

Étude expérimentale de mesures de qualité de règles d'association

Benoît Vaillant*, Philippe Lenca*
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/>

**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. La validation des connaissances extraites d'un processus d'ECD par un expert métier nécessite de filtrer ces connaissances. Pour ce faire, de nombreuses mesures ont été proposées, chacune répondant à des besoins spécifiques. Ces mesures présentent des caractéristiques variées et parfois contradictoires qu'il convient alors d'examiner. Arguant du fait que la sélection des bonnes connaissances passe aussi par l'utilisation d'un ensemble de mesures adaptées au contexte, nous présentons dans cet article une étude expérimentale de différentes mesures. Cette étude est mise en regard d'une étude formelle synthétisant les qualités des mesures.

Mots-clés : mesures d'intérêt, règles d'association, validation expérimentale, 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, ces derniers étant des experts métier, spécialistes du domaine sur lequel portent les données.

Une règle pour être intéressante doit être "*valide, nouvelle et compréhensible*" [Fayyad *et al.*, 1996]. Ces termes génériques, appliqués à des contextes précis, couvrent des réalités très différentes et l'intérêt d'une règle est une information particulièrement difficile à évaluer objectivement. De plus, le nombre particulièrement important de règles générées par les algorithmes utilisés en ECD ne permet pas aux utilisateurs-experts de faire eux-mêmes la sélection des règles pertinentes sans assistance. Le pré-filtrage automatique est un moyen pour réduire considérablement le nombre de règles qui devront être effectivement présentées au jugement de l'expert.

En permettant d'extraire un sous-ensemble des "meilleures règles" d'un ensemble trop volumineux, les mesures de qualité évaluant l'intérêt des règles jouent alors un rôle essentiel en ECD. La sélection des meilleures règles passe par l'utilisation de mesures adaptées au contexte et les mesures de qualité doivent être conçues 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. Cette difficulté est justifiée par le fait que les mesures présentent des caractéristiques variées et parfois contradictoires [Tan *et al.*, 2002], [Lenca *et al.*, 2003b], [Lallich et Teytaud, 2004]. La mesure *optimale* n'existe pas et une solution consiste à rechercher, non pas des règles optimales, mais un compromis [Francisci *et al.*, 2003].

Dans [Lenca *et al.*, 2003a], nous nous sommes intéressés à l'aide à la sélection de mesures de qualité de règles d'association en fonction du contexte (préférences de l'utilisateur, nature des données, etc.) selon des caractéristiques formelles. Nous y présentons une évaluation de 20 mesures selon 8 critères pour deux scénarios utilisateurs.

L'objectif de l'étude que nous proposons ici est de compléter cette approche formelle au moyen d'une analyse de comportement des mesures en situation sur des données. Cette double approche se justifie par le fait qu'il existe des critères difficiles à évaluer formellement et nécessitant des expériences sur des données. C'est le cas par exemple de la résistance des mesures au bruit, critère étudié par [Azé et Kodratoff, 2002].

Dans cet article, la cartographie des mesures suivant leurs propriétés établie dans [Lenca *et al.*, 2003a] est mise en regard d'une cartographie résultant d'une campagne d'expériences. Nous complétons cette comparaison avec des typologies issues de chaque cartographie. Pour ce faire nous avons développé HERBS, un outil d'analyse de bases de règles et de mesures de qualité, brièvement présenté section 2. La section 3 expose une étude détaillée sur une base de règles d'association à l'aide de HERBS, et la section 4 généralise les résultats expérimentaux sur de nombreuses bases de règles d'association. Nous y distinguons trois classes de mesures. Enfin, la section 5 présente une classification ascendante hiérarchique des mesures et compare le résultat à la classification obtenue expérimentalement. Nous concluons en section 6.

2 Présentation de la plate-forme HERBS

La plate-forme HERBS [Vaillant, 2002] permet d'étudier en situation, c'est-à-dire sur des données, le comportement des mesures de qualité de règles. Nous nous situons clairement en phase de *post-analyse* et nous ne nous préoccupons pas ici de la façon dont les règles ont été créées. Bases de règles, bases de cas et ensemble de mesures à étudier sont ici des entrées de notre analyse. HERBS est un outil graphique permettant de faciliter l'analyse de l'expert en ECD ou de l'expert des données [Vaillant *et al.*, 2003].

2.1 Cadre d'étude et notations

Considérons une base \mathcal{C} de n cas et une base \mathcal{R} de η règles. Les cas correspondent à des éléments décrits par des attributs. Nous restreignons notre étude à des bases de cas tabulaires, c'est-à-dire des bases pour lesquelles chaque cas est un k -uplet, k étant constant sur \mathcal{C} . Les domaines de valeurs considérés sont généralement $\{0,1\}$, \mathbb{N} , \mathbb{R} , ou encore un ensemble fini de chaînes de caractères. Dans de nombreuses situations, il arrive que des valeurs ne soient pas renseignées. Toutefois, nous ne traiterons pas cette problématique. Nous supposons qu'il n'y a pas de valeurs manquantes dans les bases de cas, situation vérifiée par nos exemples pratiques.

Les règles que nous considérons sont du type $A \rightarrow B$, où A (prémisse) et B (conclusion) sont des conjonctions de tests sur les valeurs prises par des attributs. Suivant le domaine de valeurs, les tests sont d'égalité, de supériorité ou d'infériorité. Dans l'algorithme AIS de calcul de règles d'association initialement proposé par [Agrawal *et al.*, 1993], les attributs sont binaires et sont donc testés au plus une fois dans la règle, et la conclusion ne fait intervenir qu'un unique attribut.

On note n_a le nombre de cas de \mathcal{C} vérifiant la prémisse de la règle, n_b le nombre de cas vérifiant sa conclusion, n_{ab} le nombre de cas vérifiant à la fois la prémisse et la conclusion de la règle (exemples), et enfin $n_{a\bar{b}}$ le nombre de cas vérifiant la prémisse mais pas la conclusion de la règle (contre-exemples). On a alors $n_{a\bar{b}} = n_a - n_{ab}$.

Il est important de bien noter à ce stade que nous dissociions complètement l'énonciation d'une règle, et son application à des données. Ainsi, les trois grandeurs caractéristiques d'une règle que sont n_a , n_b , et n_{ab} peuvent potentiellement prendre n'importe quelle valeur entre 0 et n , avec toutefois la contrainte $n_{ab} \leq \min(n_a, n_b)$. A \mathcal{C} donné, il est possible d'écarter les règles ne faisant pas sens (*i.e.* prenant une valeur nulle pour l'une des grandeurs n_a , n_b ou n_{ab}), par exemple en imposant un seuil de support minimal strictement positif pour chaque règle de \mathcal{R} .

Les mesures de qualité que nous considérons sont des fonctions de n , n_a , n_b et n_{ab} . Elles ont pour but d'ordonner les règles de la "meilleure" à la "moins bonne", cette notion de qualité dépendant fortement d'objectifs utilisateur, des données...

On appelle taux de couverture t_c la proportion d'entrées dans la base de cas qui sont exemples d'au moins une règle. On dira qu'un cas de \mathcal{C} est couvert par \mathcal{R} lorsqu'il vérifie au moins une règle de \mathcal{R} , et on note $\mathcal{C}_{|\mathcal{R}}$ le sous-ensemble de \mathcal{C} des cas couverts par \mathcal{R} . On appelle indice de recouvrement t_r le nombre moyen de règles redondantes (*i.e.* le nombre de règles vérifiées pour un cas donné, moins 1) sur $\mathcal{C}_{|\mathcal{R}}$.

2.2 Topologie de la plate-forme

HERBS est organisé autour de MySQL un SGBD relationnel (<http://www.mysql.com/>) et R un outil d'analyse statistique (<http://www.r-project.org/>). Le SGBD permet de stocker les bases de cas et les bases de règles que l'on souhaite utiliser. De plus, il sert à sauvegarder des résultats intermédiaires, tel le nombre d'exemples et de contre-exemples de chaque règle pour une base de cas donnée, ou encore les valeurs prises par des mesures de qualité. Le logiciel R permet d'effectuer des analyses statistiques sur les données et est également utilisé pour afficher et enregistrer des graphiques. Cette architecture est proche d'une structure de type client-serveur car le SGBD n'a pas nécessairement besoin d'être installé sur la machine exécutant HERBS. L'externalisation de la base de données permet d'utiliser un serveur dédié au stockage des données et de partager des résultats expérimentaux.

Chaque règle d'association $A \rightarrow B$ est stockée sous la forme d'un tuple (p,c) où p et c sont des chaînes de caractères correspondant à la description en SQL respectivement de la prémisse A et de la conclusion B de la règle. Le format interne de stockage des règles est donc générique et ne se limite pas au cas des règles d'association : HERBS réalise des opérations de comptage du nombre de cas dans la base vérifiant la prémisse, la conclusion et la règle. Il est par exemple possible de travailler avec des arbres de décision, à condition de les lire comme un ensemble de règles disjointes.

2.3 Présentation d’un schéma d’expérimentation

La plate-forme que nous proposons est une aide à la post-analyse de motifs issus d’un processus d’ECD. Elle ne génère pas de règles à partir de données. Les bases de règles et les bases de cas sont donc considérées comme des entrées.

La tâche qui incombe alors à l’utilisateur est de sélectionner les jeux de données et de spécifier les paramètres de l’expérience. HERBS détectera automatiquement à l’exécution si l’évaluation des règles a déjà été effectuée ou non, et calculera si nécessaire les valeurs prises par des mesures de qualité. Les résultats sont enfin présentés à l’utilisateur, qui peut alors modifier certaines entrées afin de préciser son étude.

L’analyse et la comparaison expérimentale de mesures de qualité que nous proposons repose sur leur application à un couple $(\mathcal{C}, \mathcal{R})$. La base de règles étant considérée comme une entrée, il est intéressant de disposer d’une synthèse de ses caractéristiques. Ainsi, la première étape d’une campagne d’expériences consiste à évaluer quelques caractéristiques globales de l’application de \mathcal{R} à \mathcal{C} . Il est par exemple intéressant de savoir combien de règles sont “extrêmes” : combien de règles n’ont pas de contre-exemple? Combien de règles n’ont pas d’exemple? Combien de règles affaiblissent les connaissances sur la conclusion (*i.e.* telles que $p_{B/A} \leq p_B$)?

Les taux de couverture et de recouvrement sont également de bons indicateurs de la structure induite par les règles sur les données. Le premier caractérise l’adéquation entre les deux bases, et le second la redondance de catégorisation des cas par les règles. Dans le cadre du schéma classique, *apprentissage* et tests en *généralisation* que nous avons mis en œuvre (section 3), il est pertinent de quantifier les deux grandeurs précédentes.

Il peut aussi être judicieux de visualiser graphiquement les comportements de mesures de qualité sur les données en représentant les distributions des valeurs prises pour chaque règle en fonction de grandeurs telles que $n_{a\bar{b}}$, n_b (figure 1).

Afin d’étudier plus finement les rangs de classement des éléments de \mathcal{R} par les différentes mesures, il est possible de sélectionner les règles de \mathcal{R} qui sont au moins k fois sélectionnées par différentes mesures comme étant parmi les N meilleures¹. Pour $k = 1$, on obtient ainsi l’union des ensembles des N meilleures règles, et pour k égal au nombre de mesures prises en compte, on obtient l’intersection de ces ensembles. De telles règles seront dénommées “règles (N, k) —sélectionnées”.

De manière plus synthétique, afin de comparer l’ordonnement des règles, on peut calculer un coefficient d’accord évaluant le degré de cohésion entre deux préordres. Nous avons retenu le coefficient τ_1 , qui est dérivé du τ de Kendall, parmi divers coefficients listés dans [Giakoumakis et Monjardet, 1987].

Finalement, un dernier moyen de comparer le degré d’accord entre les valeurs prises par deux mesures sur une base de règles consiste à tracer les distributions croisées.

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]. Nous précisons dans [Lenca *et al.*, 2003a] les critères d’éligibilité des mesures. HERBS ne reprend actuellement que 18 mesures parmi les 20 étudiées formellement. Nous rappelons la définition de ces mesures par commodité dans le tableau 1.

1. Lorsqu’il y a des règles ex-aequo, toutes les règles sont incluses dans cet ensemble. Ainsi, le cardinal de cet ensemble est supérieur ou égal à N .

Mesure	Définition	Mesure	Définition
SUP	$\frac{n_a - n_{a\bar{b}}}{n}$	SURP	$\frac{n_{ab} - n_{a\bar{b}}}{n_b}$
CONF	$1 - \frac{n_{a\bar{b}}}{n_a}$	SEB	$\frac{n_a - n_{a\bar{b}}}{n_{a\bar{b}}}$
R	$\frac{n_{ab} - n_a n_b}{\sqrt{n_a n_b n_{a\bar{b}}}}$	MC	$\frac{(n_a - n_{a\bar{b}}) n_{a\bar{b}}}{n_a n_b}$
CONFCEN	$\frac{n_{ab} - n_a n_b}{n}$	CONV	$\frac{n_a n_b}{n_{a\bar{b}}}$
PS	$\frac{1}{n} \left(\frac{n_a n_{a\bar{b}}}{n} - n_{a\bar{b}} \right)$	TEC	$\frac{n_a - 2n_{a\bar{b}}}{n_a - n_{a\bar{b}}} = 1 - \frac{1}{\frac{n_a}{n_{a\bar{b}}} - 1}$
LOE	$1 - \frac{n_{a\bar{b}}}{n_a n_b}$	IQC	$2 \frac{n_{ab} - n_a n_b - n_{a\bar{b}}}{n_{ab} + n_{a\bar{b}} - 2n_a n_b}$
ZHANG	$\frac{n_{ab} - n_a n_b}{\max\{n_{ab}, n_{a\bar{b}}, n_b n_{a\bar{b}}\}}$	GI	$\log\left(\frac{n_{ab}}{n_a n_b}\right)$
- INDIMP	$\frac{n_{a\bar{b}} - n_a n_b}{\sqrt{n_a n_b}}$	IPD ²	$P\left[\mathcal{N}(0,1) > IndImp^{CR/B}\right]$
LIFT	$\frac{n_{ab}}{n_a n_b}$	LAP	$\frac{n_{ab} + 1}{n_a + 2}$

TAB. 1 – Mesures étudiées

3 Étude de mesures de qualité : un exemple détaillé

Nous présentons dans cette section une étude comparative des 18 mesures sur une même base de règles, réalisée à partir de HERBS.

3.1 Présentation des données

Les données utilisées dans cette section proviennent de l'UCI Repository³. Ces données ont la particularité de provenir de deux bases de cas, *Solarflare1* et *Solarflare2*, que nous notons \mathcal{SF}_1 et \mathcal{SF}_2 . Elles sont décrites au moyen des mêmes attributs et contiennent respectivement 323 et 1066 cas. Nous avons utilisé l'algorithme APRIORI [Borgelt et Kruse, 2002] afin de générer une base de règles \mathcal{R}_1 à partir de \mathcal{SF}_1 et une base de règles \mathcal{R}_2 à partir de \mathcal{SF}_2 , avec un support minimal de 20% et une confiance minimale de 85%. \mathcal{R}_1 contient 5402 règles, et \mathcal{R}_2 6312. Les statistiques descriptives de l'évaluation croisée de $\mathcal{SF}_1, \mathcal{SF}_2$ sur $\mathcal{R}_1, \mathcal{R}_2$ sont données dans le tableau 2.

couple (\mathcal{C}, \mathcal{R})	η	nombre de règles telles que :			t_c	t_r
		$n_{a\bar{b}} = 0$	$n_{ab} = 0$	$p_{B/A} \leq p_B$		
($\mathcal{SF}_1, \mathcal{R}_1$)	5402	1022	0	1016	100%	1828,6
($\mathcal{SF}_2, \mathcal{R}_1$)	5402	2195	1378	2192	100%	1536,5
($\mathcal{SF}_1, \mathcal{R}_2$)	6312	226	118	1080	100%	1427,1
($\mathcal{SF}_2, \mathcal{R}_2$)	6312	2431	0	1409	100%	2276,6

TAB. 2 – Description de l'évaluation croisée des bases *Solarflare*

Il est clair que de nombreuses règles ont tendance à affaiblir la connaissance de B. De plus, lorsque l'on applique une base de règles sur une base de test n'ayant pas servi à construire la base de règles, beaucoup de règles n'ont plus d'exemple. Nous allons donc effectuer un premier traitement, afin d'éliminer les règles problématiques.

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

3. <http://ftp.ics.uci.edu/pub/machine-learning-databases/solar-flare/>

3.2 Prétraitement d'une base de règles

Etant donnée une base de règles d'association issues par exemple de l'application d'un algorithme du type APRIORI, il est logique de filtrer cette base pour ne conserver que les règles pour lesquelles la fréquence de B/A est plus grande que celle de B, soit $p_{B/A} > p_B$. On peut aller plus loin et ne vouloir conserver que les règles pour lesquelles la fréquence de B/A est significativement plus grande que celle de B, ou règles significatives du point de vue statistique. Il faut alors calculer une valeur critique pour le test de l'hypothèse d'indépendance de A et B (H_0) face à une hypothèse de dépendance positive (H_1).

Dans le cadre d'une modélisation hypergéométrique, on teste l'hypothèse d'indépendance des *itemsets* A et B en raisonnant à marges n_a et n_b fixées. Le nombre d'exemples n_{ab} suit la loi hypergéométrique $H(n, n p_a, p_b)$, ou de façon équivalente la loi $H(n, n p_b, p_a)$, sous H_0 [Lallich, 2002].

On peut opérer l'approximation normale de cette loi hypergéométrique sous des conditions peu contraignantes, à savoir $t_{ab} \geq 5$ et $t_{a\bar{b}} \geq 5$. Notant Φ la fonction de répartition de la loi normale centrée réduite $N(0,1)$, on a :

$$\frac{n_{ab} - t_{ab}}{\sqrt{\frac{n}{n-1} t_{ab} p_{\bar{b}}}} \cong N(0,1), \text{ soit } \frac{n_{ab} - t_{ab}}{\sqrt{n p_a p_b p_{\bar{a}} p_{\bar{b}}}} \cong N(0,1)$$

$$\text{d'où } P(n_{ab} \geq n_{ab \text{ obs}}) = 1 - \Phi\left(\frac{n_{ab \text{ obs}} - t_{ab}}{\sqrt{n p_a p_b p_{\bar{a}} p_{\bar{b}}}}\right) \cong 1 - \Phi\left(\sqrt{n} \frac{p_{ab} - p_a p_b}{\sqrt{p_a p_b p_{\bar{a}} p_{\bar{b}}}}\right)$$

$$\text{ainsi } P(n_{ab} \geq n_{ab \text{ obs}}) = 1 - \Phi(r\sqrt{n})$$

On voit ainsi que la façon la plus simple de caractériser les règles statistiquement significatives consiste à calculer le coefficient de corrélation linéaire entre les variables booléennes A et B, noté r , et à en déduire la *p-value* du nombre de contre-exemples qui vaut $1 - \Phi(r\sqrt{n})$. Dès lors, une règle est statistiquement significative au risque α , si $r \geq \frac{u_{1-\alpha}}{\sqrt{n}}$. Au risque 0,05 une règle est statistiquement significative si $r \geq \frac{1,645}{\sqrt{n}}$.

Une méthode classique utilisée pour valider un modèle en apprentissage est de ne prendre en compte qu'une partie des données à notre disposition afin d'effectuer l'apprentissage et de valider le modèle issu du processus sur l'autre partie des données. Nous appliquons donc cette méthode sur \mathcal{R}_1 , et la filtrons selon le principe énoncé précédemment en utilisant \mathcal{SF}_2 comme base de cas. On obtient alors une base de règles, que l'on notera \mathcal{R}_1^2 ne contenant plus que 2994 règles, et dont les caractéristiques sur les deux bases de cas sont données dans le tableau 3.

couple (\mathcal{C} , \mathcal{R})	nombre de règles telles que :			t_c	t_r
	$n_{a\bar{b}} = 0$	$n_{ab} = 0$	$p_{B/A} \leq p_B$		
(\mathcal{SF}_1 , \mathcal{R}_1^2)	803	0	155	100%	1043,6
(\mathcal{SF}_2 , \mathcal{R}_1^2)	951	0	0	100%	1181,4

TAB. 3 – Description de l'évaluation de \mathcal{R}_1^2 sur \mathcal{SF}_1 et \mathcal{SF}_2

Le nombre de règles n'ayant pas d'exemple est nul pour les deux bases de cas. Ceci est tout à fait normal. En effet, la première base de cas est celle d'apprentissage, les règles de \mathcal{R}_1^2 sont donc issues d'exemples de \mathcal{SF}_1 . Il ne peut y avoir de règles sans exemple pour cette base de cas. De plus, du fait de l'opération de filtrage, il est clair que les règles n'ayant pas d'exemple pour \mathcal{SF}_2 ont été rejetées. Par contre, il reste

des règles pour lesquelles $p_{B/A} \leq p_B$ lorsqu'on les évalue sur \mathcal{SF}_1 , mais leur nombre a fortement diminué. On remarque également que le nombre de règles n'ayant pas de contre-exemple est plus faible après filtrage.

Enfin, le taux de couverture est toujours de 100%, et le taux de recouvrement reste élevé. Il doit donc y avoir de nombreuses règles à fort support. Ceci se voit sur la figure 1, qui représente la distribution des valeurs du support en fonction de $n_{a\bar{b}}$ sur les deux bases de cas. Clairement, \mathcal{R}_1^2 contient de nombreuses règles à fort support.

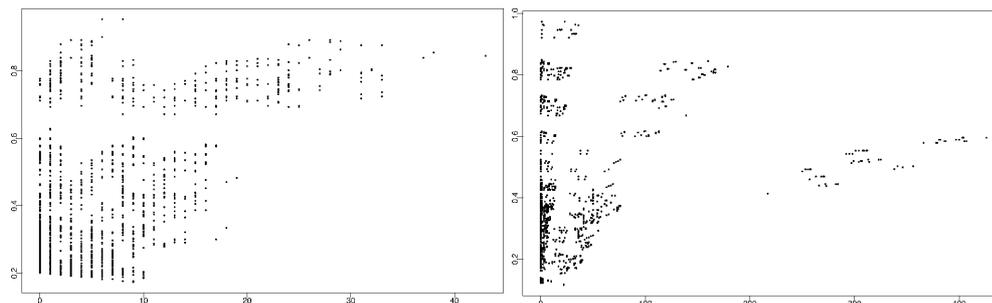
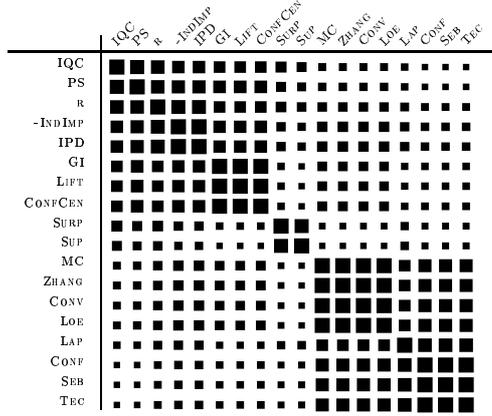


FIG. 1 – Distribution de la valeur du support en fonction du nombre de contre-exemples pour \mathcal{R}_1^2 sur \mathcal{SF}_1 et \mathcal{SF}_2

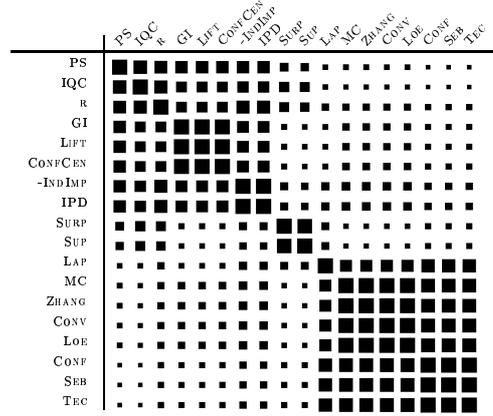
3.3 Comparaison expérimentale du comportement de mesures de qualité

A partir de cette base de règles filtrée, nous allons pouvoir comparer le comportement des mesures de qualité, et voir si en pratique celui-ci est cohérent avec l'étude formelle de leurs propriétés menée dans [Lenca *et al.*, 2003a]. Afin de visualiser les différences de comportement, les tableaux 4 et 5 présentent les valeurs du coefficient d'accord (la valeur correspondante est proportionnelle à la surface du carré) associées aux différents préordres induits par les mesures de qualité sur $(\mathcal{SF}_1, \mathcal{R}_1^2)$ et $(\mathcal{SF}_2, \mathcal{R}_1^2)$. Les lignes et les colonnes ont été réarrangées afin de mieux visualiser les structures en blocs selon la méthode AMADO [Chauchat et Risson, 1998]. A partir des deux tableaux de comparaison de préordres, on distingue clairement trois groupes de mesures bien séparés. Au sein de ces groupes, on peut discerner des rapprochements encore plus forts, dont certains s'expliquent aisément, les mesures correspondantes étant des transformées monotones les unes des autres. C'est le cas par exemple du groupe contenant CONF, SEB, TEC auquel s'ajoute également LAP, qui ne diffère que de très peu de CONF (section 5). La surprenante proximité entre SUP et SURP est liée en grande partie à un biais introduit par le mécanisme de production des règles d'association. En effet, les forts seuils de support et de confiance utilisés bornent les valeurs possibles de n_a et n_b dans des intervalles limités alors même que SUP et SURP se distinguent d'autant plus que ces intervalles sont grands.

Étude expérimentale de mesures de qualité de règles d'association



TAB. 4 – Comparaison de préordres pour $(\mathcal{SF}_1, \mathcal{R}_1^2)$



TAB. 5 – Comparaison de préordres pour $(\mathcal{SF}_2, \mathcal{R}_1^2)$

4 Autres résultats expérimentaux

Des expériences similaires ont été menées sur d'autres bases de données issues du même site, sans toutefois subdiviser les données en deux. Il n'est donc pas possible de procéder à une phase d'apprentissage puis à une phase de test. De plus, le filtrage statistique des bases n'a pas été opéré.

Les paramètres de réglages pour APRIORI ont été fixés expérimentalement, afin d'obtenir les bases de règles associées. Les caractéristiques de ces bases sont regroupées dans la table 6. Les fortes différences en termes de nombre de règles sont dues au nombre de modalités que peuvent prendre les attributs dans les bases de cas. Pour générer les bases de règles de *Cmc* et *Cmc2*, une option supplémentaire a été utilisée : l'algorithme APRIORI qui normalement n'effectue qu'un parcours limité aux premiers nœuds du treillis formé par l'ensemble des modalités des attributs a été forcé de parcourir la totalité du treillis, sous contrainte de support suffisant.

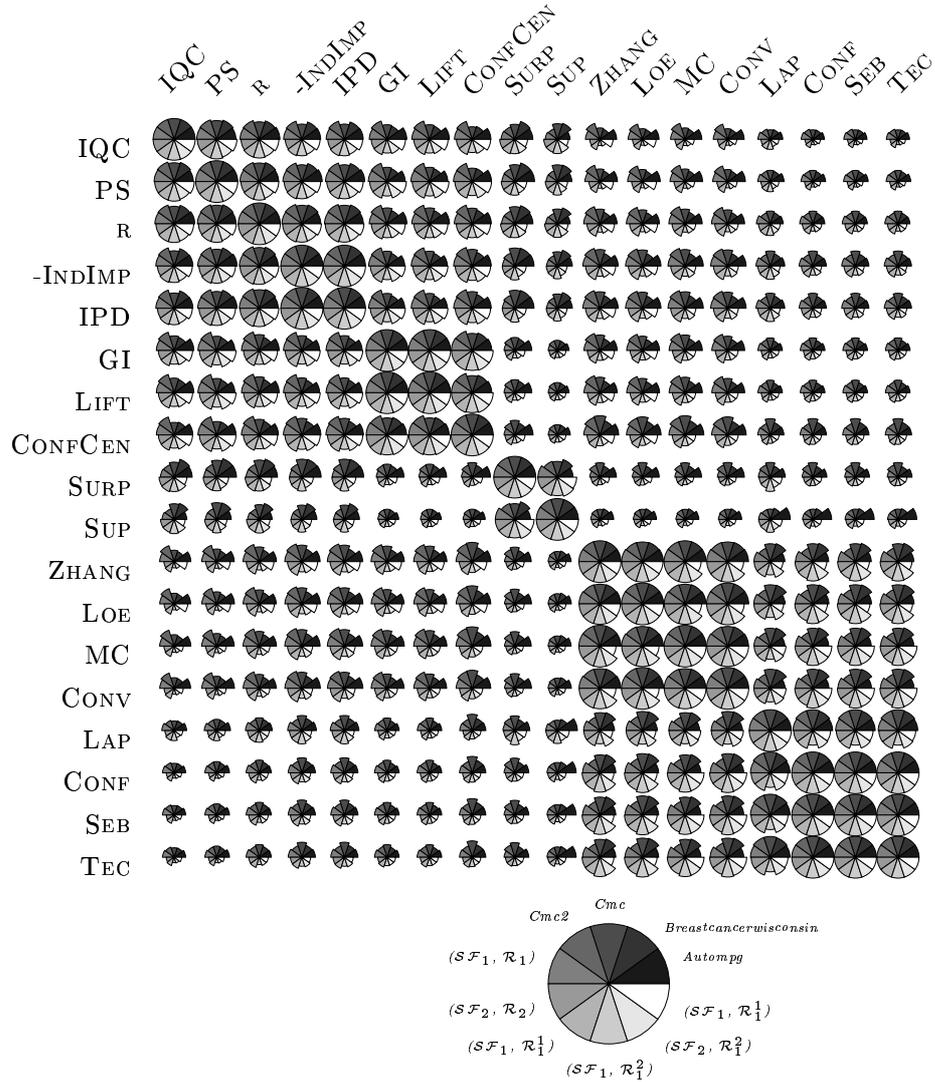
nom	n	sup_{min}	$conf_{min}$	η	t_c	t_r
<i>Autompg</i>	392	5	50	49	81,37%	0,79
<i>Breastcancerwisconsin</i>	683	10	70	3095	96,19%	646
<i>Cmc</i>	1473	5	60	2878	100%	259
<i>Cmc2</i> ⁴	1473	5	60	825	92,93%	84

TAB. 6 – Résumé des différentes bases utilisées pour les expériences

Nous avons de plus calculé le coefficient d'accord τ_1 pour les couples issus de *Solarflare1* et *Solarflare2* suivants : $(\mathcal{SF}_1, \mathcal{R}_1)$, $(\mathcal{SF}_2, \mathcal{R}_2)$, $(\mathcal{SF}_1, \mathcal{R}_1^1)$, $(\mathcal{SF}_1, \mathcal{R}_1^2)$, $(\mathcal{SF}_2, \mathcal{R}_2^1)$, $(\mathcal{SF}_2, \mathcal{R}_2^2)$. Au total, nous avons donc généré 10 matrices de comparaison de préordres. Le tableau 7 synthétise l'ensemble des résultats (la valeur correspondante est proportionnelle à la distance au centre du secteur). La méthode AMADO a été ap-

4. La base de règle de *cmc2* a été obtenue à partir de celle de *cmc* en effectuant un filtrage pour ne conserver que les règles ayant un bon lift, en imposant un seuil minimal de 1,2.

pliquée à une matrice constituée des valeurs moyennes de chaque coefficient d'accord. Les résultats ainsi représentés sont globalement concordants entre eux. On retrouve bien une séparation des mesures en trois classes. Il existe des différences en fonction des données utilisées. Ces différences justifient une fois de plus la nécessité d'étudier les propriétés et les comportements des mesures en situation.



TAB. 7 – Comparaisons de préordres induits par 18 mesures pour 10 couples $(\mathcal{C}, \mathcal{R})$.

5 Mise en regard des classifications issues de l'étude expérimentale et d'approches formelles

Les 8 propriétés formelles (tableau 8) proposées dans [Lenca *et al.*, 2003a] nous ont permis de construire une matrice de décision évaluant les mesures. Leur évaluation a fondé la construction d'une matrice de distance entre les mesures après avoir passé la matrice de décision sous forme disjonctive complète et calculé le carré des distances euclidiennes. Cette façon de procéder n'émet aucun jugement de valeur, contrairement à notre approche *aide à la décision*. Nous n'avons gardé ici que les critères ayant une influence sur les résultats expérimentaux (g_7 et g_8 n'ayant aucun impact sur l'ordonnement des règles et donc sur les résultats expérimentaux discutés ici, sont exclus).

Propriétés	Sémantique
g_1	non symétrie
g_2	décroissance avec n_b
g_3	situation à l'indépendance
g_4	situation à la règle logique
g_5	non-linéarité autour de 0^+
g_6	prise en compte de n
g_7	facilité à fixer un seuil
g_8	intelligibilité

TAB. 8 – Propriétés des mesures

En appliquant une classification ascendante hiérarchique (CAH) avec le critère de WARD on distingue trois classes principales : {PS, IQC, GI, LIFT, R, -INDIMP, IPD }, {CONF CEN, SURP, SUP, LAP, CONF, SEB, TEC } et {MC, ZHANG, CONV, LOE }. Le tableau 9 compare les deux approches. Du point de vue formel, seule la classe 2 présente des désaccords avec les résultats expérimentaux. {LAP, CONF, SEB, TEC } rejoint la classe 3. Les mesures CONF, SEB et TEC présentent en particulier la propriété d'avoir une valeur fixe à la règle logique, comme MC, ZHANG, CONV et LOE, au contraire de SUP et SURP. Cette propriété est déterminante dans le comportement expérimental des mesures. En effet, l'évaluation donnée aux règles logiques à très fort conséquent est très différente suivant la modalité de g_4 . Le fait que LAP suive CONF expérimentalement est simplement lié au fait que LAP se différencie de CONF uniquement si n_a est proche de 0 (or cette situation est impossible étant donné le taux de support que nous avons utilisé pour APRIORI ; ce cas révèle un biais algorithmique).

CAH vs Expérimentale	Classe 1	Classe 2	Classe 3
Classe 1	PS, IQC, GI, CONF CEN LIFT, R, -INDIMP, IPD		
Classe 2	SURP, SUP		LAP, CONF, SEB, TEC
Classe 3	MC, ZHANG, CONV, LOE		

TAB. 9 – Comparaison des classes entre les approches CAH et expérimentale

6 Conclusions

Les mesures d'intérêt des règles d'association jouent un grand rôle en ECD mais elles possèdent des caractéristiques diverses qu'il faut étudier afin d'utiliser la (ou les) mesure(s) la (les) plus adaptée(s) au contexte. Nous présentons une plate-forme, HERBS, permettant de mener des campagnes d'expérimentation du comportement des mesures et exposons les résultats pour 18 mesures sur 10 bases de données.

Ces campagnes nous permettent d'isoler 3 classes de mesures. Nous avons précédemment étudié formellement 20 mesures selon 8 critères et distingué les comportements des mesures selon des scénarios utilisateur. Exploitant une partie de ce travail, nous formons une matrice de distance entre les mesures. Nous construisons une seconde classification à l'aide d'une CAH avec le critère de WARD et comparons les deux systèmes de classes. L'approche expérimentale s'avère être un complément indispensable de l'étude formelle, elle nous a permis de vérifier la pertinence des critères proposés mais aussi de remettre en cause certains choix précédemment effectués. Par exemple, le fait qu'une mesure ait une valeur fixe pour la règle logique a un effet pervers en autorisant la sélection des règles logiques à fort conséquent. Ceci peut être important dans certains secteurs, par exemple le domaine médical, mais être sans intérêt dans d'autres.

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 International 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.
- [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.
- [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.
- [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.
- [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.

- [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. A paraître.
- [Lallich, 2002] S. Lallich. Mesure et validation en extraction des connaissances à partir des données. Habilitation à Diriger des Recherches – Université Lyon 2, 2002.
- [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 (Entroposage 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.
- [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 International Conference on Knowledge Discovery and Data Mining*, pages 32–41, 2002.
- [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

It is an accepted fact that the validation of knowledge extracted within a KDD process by a field expert requires a filtering step. This is usually done using quality measures, each of these responding to specific needs. The measures can be described by various and sometimes contradictory properties that have to be studied. Stating that extracting interesting rules also means using an adapted measure, we present in this article an experimental study of the behaviour of different quality measures. This study is compared to a previous analysis of formal properties, based on a multi-criteria decision aid process, assessing the issue of selecting a quality measure fitting an expert's expectations.

Keywords : interestingness measures, association rules, experimental validation, decision aid.