

Un cadre formel pour l'étude des mesures d'intérêt des règles d'association

Yannick Le Bras^{1,3}, Stéphane Lallich², Philippe Lenca^{1,3}

¹Institut Télécom, Télécom Bretagne
UMR CNRS 3192 Lab-STICC

³Université Lyon, Lyon 2
Laboratoire ERIC

³Université européenne de Bretagne
{prenom.nom}@telecom-bretagne.eu stephane.lallich@univ-lyon2.fr

Depuis la définition du problème de l'extraction des règles d'association et la proposition de l'algorithme Apriori [1], beaucoup de travaux se sont focalisés sur l'extraction et l'évaluation des règles à partir de mesures objectives. Ces dernières sont essentiellement utilisées pour filtrer les règles extraites. Ces travaux ont donné lieu à quelques améliorations algorithmiques, un nombre important d'améliorations techniques et de très nombreuses nouvelles mesures.

Cependant, peu d'entre eux se sont attaché à mêler extraction et évaluation. Nous introduisons ici un cadre formel d'étude des règles d'association et des mesures d'intérêt qui permet une étude analytique des ces objets. Ce cadre s'appuie sur la notion de table de contingence d'une règle et via la modélisation par une fonction de trois variable des mesures d'intérêt, permet une étude mathématique des mesures et de leurs propriétés algorithmiques. Nous détaillons le cas de trois de ces propriétés : la *all-confidence* [6], la *Universal Existential Upward Closure* [7], et la propriété d'élagage pour les *règles optimales* [5]. Chacune des ces propriétés est dans un premier temps généralisée, puis nous proposons à partir du cadre formel des conditions d'existence nécessaire [4, 3], suffisante [4, 3], ou nécessaire et suffisante [2]. Ces conditions sont alors appliquées à 42 mesures et permettent pour chaque mesure de proposer un ensemble de propriétés algorithmiques (et les algorithmes sous-jacents) qu'elles vérifient. L'impact de ces propriétés sera illustré à partir d'expérimentations sur différents jeux de données.

References

- [1] Rakesh Agrawal and Ramakrishnan Srikant. Fast algorithms for mining association rules in large databases. In Jorge B. Bocca, Matthias Jarke, and Carlo Zaniolo, editors, *20th International Conference on Very Large Data Bases*, pages 478–499. Morgan Kaufmann, 1994.
- [2] Yannick Le Bras, Philippe Lenca, and Stéphane Lallich. On optimal rule mining: A framework and a necessary and sufficient condition of antimonotonicity. In Thanaruk Theeramunkong, Boonserm Kijsirikul, Nick Cercone, and Ho Tu Bao, editors, *13th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, volume 5476 of *Lecture Notes in Computer Science*, pages 705–712. Springer, 2009.
- [3] Yannick Le Bras, Philippe Lenca, and Stéphane Lallich. Mining interesting rules without support requirement: a general universal existential upward closure property. In Robert Stahlbock, Sven F. Crone, and Stefan Lessmann, editors, *Data Mining*, volume 8 of *Annals of Information Systems*, pages 75–98. Springer US, 2010.
- [4] Yannick Le Bras, Philippe Lenca, Sorin Moga, and Stéphane Lallich. All-monotony: A generalization of the all-confidence antimonotony. *4th International Conference on Machine Learning and Applications*, pages 759–764, 2009.
- [5] Jiuyong Li. On optimal rule discovery. *IEEE Transactions on Knowledge and Data Engineering*, 18(4):460–471, 2006.
- [6] Edward Omiecinski. Alternative interest measures for mining associations in databases. *IEEE Transactions on Knowledge and Data Engineering*, 15(1):57–69, 2003.
- [7] Ke Wang, Yu He, and David W. Cheung. Mining confident rules without support requirement. In *10th International Conference on Information and Knowledge Management*, pages 89–96. ACM, 2001.