

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. Cet article présente nos travaux sur la mesure de l'intérêt des règles d'association. 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 mesures et la typologie expérimentale. 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 $A \rightarrow B$ telles que définies dans (Agrawal *et al.*, 1993) : dans une base de données transactionnelles, $A \rightarrow B$ signifie que si les articles qui constituent A sont dans *le panier d'une ménagère*, alors le plus souvent les articles qui constituent B le sont aussi. 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.

Cet article présente une synthèse de l'ensemble de nos travaux sur la qualité des règles d'association. C'est un résumé de (Vaillant *et al.*, 2005).

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 plate-forme 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

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 $A \rightarrow B$ 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 $A \rightarrow B$ doit être distinguée de l'implication $A \Rightarrow B$ et de l'équivalence $A \Leftrightarrow B$. Ainsi à la différence de (Tan *et al.*, 2002), (Hilderman et Hamilton, 2001), nous nous sommes limités aux mesures (nous en avons étudié 20) qui sont décroissantes avec $n_{a\overline{b}}$ (resp. croissantes avec n_{ab})¹.

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_a ², 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 qui ne dépend pas des marges 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. 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.

¹Soit E l'ensemble des enregistrements, on note $n = |E|$; pour $A \rightarrow B$, on note $n_a = |A|$, le nombre d'enregistrements vérifiant A , $n_b = |B|$, le nombre d'enregistrements vérifiant B , $n_{ab} = |A \cap B|$, le nombre d'exemples de la règle et $n_{a\overline{b}} = |A \cap \overline{B}|$, le nombre de contre-exemples à la règle.

²Pour $X \subseteq E$, on remplace n_X/n par p_X lorsque l'on considère les fréquences relatives plutôt que les fréquences absolues

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

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 :

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

- Évaluation 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.

A partir de HERBS nous avons obtenu une classification expérimentale des mesures. Cette dernière est mise en regard (section 5) d'une classification obtenue à partir de propriétés formelles (Vaillant *et al.*, 2004).

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 multicritères 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.

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. Cette partition formelle est comparée à la partition issue du comportement expérimentale des mesures dans (Vaillant *et al.*, 2004).

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. Telle mesure dépasse-t-elle significativement une valeur fixée, par exemple 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 $\alpha = P(\text{décider } H_1/H_0)$, 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 développées 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 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.
- Blanchard J., Guillet F. et Briand H. (2004). Une visualisation orientée qualité pour la fouille anthropocentrée de règles d'association. *Cahiers Romains de Sciences Cognitives – In Cognito*, 1(3) :79–100.
- 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.
- Hilderman Robert J. et Hamilton Howard J. (2001). Evaluation of interestingness measures for ranking discovered knowledge. *Lecture Notes in Computer Science*, 2035 :247–259.
- 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 (AS 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.

Mesurer l'intérêt des règles d'association

- 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 (AS STIC fouille de bases de données), 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.
- 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.
- 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.
- 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., Meyer P., Prudhomme E., Lallich S. et Lenca P. (2005). Mesurer l'intérêt des règles d'association. In *Atelier Qualité des Données et des Connaissances (EGC 2005)*, pp 69–78.
- Vaillant B. (2002). Evaluation de connaissances : le problème du choix d'une mesure de qualité en ECD. Rapport de DEA, ENST Bretagne.