



HAL
open science

On optimal rules mining: a framework and a necessary and sufficient condition for optimality

Yannick Le Bras, Philippe Lenca, Stéphane Lallich

► **To cite this version:**

Yannick Le Bras, Philippe Lenca, Stéphane Lallich. On optimal rules mining: a framework and a necessary and sufficient condition for optimality. PAKDD 2009 : Pacific-Asia Conference on Knowledge Discovery and Data Mining, Apr 2009, Bangkok, Thailand. pp.705 - 712, 10.1007/978-3-642-01307-2_71 . hal-01881309

HAL Id: hal-01881309

<https://hal.science/hal-01881309v1>

Submitted on 25 Sep 2018

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.

On Optimal Rule Mining: A Framework and a Necessary and Sufficient Condition of Antimonotonicity

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

¹ Institut Telecom; TELECOM Bretagne; UMR CNRS 3192 Lab-STICC
Technopôle Brest-Iroise - CS 83818 - 29238 Brest Cedex 3 - France

{yannick.lebras, philippe.lenca}@telecom-bretagne.eu

² Université de Lyon, Laboratoire ERIC, Lyon 2, France

stephane.lallich@univ-lyon2.fr

³ Université européenne de Bretagne, France

Abstract. Many studies have shown the limits of support/confidence framework used in Apriori-like algorithms to mine association rules. There are a lot of efficient implementations based on the antimonotony property of the support but candidate set generation is still costly. In addition many rules are uninteresting or redundant and one can miss interesting rules like nuggets. One solution is to get rid of frequent item-set mining and to focus as soon as possible on interesting rules. For that purpose algorithmic properties were first studied, especially for the confidence. They allow all confidence rules to be found without a preliminary support pruning. More recently, in the case of class association rules, the concept of optimal rules gave a pruning strategy compatible with more measures. However, all these properties have been demonstrated for a limited number of interestingness measures. We present a new formal framework which allows us to make the link between analytic and algorithmic properties of the measures. We apply this framework to optimal rules, and we demonstrate a necessary and sufficient condition of existence for this pruning strategy, which can be applied to any measure.

Keywords: Association rules, pruning, algorithms, analytic, nuggets, redundance, optimal rules.

1 Introduction

Rule discovery has attracted a lot of attention from data mining researchers probably because rules are an understandable representation of knowledge. In unsupervised paradigm, association rule mining is certainly the most popular method. Since its initial formulation by [1], the problem of association rule mining -and the underlying problem of the frequent set mining- has focused many works. Our paper focuses on association rule mining, especially on classification rule mining as proposed by [2] i.e. with one predetermined target like in supervised learning.

An association rule is a rule $A \rightarrow B$, where A and B are two sets of items (also called itemsets) such that $A \neq \emptyset$, $B \neq \emptyset$ and $A \cap B = \emptyset$, meaning that given a database \mathcal{D} of transactions (where each transaction is a set of items) whenever a transaction T contains A , then T probably contains B also [1]. The problem of mining association rules is then to generate all association rules that have support and confidence greater than a user-specified minimum support and minimum confidence respectively. Support is defined as the proportion of transactions containing A and B in \mathcal{D} (noted $\mathbb{P}(AB)$ or *supp*(AB)), while confidence is the proportion of transactions containing A and B inside the set of transactions containing A in \mathcal{D} (noted $\mathbb{P}(B|A)$ or *conf*($A \rightarrow B$)). Association rule mining is a two-step process: minimum support constraint is firstly applied to find all frequent itemsets in \mathcal{D} and secondly, the frequent itemsets and the minimum confidence constraint are used to form rules.

Finding all frequent itemsets in \mathcal{D} is computationally expensive since there are 2^k possible itemsets where k is the number of items in \mathcal{D} . However using the downward-closure property of support (also called antimonotonicity), very efficient algorithms [3,4] can find all frequent itemsets. An interesting survey may be found in [5]. This step gives in any circumstances an important role to the support constraint and mining interesting rules without the support requirement has been identified as an important problem. Some researches try to avoid the use of support [6] or to avoid the task of threshold fixation [7].

Finding all and only all interesting rules is also a problem because the generated rule sets are quite large, especially within the support-confidence framework while the percentage of interesting rules is often only a very small fraction of the all rules. A first strategy to reduce the number of mined rules consists of increasing the user-specified minimum support. Unfortunately with this strategy many interesting rules and especially nuggets will be missed. A second strategy consists of increasing the user-specified minimum confidence. This will favor rules with large consequent which may give many uninteresting rules.

Another popular strategy is to rank the rules in a post-analysis phase with additional objective measures of interest. A large number of interestingness measures were proposed. Interesting surveys and comparisons may be found in [8,9,10]. Another way of reducing the rule sets is to allow the user to specify which items have to or cannot be part of the left-hand-side or the right-hand-side of the rules. This is the case for associative classification which focuses on association rules whose right-hand-side are restricted to a class attribute [2,11]. Unfortunately these strategies are always subject to the dictatorship of support. A more efficient approach is to apply additional constraints [12] on item appearance or to use additional interestingness measures as soon as possible to reduce both the time to mine databases and the number of founded itemsets [13]. In particular we are here interested in measures that can reduce the search space and also be useful for the evaluation of the quality of mined patterns. We believe that this approach is certainly the most promising one. To overcome the

previous mentioned problems with the dictatorship of the support constraint different solutions were proposed at the algorithmic level and mainly for the confidence measure.

The rest of the paper is organized as follows. In Section 2, we present an overview of recent works that focus on the algorithmic properties of measures. In Section 3, we present a new formal framework which allows us to make the link between analytic and algorithmic properties of the measures. We apply in Section 4 this framework to optimal rules, and we demonstrate a necessary and sufficient condition of existence for applying a pruning strategy with a large set of measures. We conclude in Section 5.

2 State of Art

To get rid of support constraint, several authors proposed to focus on the algorithmic properties of confidence. The goal is to succeed in getting all the high confident rules, even the nuggets of knowledge. We here briefly review some of these works. Some of these works are mainly based on algorithmic properties of confidence, while others focus on different measures, by exploiting their intrinsic properties.

In [14] the authors introduce the *h-confidence* a new measure over itemsets mathematically identical to *all-confidence* proposed by [15]. Both of these measures are antimonotone. They also introduce the concept of cross-support patterns i.e. - uninteresting - patterns involving items with substantially different support levels. The authors thus propose *hyperclique miner*, an efficient algorithm that utilizes both the cross-support and anti-monotone properties of the h-confidence measure. In [16] the authors introduce the *Universal Existential Upward Closure* property based on a certain monotonicity of the confidence. This property applies to classification rules and allows to examine only confident rules of larger size for generating confident rules of smaller size. The authors deduce from this property a top-down confidence-based pruning strategy.

In [17] the authors adapt a technique by [18] to propose a branch-and-bound algorithm for associative classification, based on an antimonotonic property of convexity of the χ^2 . In [19] the authors introduce a new type of antimonotony, called Loose Anti-Monotony, that can be applied for statistical constraints. This concept is then efficiently exploited in an Apriori-like algorithm. In [20] the author introduces the notion of optimal rule set for classification rules. A rule set is optimal if it contains all rules, except those with no greater interestingness than one of its more general rules. This concept defines a pruning strategy that applies to a large set of measures. The author gives an individual proof for 12 measures.

In the following, we introduce a new framework and use it to link together analytic properties of measures and the antimonotone property of [20].

3 A Framework for Measure’s Study

3.1 Adapted Functions of Measure

Firstly, we will precise the concept of associated measure introduced in [21]. In this article, measures are considered as functions of $\mathbb{R}^3 \rightarrow \mathbb{R}$. The authors only focus on the parametrization of interestingness measures in function of the number of examples, antecedents, and consequents. Since the antimonotone property of optimal rule mining relies on the number of counter-examples, we will study the behavior of measures according to this quantity. Similar approaches can be found in [22,9,23,10]. We here give a general framework to study analytic properties of measures.

Definition 1. (*adapted function of measure*) Let call function of measure adapted to a given measure of interest m a couple (\mathcal{D}, ϕ_m) , where $\mathcal{D} \subset \mathbb{Q}^3$ and $\phi_m : \mathcal{D} \rightarrow \mathbb{R}$, verifying:

- (1) \forall association rule $r, \exists X \in \mathcal{D}$ such that $\phi_m(X) = m(r)$;
- (2) $\forall X \in \mathcal{D}, \exists$ an association rule r such that $m(r) = \phi_m(X)$.

The domain \mathcal{D} is then called an adapted domain.

We thus define a surjection from the space of association rules, over all the databases, in \mathcal{D} . The specification of \mathcal{D} allows to make the difference between all the parametrizations: support of examples, support of counterexamples, or confidence of the rule. The adapted function depends of the parametrization. The second condition assures of the usability of \mathcal{D} by assuming that every point in \mathcal{D} corresponds to a real situation.

Example 1. (Recall Measure) $Rec(A \rightarrow B) = \frac{y-x}{z}$ where $x = \mathbb{P}(A \rightarrow B)$, $y = \mathbb{P}(A)$ and $z = \mathbb{P}(B)$. If we study its variations according to z , we find that it increases if $y - x < 0$ and decreases if $y - x > 0$. But constraints of support let us assume that $y > x$, thus Rec decreases with z .

3.2 A Domain for Counterexamples

We present now a domain according to counterexamples. Using the constraints $0 < \mathbb{P}(A), \mathbb{P}(B) < 1, 0 \leq \mathbb{P}(A \rightarrow B) \leq \mathbb{P}(A)$ and $\mathbb{P}(A) - \mathbb{P}(B) \leq \mathbb{P}(A \rightarrow B) \leq \mathbb{P}(\neg B)$, we define the following domain, and prove that it is adapted:

$$\mathcal{D} = \left\{ \left(\begin{matrix} x \\ y \\ z \end{matrix} \right) \in \mathbb{Q}^3 \mid \begin{matrix} 0 & < y < & 1 \\ 0 & < z < & 1 \\ \max\{0, y - z\} \leq x \leq \min\{y, 1 - z\} \end{matrix} \right\}.$$

Table 1. Database for the domain \mathcal{D}

	1	$x \cdot n$		$y \cdot n$		$(z + x) \cdot n$		n
A	1	1	0	0
B	0	0	1	1	0	0

Property 1. Let m be an interestingness measure such that there exists ϕ_m with $m(\mathbf{A} \rightarrow \mathbf{B}) = \phi_m(\mathbb{P}(\mathbf{A} \rightarrow \mathbf{B}), \mathbb{P}(\mathbf{A}), \mathbb{P}(\mathbf{B}))$. Then the couple (\mathcal{D}, ϕ_m) is an adapted function of measure for m .

Proof. The first condition comes immediately from definitions by taking $X = (\mathbb{P}(\mathbf{A} \rightarrow \mathbf{B}), \mathbb{P}(\mathbf{A}), \mathbb{P}(\mathbf{B}))$. For the second one, considering an element (x, y, z) of \mathcal{D} , we need to construct a database \mathcal{DB} containing an association rule $\mathbf{A} \rightarrow \mathbf{B}$, such that $m(\mathbf{A} \rightarrow \mathbf{B})$ equals to $\phi_m(x, y, z)$. We define n as an integer such that $(x \times n, y \times n, z \times n) \in \mathbb{N}^3$. Our database should verify the following equalities $\mathbb{P}(\mathbf{A} \rightarrow \mathbf{B}) = x$, $\mathbb{P}(\mathbf{A}) = y$, $\mathbb{P}(\mathbf{B}) = z$. The constraints on the domain assure that $x \leq y \leq x + z \leq 1$ holds. We can thus construct the database of table 1 satisfying the equalities.

We now benefit of a rigorous formal framework for the study of interestingness measures. We will use it in the following to establish a necessary and sufficient condition (NSC) for existence of an antimonotone property. We thus focus on the parametrization with counterexamples.

4 General Optimal Rule Discovery

4.1 A Condition of Optimotony

From now, we use the same notations as in [20]: \mathbf{P} is an itemset, \mathbf{X} is an item not in \mathbf{P} and \mathbf{c} a class item in a supervised context.

Theorem 1. (*Antimonotone property, [20]*) *If $\text{supp}(\mathbf{PX} \rightarrow \mathbf{c}) = \text{supp}(\mathbf{P} \rightarrow \mathbf{c})$ then rule $\mathbf{PX} \rightarrow \mathbf{c}$ and all its more-specific rules will not occur in an optimal rule set defined by confidence, odds ratio, lift, gain, added-value, Kloggen, conviction, p-s, Laplace, cosine, certainty factor, or Jaccard.*

In addition Li et al. [24] prove that the Relative Risk verifies this property. We introduce the general notion of *optimotony*, the antimonotone property for optimal rule mining, and formulate a NSC for it.

Definition 2. (*optimotony*) *A measure of interest m is optimotone if, given a rule $\mathbf{P} \rightarrow \mathbf{c}$ and a specification $\mathbf{PX} \rightarrow \mathbf{c}$, we have $(\text{supp}(\mathbf{PX} \rightarrow \mathbf{c}) = \text{supp}(\mathbf{P} \rightarrow \mathbf{c})) \implies m(\mathbf{PX} \rightarrow \mathbf{c}) \leq m(\mathbf{P} \rightarrow \mathbf{c})$.*

An optimotone measure can be clearly added to the list of theorem 1.

Theorem 2. (*NSC of optimotony*) *Let m be an interestingness measure, and (\mathcal{D}, ϕ_m) its adapted function of measure. m is optimotone iff ϕ_m increases with the second variable (associated with $\text{supp}(\mathbf{P})$).*

Proof. Sufficiency. Let ϕ_m increases with the second variable, \mathbf{P} be an itemset, \mathbf{c} a class attribute, \mathbf{X} an item $\notin \mathbf{P}$, and $\text{supp}(\mathbf{P} \rightarrow \mathbf{c}) = \text{supp}(\mathbf{PX} \rightarrow \mathbf{c})$, we show

$m(\text{PX} \rightarrow c) \leq m(\text{P} \rightarrow c)$. Since $\text{supp}(\text{P}) \geq \text{supp}(\text{PX})$, we write:

$$\begin{aligned}
 m(\text{PX} \rightarrow c) &= \phi_m(\text{supp}(\text{PX} \neg c), \text{supp}(\text{PX}), \text{supp}(c)) \\
 &\stackrel{\text{hypothesis}}{=} \phi_m(\text{supp}(\text{P} \neg c), \text{supp}(\text{PX}), \text{supp}(c)) \\
 &\stackrel{\text{monotony}}{\leq} \phi_m(\text{supp}(\text{P} \neg c), \text{supp}(\text{P}), \text{supp}(c)) = m(\text{P} \rightarrow c)
 \end{aligned}$$

Necessity. We want to prove that, given two elements (x, y, z) and (x, y', z) of \mathcal{D} such that $y' \leq y$, we have $\phi_m(x, y', z) \leq \phi_m(x, y, z)$. Let n be an integer such that $n \times x$, $n \times y$, $n \times y'$ and $n \times z$ are integers. We construct the database \mathcal{DB} of table 2, in which the supports are the followings: $\text{supp}(\text{P} \neg c) = x = \text{supp}(\text{PX} \neg c)$, $\text{supp}(c) = z$, $\text{supp}(\text{P}) = y$, $\text{supp}(\text{PX}) = y'$. Thus, we do have the equality $\text{supp}(\text{PX} \neg c) = \text{supp}(\text{P} \neg c)$, and deduce, following the hypothesis of implication, that $m(\text{PX} \rightarrow c) \leq m(\text{P} \rightarrow c)$. Since (D, ϕ_m) is a function of measure

Table 2. Database \mathcal{DB}

	1	$(z - (y - x)) \times n$		$(z - (y' - x)) \times n$		$z \times n$		$(z + x) \times n$		n		
P	0	...	0	1	1	0	...	0	
X	0	0	1	1	0	...	0
c	1	1	0	0

Table 3. Expression of measures in \mathcal{D} and relation to optimotony.

measure	$(x, y, z) \in \mathcal{D}$	optimotone	measure	$(x, y, z) \in \mathcal{D}$	optimotone
support	$y - x$	Y	confidence	$1 - \frac{x}{y}$	Y
cover	y	Y	prevalence	z	Y
recall	$\frac{y - x}{z}$	Y	specificity	$\frac{1 - z - x}{1 - y}$	Y
relative specificity	$\frac{1 - z - x}{1 - y} - 1 + y$	Y	precision	$y - 2x + 1 - z$	Y
lift	$\frac{1}{z} - \frac{x}{yz}$	Y	leverage	$1 - \frac{x}{y} - yz$	N
centered confidence	$1 - \frac{x}{y} - z$	Y	relative risk	$1 + \frac{(1 - z) - x}{1 - y}$	Y
Jaccard	$\frac{y - x}{z + x}$	Y	positive confidence	$1 - \frac{x}{y(1 - z)}$	Y
interest	$ y(1 - z) - x $	N	odds ratio	$1 + \frac{y(1 - z) - x}{x(z - y + x)}$	Y
Klogsen	$\sqrt{y - x(1 - z - \frac{x}{y})}$	N	added value	$\max(1 - z - \frac{x}{y}, \frac{y(1 - z) - x}{z})$	Y
conviction	$\frac{y(1 - z)}{x}$	Y	one way support	$(1 - \frac{x}{y}) \log \frac{y - x}{yz}$	N
J ₁ -measure	$(y - x) \log \frac{y - x}{yz}$	N	Piatetsky-Shapiro	$y(1 - z) - x$	Y
cosine	$\frac{y - x}{\sqrt{yz}}$	Y	Loevinger	$1 - \frac{x}{y(1 - z)}$	Y
information gain	$\log \frac{y - x}{yz}$	Y	Sebag Shoenuer	$\frac{y - x}{x}$	Y
contramin	$\frac{y - 2x}{z}$	Y	Bayesian factor	$\frac{(y - x)(1 - z)}{zx}$	Y
example-counter-example rate	$1 - \frac{x}{y - x}$	Y	Zhang	$\frac{y(1 - z) - x}{\max((y - x)(1 - z), zx)}$	Y
Kappa	$2 \frac{y(1 - z) - x}{y + z - 2yz}$	Y	Interestingness Wheighting Dependency	$\left(\left(\frac{y - x}{y - z} \right)^k - 1 \right) \times (y - x)^m$	N

adapted to m , $\phi_m(x, y', z) \leq \phi_m(x, y, z)$. Therefore, ϕ_m does increase with the second variable.

4.2 Classification of the Measures

We thus now have a NSC for the optimonotony. This condition is applied to 32 different measures. The results are presented in table 3. These results show that 26 measures are compatible with the optimonotony, making this pruning strategy very interesting.

Example 2. In table 3 we can notice that, in contrary to the original theorem 1 [20], Klogsen measure K doesn't verify the property. A 3-attributes-database containing the three following transactions: $\{P, X\}$, $\{P, c\}$, $\{c\}$ gives a counterexample. We have $\mathbb{P}(PX \neg c) = \mathbb{P}(P \neg c)$ but $K(P \rightarrow c) = -0.096 < 0$ and $K(PX \rightarrow c) = 0$ i.e. $K(PX \rightarrow c) \not\leq K(P \rightarrow c)$.

5 Conclusion

Measures may possess many algorithmic properties, that have not been exploited. We here propose a formal framework for the analytical study of measures and use it to study the optimonotony property we define. We then demonstrate a NSC for optimonotony. Finally we show that many measures are optimonotone and thus could be used with the underlying pruning strategy.

References

1. Agrawal, R., Imieliski, T., Swami, A.: Mining association rules between sets of items in large databases. In: ICMD, pp. 207–216 (1993)
2. Liu, B., Hsu, W., Ma, Y.: Integrating classification and association rule mining. In: KDD, pp. 80–86 (1998)
3. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: VLDB, pp. 478–499 (1994)
4. Han, J., Pei, J., Yin, Y.: Mining frequent patterns without candidate generation. In: ICMD, pp. 1–12 (2000)
5. Goethals, B.: Frequent set mining. In: The Data Mining and Knowledge Discovery Handbook, pp. 377–397 (2005)
6. Cohen, E., Datar, M., Fujiwara, S., Gionis, A., Indyk, P., Motwani, R., Ullman, J.D., Yang, C.: Finding interesting associations without support pruning. In: TKDE, pp. 64–78 (2001)
7. Koh, Y.S.: Mining non-coincidental rules without a user defined support threshold. In: Washio, T., Suzuki, E., Ting, K.M., Inokuchi, A. (eds.) PAKDD 2008. LNCS, vol. 5012, pp. 910–915. Springer, Heidelberg (2008)
8. Tan, P.N., Kumar, V., Srivastava, J.: Selecting the right objective measure for association analysis. In: IS, pp. 293–313 (2004)
9. Geng, L., Hamilton, H.J.: Interestingness measures for data mining: A survey. ACM Computing Surveys 38(3, Article 9) (2006)

10. Lenca, P., Meyer, P., Vaillant, B., Lallich, S.: On selecting interestingness measures for association rules: user oriented description and multiple criteria decision aid. In: *EJOR*, pp. 610–626 (2008)
11. Li, W., Han, J., Pei, J.: CMAR: Accurate and efficient classification based on multiple class-association rules. In: *ICDM*, pp. 369–376 (2001)
12. Pei, J., Han, J.: Can we push more constraints into frequent pattern mining? In: *KDD*, pp. 350–354 (2000)
13. Bayardo Jr., R.J., Agrawal, R., Gunopulos, D.: Constraint-based rule mining in large, dense databases. In: *ICDE*, pp. 188–197 (1999)
14. Xiong, H., Tan, P.N., Kumar, V.: Mining strong affinity association patterns in data sets with skewed support distribution. In: *ICDM*, pp. 387–394 (2003)
15. Omiecinski, E.: Alternative interest measures for mining associations in databases. In: *TKDE*, pp. 57–69 (2003)
16. Wang, K., He, Y., Cheung, D.W.: Mining confident rules without support requirement. In: *IKM*, pp. 89–96 (2001)
17. Zimmermann, A., De Raedt, L.: CorClass: Correlated association rule mining for classification. In: Suzuki, E., Arikawa, S. (eds.) *DS 2004*. LNCS, vol. 3245, pp. 60–72. Springer, Heidelberg (2004)
18. Morishita, S., Sese, J.: Transversing itemset lattices with statistical metric pruning. In: *PODS*, pp. 226–236 (2000)
19. Bonchi, F., Lucchese, C.: Pushing tougher constraints in frequent pattern mining. In: Ho, T.-B., Cheung, D., Liu, H. (eds.) *PAKDD 2005*. LNCS, vol. 3518, pp. 114–124. Springer, Heidelberg (2005)
20. Li, J.: On optimal rule discovery. In: *TKDE*, pp. 460–471 (2006)
21. Hébert, C., Crémilleux, B.: A unified view of objective interestingness measures. In: Perner, P. (ed.) *MLDM 2007*. LNCS, vol. 4571, pp. 533–547. Springer, Heidelberg (2007)
22. Lallich, S., Vaillant, B., Lenca, P.: Parametrised measures for the evaluation of association rule interestingness. In: *AMSDA*, pp. 220–229 (2005)
23. Lallich, S., Vaillant, B., Lenca, P.: A probabilistic framework towards the parameterization of association rule interestingness measures. In: *MCAP*, pp. 447–463 (2007)
24. Li, J., Chee Fu, A.W., He, H., Chen, J., Jin, H., McAullay, D., Williams, G., Sparks, R., Kelman, C.: Mining risk patterns in medical data. In: *KDD*, pp. 770–775 (2005)