeurs entre 0 et 1, et un réarrangement de l'ordre des lignes et des colonnes afin de mieux mettre en évidence les structures de blocs, on obtient une classification expérimentale des mesures de qualité, à mettre en regard avec la classification obtenue à partir de propriétés formelles.

4 Assistance au choix des mesures

Les mesures d'intérêt possèdent des propriétés diverses (Lallich et Teytaud, 2002) et l'ensemble des n meilleures règles résultant d'un préfiltrage d'une base de $m > n$ règles peut varier grandement selon la mesure utilisée (Vaillant, 2002). Ainsi, lorsque l'utilisateur est confronté à la sélection du sous-ensemble des n meilleures règles il est aussi confronté au choix des mesures d'intérêt à appliquer : choisir les *bonnes* règles c'est aussi choisir les *bonnes* mesures (Lenca *et al.*, 2002), (Lenca *et al.*, 2003b).

Ce choix doit être guidé par les préférences et les objectifs du principal intéressé, l'utilisateur expert des données. L'utilisateur est au cœur du processus et les travaux l'impliquant fortement sont à notre avis fort prometteurs, par exemple (Poulet, 1999), (Lehn *et al.*, 1999) et (Blanchard *et al.*, 2004). Partant des huit propriétés présentées section 2 nous avons identifié celles reposant sur les préférences de l'utilisateur et celles plus normatives afin de définir des critères de décision sur les mesures. Différentes méthodes d'*aide multi-critères à la décision* ont été appliquées (Lenca *et al.*, 2003b), (Lenca *et al.*, 2003a) afin d'obtenir, selon la méthode, un sous-ensemble de *bonnes* mesures ou un classement des mesures. Dans (Lenca *et al.*, 2004) nous précisons six éléments définissant le contexte et à prendre en compte pour l'assistance au choix des mesures : l'ensemble de données, l'ensemble de règles, l'ensemble de mesures, l'ensemble de propriétés des mesures, l'ensemble de préférences de l'utilisateur, l'ensemble de critères de décision. Nous y présentons une étude détaillée de deux scénarios utilisateur (tolérance –**Sc1**– ou non –**Sc2**– de l'apparition de contre-exemples dans les règles) et des classements des mesures selon ces scénarios. Le tableau 2 donne le classement des mesures pour ces deux scénarios avec des poids égaux pour les critères, obtenus avec la méthode PROMETHEE (Brans et Mareschal, 1994).

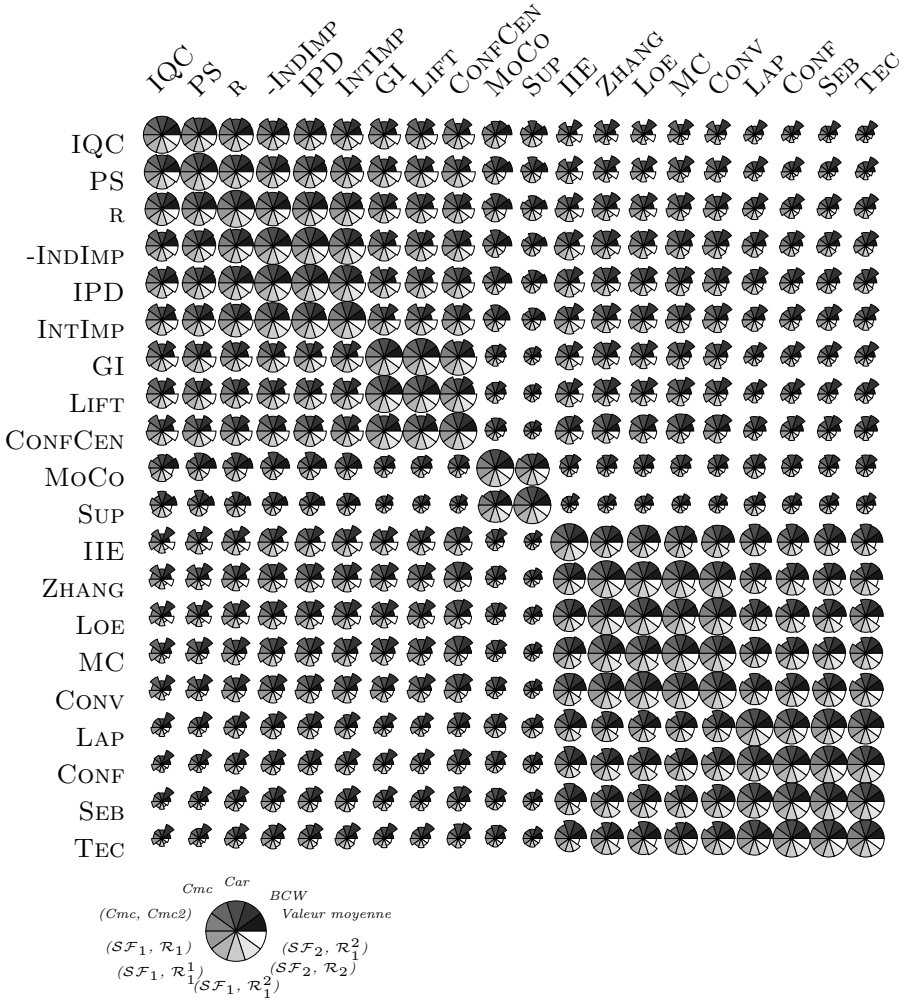


Fig. 2 – Comparaison de préordres

5 Typologies formelle et expérimentale des mesures

Nous avons extrait un sous-ensemble pertinent de 6 propriétés formelles, détaillées dans (Lenca *et al.*, 2003a). Ce sous-ensemble de propriétés nous a permis de construire une matrice de décision évaluant les mesures. A partir de ces évaluations, on peut ainsi construire une matrice de distance entre les mesures. En appliquant une classification ascendante hiérarchique avec le critère de Ward on distingue quatre classes principales : {PS, IQC, GI, ConfCen, Lift, R, -IndImp, IPD}, {IntImp, IIE, Loe, Zhang, MC, Conv}, {Conf, Seb, Tec}, et {Lap, Sup, MoCo}.

Le tableau 3 compare les deux approches. Du point de vue expérimental, seule la classe 3 présente de forts désaccords avec les résultats formels.

Rang :	1	2	3	4	5	6	7
Sc1 :	IntImp	Loe	MC	ConfCen	Conv	-IndImp,IPD	
Sc2 :	MC	Conv	Loe	ConfCen	IntImp	-IndImp, IPD	
Rang :	8	9	10	11	12	13	14
Sc1 :	IIE,Zhang		PS	Tec	Conf	GI	r, Lift
Sc2 :	PS	Seb	Conf	r, Lift		MoCo	IIE
Rang :	15	16	17	18	19	20	
Sc1 :		MoCo	Seb	IQC	Sup	Lap	
Sc2 :	Zhang	IQC	Tec	Sup	GI	Lap	

Tab. 2 – Rangements totaux pour les scénarios **Sc1** et **Sc2**.

Formelle \ Expérimentale	Classe 1	Classe 2	Classe 3	Classe 4
Classe 1	PS, IQC, GI ConfCen, Lift, r -IndImp, IPD			
Classe 2	IntImp	IIE	Loe, Zhang, MC, Conv	
Classe 3			Conf, Seb, Tec	
Classe 4			Lap	MoCo, Sup

Tab. 3 – Comparaison des classes entre l'approche formelle et expérimentale

6 Validation de règles

Le plus souvent, les *transactions* à partir desquelles sont extraites les règles d'association ne sont qu'un échantillon d'une population plus vaste. Au terme de la procédure d'extraction et d'évaluation des règles, on dispose d'une multitude de règles décrites par différentes mesures, au minimum le support et la confiance, ainsi qu'une mesure de l'intérêt de la règle. Différentes questions se posent classiquement : ces mesures, en particulier le support et la confiance, dépassent-elles significativement pour toutes les règles le seuil requis ? Telle mesure dépasse-t-elle significativement une valeur fixée ? La confiance $p_{b/a}$ d'une règle $A \rightarrow B$ est-elle significativement supérieure à sa fréquence *a priori* ? Dans ce dernier cas, il s'agit de tester l'hypothèse d'indépendance (H_0) du conséquent B et de l'antécédent A contre une hypothèse de dépendance positive (H_1), afin de ne retenir que les règles statistiquement significatives. On doit donc pratiquer une multitude de tests, ce qui pose le problème du contrôle du risque multiple. Par exemple, si l'on effectue le test d'indépendance de A et B pour 10000 règles successivement, en fixant à 0.05 le niveau du risque de 1^{re} espèce $\alpha = P(\textit{décider } H_1/H_0)$, alors même qu'aucune règle ne serait pertinente, on sélectionne quand même 500 règles en moyenne.

Dans le but de contrôler le risque multiple, nous avons d'abord utilisé des outils de la théorie de l'apprentissage statistique, fondés pour l'essentiel sur la dimension de Vapnik, afin de proposer des bornes uniformes non asymptotiques pour toutes les règles et toutes les mesures considérées (Teytaud et Lallich, 2001). Par la suite (Lallich et Teytaud, 2004), nous avons proposé *BS*, un algorithme fondé sur le bootstrap qui contrôle le risque de 1^{re} espèce sur l'ensemble des tests. Ces méthodes assurent que le risque de faire la moindre fausse découverte soit égal à un seuil fixé, mais elles ont l'inconvénient d'exprimer un point de vue très sévère sur les erreurs, ce qui les rend

peu puissantes et les amène à manquer plus souvent de vraies découvertes.

Pour remédier à ce défaut, nous avons choisi de contrôler non pas le risque, mais le nombre V de fausses découvertes, suivant les procédures de sélection de gènes dévelopées en biostatistique. Nous avons proposé (Lallich *et al.*, 2004) un critère original, *User Adjusted Family Wise Error Rate,* $UAFWER = \Pr(V > V_0)$, que nous contrôlons au risque δ grâce à une procédure fondée sur le bootstrap. Ce critère est plus tolérant au sens où il assure au risque δ d'avoir au maximum V_0 fausses découvertes. Appliquée à différentes bases de règles, cette procédure a permis d'éliminer jusqu'à 50% de règles non significatives. La méthode proposée a le double avantage de gérer la dépendance entre les différentes règles, grâce au bootstrap, et de ne pas nécessiter la connaissance de la loi de la statistique de test sous H_0, exigeant seulement une valeur fixe sous H_0.

7 Conclusion

Le principe de notre démarche commune est de mettre en regard deux approches complémentaires du problème de la mesure de l'intérêt des règles d'association, l'une formelle, l'autre empirique. Dans le cadre de l'approche formelle, nous proposons un certain nombre de propriétés opérationnelles qui permettent d'évaluer les mesures. Pour mener à bien l'approche empirique, nous avons développé HERBS, une plateforme d'expérimentation des mesures sur des bases de règles et nous nous sommes donnés de contrôler la validation statistique des règles retenues. Nous avons aussi fait le lien avec l'utilisateur en lui donnant le moyen de choisir la mesure qui correspond le mieux à ses préférences en termes de propriétés. D'autres propriétés intéressantes sont à l'étude et devraient être intégrées dans notre plateforme, ainsi que dans le module d'aide à la décision.

Références

Agrawal R., Imielinski T. et Swami A.N., (1993). Mining association rules between sets of items in large databases. In *ACM SIGMOD Int. Conf. on Management of Data*, pp 207–216.

Azé J. et Kodratoff Y., (2002). Evaluation de la résistance au bruit de quelques mesures d'extraction de règles d'assocation. *EGC 2002*, 1(4) :143–154.

Berti-Equille L., (2004). La qualité des données comme condition vers la qualité des connaissances : un état de l'art. *RNTI-E-1*, pp 95–118.

Blanchard J., Guillet F. et Briand H., (2004). Une visualisation orientée qualité pour la fouille anthropocentrée de règles d'association. *Cahiers Romans de Sciences Cognitives – In Cognito*, 1(3) :79–100.

Brans J.P. et Mareschal B., (1994). The PROMETHEE-GAIA decision support system for multicriteria investigations. *Investigation Operativa*, 4(2) :102–117.

Briand H., Sebag M., Gras R. et Guillet F. (éditeurs), (2004). Mesures de qualité pour la fouille de données. *RNTI-E-1*.

Brin Sergey, Motwani Rajeev et Silverstein Craig, (1997). Beyond market baskets : generalizing association rules to correlations. In *ACM SIGMOD/PODS'97*, pp 265–276.

Brin Sergey, Motwani Rajeev, Ullman Jeffrey D. et Tsur Shalom, (1997). Dynamic itemset counting and implication rules for market basket data. In Peckham Joan, editor, *ACM SIGMOD 1997 Int. Conf. on Management of Data*, pp 255–264.

Church Kenneth Ward et Hanks Patrick, (1990). Word association norms, mutual information an lexicography. *Computational Linguistics*, 16(1) :22–29.

Cohen J., (1960). A coefficient of agreement for nominal scale. *Educational and Psychological Measurement*, 20 :37–46.

Good Irving John. The estimation of probabilities : An essay on modern bayesian methods. The MIT Press, Cambridge, MA, (1965).

Gras R., Ag. Almouloud S., Bailleuil M., Larher A., Polo M., Ratsimba-Rajohn H. et Totohasina A., (1996). *L'implication Statistique, Nouvelle Méthode Exploratoire de Données. Application à la Didactique, Travaux et Thèses*. La Pensée Sauvage.

Gras R., Kuntz P., Couturier R. et Guillet F., (2001). Une version entropique de l'intensité d'implication pour les corpus volumineux. *EGC 2001*, 1(1-2) :69–80.

Gras R., Couturier R., Blanchard J., Briand H., Kuntz P. et Peter P., (2004). Quelques critères pour une mesure de qualité de règles d'association. *RNTI-E-1*, pp 3–31.

Lallich S., Prudhomme E. et Teytaud O., (2004). Contrôle du risque multiple en sélection de règles d'association significatives. *RNTI-E-2*, 2 :305–316.

Lallich S. et Teytaud O., (2002). Évaluation 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.

Lallich S. et Teytaud O., (2004). Évaluation et validation de l'intérêt des règles d'association. *RNTI-E-1*, pp 193–217.

Lallich S., (2002). Mesure et validation en extraction des connaissances à partir des données. Habilitation à Diriger des Recherches – Université Lyon 2.

Lehn R., Guillet F., Kuntz P., Briand H. et Philippé J., (1999). Felix : An interactive rule mining interface in a kdd process. In Lenca P., editor, *Proceedings of the Human Centered Processes Conference*, pp 169–174, Brest, France.

Lehn R., (2000). *Un système interactif de visualisation et de fouille de règles pour l'extraction de connaissances dans les bases de données*. Thèse de Doctorat, Université de Nantes.

Lenca P., Meyer P., Vaillant B. et Picouet P., (2002). Aide multicritère à la décision pour évaluer les indices de qualité des connaissances - modélisation des préférences de l'utilisateur. Rapport de recherche pour le groupe de travail GAFOQUALITÉ de l'action spécifique STIC fouille de bases de données, Département IASC, ENST Bretagne.

Lenca P., Meyer P., Picouet P., Vaillant B. et Lallich S., (2003a). Critères d'évaluation des mesures de qualité en ECD. *Entreposage et Fouille de données*, (1) :123–134.

Lenca P., Meyer P., Vaillant B. et Picouet P., (2003b). 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.

Lenca P., Meyer P., Vaillant B., Picouet P. et S. Lallich, (2004). Évaluation et analyse multicritère des mesures de qualité des règles d'association. *Mesures de Qualité pour la Fouille de Données*, (RNTI-E-1) :219–246.

Lerman I.C., Gras R. et 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.

Lerman I.C. et Azé J., (2003). Une mesure probabiliste contextuelle discriminante de qualité des règles d'association. *EGC 2003*, 1(17) :247–262.

Loevinger J., (1947). A systemic approach to the construction and evaluation of tests of ability. *Psychological monographs*, 61(4).

IBM, (1996). *IBM Intelligent Miner User's Guide, Version 1 Release 1, SH12-6213-00*.

Pasquier N., (2000). *Data Mining : Algorithmes d'extraction et de réduction des règles d'association dans les bases de données*. Thèse de Doctorat, Université Blaise Pascal - Clermont-Ferrand II.

Pearson Karl, (1896). Mathematical contributions to the theory of evolution. regression, heredity and panmixia. *Philosophical Trans. of the Royal Society*, A.

Piatetsky-Shapiro G., (1991). Discovery, analysis and presentation of strong rules. In Piatetsky-Shapiro G. et Frawley W.J., editors, *Knowledge Discovery in Databases*, pp 229–248. AAAI/MIT Press.

Poulet F., (1999). Visualization in data-mining and knowledge discovery. In Lenca P., editor, *Proceedings of the Human Centered Processes Conference*, pp 183–191, Brest, France.

Sebag M. et Schoenauer M. Generation of rules with certainty and confidence factors from incomplete and incoherent learning bases. In Boose J., Gaines B. et Linster M., editors, *EKAW'88*, pp 28–1 – 28–20. (1988).

Tan P.-N., Kumar V. et Srivastava J., (2002). Selecting the right interestingness measure for association patterns. In *Eighth ACM SIGKDD Int. Conf. on KDD*, pp 32–41.

Terano Takao, Liu Huan et Chen Arbee L. P., editors. *Association Rules*, volume 1805 of *Lecture Notes in Computer Science*. Springer, April 2000.

Teytaud O. et Lallich S., (2001). Bornes uniformes en extraction de règles d'association. In *Conférence d'Apprentissage, CAp'01*, pp 133–148.

Vaillant B., Lenca P. et Lallich S., (2004). A clustering of interestingness measures. In *Discovery Science*, volume 3245 of *Lecture Notes in Artificial Intelligence*, pp 290–297. Springer-Verlag.

Vaillant B., (2002). Evaluation de connaissances : le problème du choix d'une mesure de qualité en ECD. Rapport de DEA, ENST Bretagne.