



**HAL**  
open science

## Critères d'évaluation des mesures de qualité en ECD

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

► **To cite this version:**

Philippe Lenca, Patrick Meyer, Philippe Picouet, Benoît Vaillant, Stéphane Lallich. Critères d'évaluation des mesures de qualité en ECD. XXXV Journées de Statistique, Jun 2003, Lyon, France. pp.647 - 650. hal-02301475

**HAL Id: hal-02301475**

**<https://hal.science/hal-02301475>**

Submitted on 4 Jul 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

# Critères d'évaluation des mesures de qualité en ECD

**Philippe Lenca\*** — **Patrick Meyer\*\*** — **Philippe Picouet\***  
**Benoît Vaillant\*** — **Stéphane Lallich\*\*\***

\* *Département IASC*  
*ENST Bretagne BP 832*  
*29285 Brest Cedex*  
*{prenom.nom}@enst-bretagne,*

\*\*\* *Laboratoire ERIC*  
*Université Lumière - Lyon 2*  
*5 avenue Pierre Mendès-France*  
*69676 Bron Cedex*  
*lallich@univ-lyon2.fr*

\*\* *Département de Mathématique*  
*Université de Liège*  
*Bât. B37 Grande Traverse, 12*  
*B-4000 Liège, Belgique*  
*P.Meyer@ulg.ac.be*

---

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

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

*MOTS-CLÉS : mesure de qualité de règle, objectif, expert*

*KEYWORDS: rule quality measure, goal, expert*

---

## 1. Introduction

L'un des objectifs majeurs de l'extraction de connaissances à partir de données (ECD) est de produire des règles intéressantes du point de vue de l'utilisateur. L'intérêt d'une règle peut alors revêtir de nombreuses formes. Celle-ci doit être valide, nouvelle, compréhensible [FAY 96] et respecter des exigences minimales [PIA 91]. De son côté, l'utilisateur peut rechercher des règles aux comportements très divers lorsqu'elles s'appliquent concrètement à ses données : maximiser le nombre de données concernées, minimiser le nombre de contre-exemples, etc.

Les mesures d'évaluation de la qualité des règles jouent alors un rôle essentiel dans les processus d'ECD. Elles permettent, au sortir d'un processus automatisé qui a généré des règles en grandes quantités, de sélectionner pour un utilisateur celles qui sont adaptées à son besoin. Des travaux [HIL 01, LAL 02] ont mis en évidence les caractéristiques et qualités très diverses de ces mesures et il est aisé de constater que bien souvent l'ensemble des  $n$  meilleures règles varie selon la mesure utilisée [TAN 02, VAI 03]. Il apparaît aujourd'hui que le choix de la mesure est une phase cruciale si on souhaite filtrer efficacement les règles produites en trop grand nombre [LEN 03].

Dans le contexte précis des règles d'association, l'objectif de notre travail est triple : produire une liste de propriétés, qu'il s'agit d'exprimer tantôt sémantiquement du point de vue des objectifs de l'utilisateur tantôt du point de vue statistique ; produire une liste pertinente de mesures et les évaluer sur les propriétés précédentes ; justifier et mettre en œuvre un processus d'aide multicritère à la décision permettant à l'expert de choisir parmi les mesures disposition, celle(s) qui est (sont) le plus susceptible(s) de filtrer efficacement les règles produites.

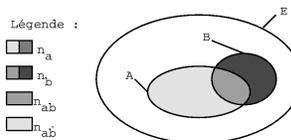
## 2. Mesures éligibles et critères de choix

Nous nous limitons volontairement aux mesures relatives à l'intérêt des règles d'association telles que celles-ci ont été définies par [AGR 93]. On distingue ainsi les règles d'association  $A \rightarrow B$ , qui se focalisent sur les coprésences en traitant les variables de façon non symétrique, de l'implication  $A \Rightarrow B$  et de l'équivalence  $A \Leftrightarrow B$ . Une bonne façon de préciser le critère d'éligibilité des mesures de l'intérêt des règles d'association consiste à ne retenir que les mesures qui sont décroissantes avec  $n_{a\bar{b}}$  (resp. croissantes avec  $n_{ab}$ ), les effectifs marginaux étant fixés. C'est ainsi que sont d'emblée exclues certaines mesures comme le *khi 2* ou le  $r^2$  de Pearson, ainsi que la mesure de Pearl.

Nous avons retenu les principales mesures rencontrées dans la littérature pour évaluer l'intérêt des règles d'association : 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. Une première catégorie de mesures est caractérisée par le fait que celles-ci sont des transformées affines de la confiance ayant pour objet de comparer celle-ci à  $p_b$ . Cette comparaison se fait le plus souvent en centrant la confi-

## Critères d'évaluation des mesures de qualité en ECD

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



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

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

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.

**Figure 1.** Notations.

ance sur  $p_b$  avec différents coefficients d'échelle (confiance centrée, mesure de Pearl, mesure de Piattetsky-Shapiro, mesure de Loevinger, mesure de Zhang, coefficient de corrélation, indice d'implication). Elle peut aussi se faire en divisant la confiance par  $p_b$  (lift ou le 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  $\frac{p_{\bar{b}|a}}{p_{\bar{b}}}$  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. La liste détaillée de ces mesures et leurs formules se trouve dans la version longue de notre papier.

Nous proposons huit critères possibles pour évaluer, sur des échelles ordinales, l'ensemble des mesures éligibles. Nous en présentons brièvement trois ci-dessous, les autres seront développés dans la version longue du papier.

L'antécédent et le conséquent d'une règle n'ayant pas le même rôle, il est souhaitable que la mesure évalue de façon différente les règles  $A \rightarrow B$  et  $B \rightarrow A$  (critère de **dissymétrie**). Pour une même proportion d'exemples  $p_{ab}$ , une règle est d'autant plus intéressante que  $p_b$  est faible (critère de **décroissance avec  $p_b$** ). De plus, 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 (critère de **facilité à fixer un seuil**). 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 ([LAL 02]), mais comme un paramètre de contrôle. Dans certains cas, la mesure est elle-même définie comme une telle probabilité critique, ainsi l'intensité d'implication ([GRA 79]) ou la mesure IPD ([LER 03]). Sinon, on peut le plus souvent définir un tel seuil en partant des trois types de modélisation proposée par [LER 70] pour établir la loi de  $n_{ab}$  sous l'hypothèse d'absence de lien ( $H_0$ ).

### 3. Perspectives

Disposant d'un ensemble représentatif des mesures éligibles couramment utilisées pour évaluer la qualité des règles d'association et d'un ensemble de critères pertinents du point de vue de l'utilisateur nous nous proposons d'évaluer (voir de classer) les mesures en fonction des objectifs de ce dernier. Une telle étude, qui a déjà montré de belles perspectives ([TAN 02, LEN 03]) sera présentée dans la version longue de notre papier. L'ensemble des critères reste bien évidemment à élargir. Par exemple, des critères importants mais difficiles à évaluer formellement doivent être étudiés concrètement en situation sur des données. C'est le cas, par exemple du pouvoir discriminant d'une mesure. Enfin n'oublions pas le rôle certainement essentiel de l'utilisateur et les approches dites subjectives.

### 4. Bibliographie

- [AGR 93] AGRAWAL R., IMIELINSKI T., SWAMI A. N., « Mining Association Rules between Sets of Items in Large Databases », BUNEMAN P., JAJODIA S., Eds., *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, Washington, D.C., 26–28 1993, p. 207–216.
- [FAY 96] FAYYAD U., PIATETSKY-SHAPIO G., SMYTH P., UTHURUSAMY R., Eds., *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, 1996.
- [GRA 79] GRAS R., « Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques », Master's thesis, Université de Rennes I, 1979.
- [HIL 01] HILDERMAN R. J., HAMILTON H. J., « Evaluation of Interestingness Measures for Ranking Discovered Knowledge », *Lecture Notes in Computer Science*, vol. 2035, 2001, p. 247–259.
- [LAL 02] LALLICH S., TEYTAUD O., « Evaluation et validation de l'intérêt des règles d'association », Rapport de recherche pour le groupe de travail GAFOQUALITÉE de l'action spécifique STIC fouille de bases de données, 2002, E.R.I.C., Université Lyon 2.
- [LEN 03] LENCA P., MEYER P., VAILLANT B., PICOUEP P., « Aide multicritère à la décision pour évaluer les indices de qualité des connaissances - Modélisation des préférences de l'utilisateur », *RSTI-RIA*, vol. 1, n° 17, 2003, p. 271–282, Lavoisier.
- [LER 70] LERMAN I., *Classification et analyse ordinaire des données*, Dunod, 1970.
- [LER 03] LERMAN I., AZÉ J., « Une Mesure Probabiliste Contextuelle Discriminante de Qualité des Règles d'Association », *RSTI-RIA*, vol. 1, n° 17, 2003, p. 247–262, Lavoisier.
- [PIA 91] PIATETSKY-SHAPIO G., « Discovery, analysis and presentation of strong rules », PIATETSKY-SHAPIO G., FRAWLEY W., Eds., *Knowledge Discovery in Databases*, AAAI/MIT Press, 1991, p. 229-248.
- [TAN 02] TAN P.-N., KUMAR V., SRIVASTAVA J., « Selecting the Right Interestingness Measure for Association Patterns », *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2002, p. 32–41.
- [VAI 03] VAILLANT B., PICOUEP P., LENCA, « Une plateforme extensible pour l'évaluation de mesures de qualité de règles », *RSTI-RIA*, vol. 1, n° 17, 2003, Page 556, Lavoisier.