Semantics-based classification of rule interestingness measures

Julien Blanchard, Fabrice Guillet, Pascale Kuntz

Semantics-based classification of rule interestingness measures

Julien Blanchard, Fabrice Guillet, Pascale Kuntz
KnOwledge & Decision Team (KOD)
LINA (UMR CNRS 6241) – Polytechnic School of Nantes University, France
{julien.blanchard,fabrice.guillet,pascale.kuntz}@univ-nantes.fr

Assessing rules with interestingness measures is the cornerstone of successful applications of association rule discovery. However, as numerous measures may be found in the literature, choosing the measures to be applied for a given application is a difficult task. In this chapter, we present a novel and useful classification of interestingness measures according to three criteria: the subject, the scope, and the nature of the measure. These criteria seem to us essential to grasp the meaning of the measures, and therefore to help the user to choose the ones (s)he wants to apply. Moreover, the classification allows one to compare the rules to closely related concepts such as similarities, implications, and equivalences. Finally, the classification shows that some interesting combinations of the criteria are not satisfied by any index.

Keywords: Association rules, Classification of interestingness measures, Meaning of measures, Subject, Scope, Nature.

INTRODUCTION

Most of association rule mining algorithms are unsupervised algorithms, i.e. they do not need any endogenous variable but search all the valid associations existing in the data. This makes the main interest of association rules, since the algorithms can discover relevant rules that the user didn't even think of beforehand. However, the unsupervised nature of association rules causes their principal drawback too: the number of rules generated increases exponentially with the number of variables. Then a very high number of rules can be extracted even from small datasets.

To help the user to find relevant knowledge in this mass of information, many Rule Interestingness Measures (RIM) have been proposed in the literature. RIMs allow one to assess, sort, and filter the rules according to various points of view. They are often classified into two categories: the subjective (user-oriented) ones and the objective (data-oriented) ones. Subjective RIMs take into account the user's goals and user's beliefs of the data domain (Silberschatz & Tuzhilin, 1996; Padmanabhan & Tuzhilin, 1999; Liu et al., 2000). On the other hand, the objective RIMs do not depend on the user but only on objective criteria such as data cardinalities or rule complexity. In this chapter, we are interested in the objective RIMs. This category is very heterogeneous: one can find both elementary measures based on frequency and sophisticated measures based on probabilistic models, as well as information-theoretic measures or statistical similarity measures. In practice, the use of RIMs is problematic since:

- the RIMs are too numerous, and sometimes redundant (Bayardo & Agrawal, 1999; Tan et al., 2004; Blanchard et al., 2005a; Huynh et al., 2006; Lenca et al., 2007);

- the meanings of the RIMs are often unclear, so that it is hard to know precicely what is measured;
- finally, choosing the RIMs to apply for a given study remains a difficult task for the user.

The main contribution of this chapter is to present a novel and useful classification of RIMs according to three criteria: the subject, the scope, and the nature of the measure. These criteria seem to us essential to grasp the meaning of the RIMs, and therefore to help the user to choose the ones (s)he wants to apply. Moreover, the classification allows one to compare the rules to closely related concepts such as similarities, implications, and equivalences. Finally, the classification shows that some interesting combinations of the criteria are not satisfied by any index.

The remainder of the chapter is organized as follows. In the next section, after introducing the notations, we formalize the concepts of *rule* and *interestingness measure*, and then take inventory of numerous measures traditionally used to assess rules. Section 3 defines the three classification criteria, presents our classification of rule interestingness measures, and describes two original measures that we specifically developed to complement the classification. Section 4 discusses the related works. Finally, we give our conclusion in section 5.

RULES AND INTERESTINGNESS MEASURES

Notations

Table 1. Contingency table for two boolean variables a and b. 0 and 1 refer to true and false.

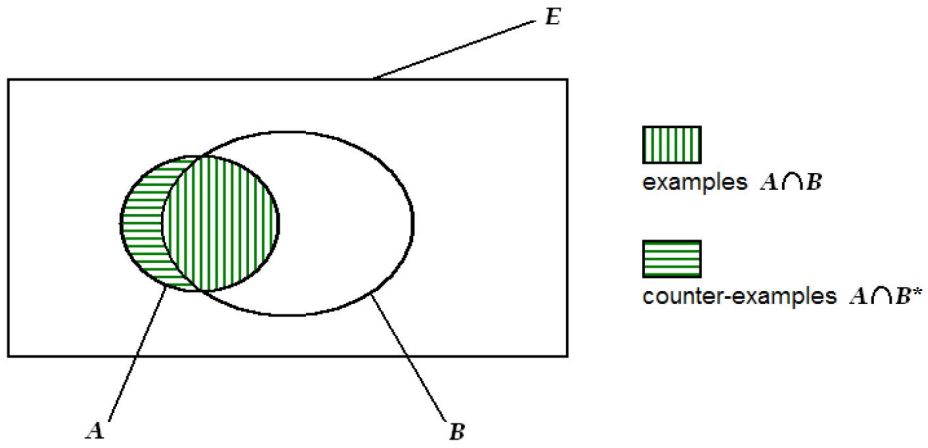
a \ b	1	0	
1	n_{ab}	n_{ab^*}	n_a
0	n_{a^*b}	$n_{a^*b^*}$	n_{a^*}
	n_b	n_{b^*}	n

We consider a set O of n objects described by boolean variables. In the association rule terminology, the objects are transactions stored in a database, the variables are called items, and the conjunctions of variables are called itemsets.

Let a be a boolean variable which is either an itemset, or the negation of an itemset[1]. The variable a^* is the negation of a. We note A the set of objects that verify a, and n_a the cardinality of A. The complementary set of A in O is the set A^* with cardinality n_{a^*}. The probability of the event "*a is true*" is noted $P(a)$. It is estimated by the empirical frequency: $P(a)=n_a/n$.

[1] In general, association rule mining algorithms do not handle negations of items or itemsets. In this chapter, in order to study the meaning of the interestingness measures, we have to consider negations too.

Figure 1. Venn diagram for the rule $a \rightarrow b$.



In the following, we study two boolean variables a and b. The repartition of the n objects in O with regard to a and b is given by the contingency Table 1, where the value n_{ab} is the number of objects that verify both a and b.

Rules

In this chapter, we study the rule interestingness measures as mathematical functions. To do so, we need a general mathematical definition of the concept of *rule* that does not rely on any data mining algorithm[2].

Definition 1. A **rule** is a couple of boolean variables (a, b) noted $a \rightarrow b$. The examples of the rule are the objects which verify the antecedent a and the consequent b, while the counter-examples are the objects which verify a but not b (Figure 1). A rule is better when it has many examples and few counter-examples.

With this definition, an association rule is a special kind of rule. This is simply a rule where a and b are two itemsets which have no items in common.

Rule connections

From two variables a and b, one can build eight different rules:

- $a \rightarrow b$,
- $b \rightarrow a$,
- $a \rightarrow b^*$,
- $b \rightarrow a^*$,
- $a^* \rightarrow b$,
- $b^* \rightarrow a$,
- $a^* \rightarrow b^*$,
- $b^* \rightarrow a^*$.

For a rule $a \rightarrow b$, $a \rightarrow b^*$ is the opposite rule, $b \rightarrow a$ is the converse rule, and $b^* \rightarrow a^*$ is the contrapositive rule.

[2] The concept of *rule* defined here can refer to association rules as well as classification rules. Classification rules are generated for example by induction algorithms such as CN2 (Clark & Boswell, 1991) or decision tree algorithms such as C4.5 (Quinlan, 1993).

Rule modeling

In the same way that the contingency table of two boolean variables is determined by four independent cardinalities, a rule can be modeled with four parameters. Commonly, in the literature, the parameters are n_a, n_b, n and one cardinality of the joint distribution of the two variables such as n_{ab} or n_{ab*}[3]. Like Piatetsky-Shapiro (1991), we choose as fourth parameter the number of examples n_{ab}. So each rule $a \rightarrow b$ is modeled by (n_{ab}, n_a, n_b, n). In the following, we do not differentiate between a rule and its model: $(a \rightarrow b)=(n_{ab}, n_a, n_b, n)$. The set R of all the possible rules is the following subset of $\aleph^4$:

$$R=\{ (n_{ab}, n_a, n_b, n) \mid n_a \leq n, n_b \leq n, max(0, n_a+n_b-n) \leq n_{ab} \leq min(n_a, n_b) \}$$

The choice of the modeling parameters is important since it determines the way of studying rules by inducing a particular point of view for their variations. For example, let us assume that we are interested in the behavior of a rule when n_{ab} varies. If n_{ab} and n_a are among the chosen modeling parameters, then one will tend to fix n_a and therefore to consider that $n_{ab*}=n_a-n_{ab}$ decreases when n_{ab} increases. On the other hand, if n_{ab} and n_{ab*} are among the chosen modeling parameters, then one will tend to fix n_{ab*} and therefore to consider that $n_a=n_{ab}+n_{ab*}$ increases with n_{ab}, which is a totally different scenario. Unfortunately, the choice of the modeling parameters is generally not specified in the literature about rule interestingness (this is an implicit assumption). Few authors have alluded to this problem (see for example (Freitas, 1999)).

Rule interestingness

A huge number of indexes can be found in the literature to assess associations between categorical variables. Before giving a definition of the rule interestingness measures, we explain why some of these indexes are not appropriate to assess rules.

Associations measures between categorical variables

We differentiate two kinds of association measures between two nominal categorical variables (see Table 2):

- the association measures between (possibly) multivalued variables, in which all the values of the variables are treated in the same way;
- the association measures between (necessarily) binary variables, in which the two values are not treated in the same way.

An association measure between two multivalued variables is not altered by a permutation among the values of a variable. On the other hand, an association measure between binary variables is altered. For such a measure M and two binary variables v_1 and v_2, let us denote v_1* and v_2* the variables coming from v_1 and v_2 by permutation of the values (v_1* and v_2* are the negations of v_1 and v_2 in the boolean case). Then we have:

$$M(v_1, v_2) \neq M(v_1, v_2*) \quad \text{and} \quad M(v_1, v_2) \neq M(v_1*, v_2)$$

[3] Certain authors assume that n is constant, therefore they use only three parameters. However, this prevents from comparing rules coming from different datasets.

Table 2. Examples of association measures between two nominal categorical variables.

	symmetrical	**directed**
measures between multivalued variables	χ^2, Cramer's V, Tschuprow's T, mutual information	Goodman and Kruskal's λ and τ, Theil uncertainty coefficient, J-measure, Gini index
measures between binary variables	Jaccard index, Dice index, Kulczynski index, Yule index	confidence, Loevinger index, implication intensity, conviction

As a rule concerns boolean variables, measures between multivalued variables are few appropriate to rule assessment. Indeed, by treating identically all the values, these measures do not differentiate *true* and *false*, and in particular examples and counter-examples. They systematically give the same score to $a \rightarrow b$, $a \rightarrow b^*$, and $a^* \rightarrow b$. However, if $a \rightarrow b$ is a strong rule, then intuitively $a \rightarrow b^*$ and $a^* \rightarrow b$ should be weak ($a \rightarrow b$ and $a \rightarrow b^*$ have opposite meanings). As pointed out by Jaroszewicz and Simovici (2001), a rule should not be assessed on the full joint distribution of the antecedent and consequent.

Interestingness measures

The two basic rule interestingness measures are support and confidence (Agrawal et al., 1993). Support evaluates the generality of the rule; it is the proportion of objects which satisfy the rule in the dataset:

$$support(a \rightarrow b) = n_{ab}/n$$

Confidence evaluates the validity of the rule (success rate); it is the proportion of objects which satisfy the consequent among those which satisfy the antecedent:

$$confidence(a \rightarrow b) = n_{ab}/n_a$$

Support and confidence are simple measures, but they are very commonly used. This populariry can be explained by two major reasons: they are highly intelligible, and they are at the root of the association rule mining algorithms.

Nevertheless, it is now well-known that the support-confidence framework is rather poor to evaluate rule interestingness (Brin et al., 1997a; Bayardo & Agrawal, 1999; Tan et al., 2004). Numerous rule interestingness measures have been proposed to complement this framework (lists can be found for example in (Tan et al., 2004; Geng & Hamilton, 2007)). To unify all the approaches, we propose below a definition of the concept of *rule interestingness measure*.

Definition 2. A **rule interestingness measure (RIM)** is a function $M(n_{ab},n_a,n_b,n)$ from $\mathbb{R}$ to $\mathfrak{R}$ which increases with n_{ab} and decreases with n_a when the other parameters are fixed. The variations are not strict.

In this chapter, we have taken inventory of numerous measures which are traditionally used as RIM. They are listed in the Table 3. The Definition 2 is general enough to include all the measures of the Table 3. The definition is also specific enough to discard all the association measures between multivalued variables (these measures cannot satisfy the variations in the

Table 3. The main RIMs.

Measure M	$M(a \rightarrow b)=$	**References**
confidence	$\frac{n_{ab}}{n_a}$	[Agrawal et al., 1993]
Laplace	$\frac{n_{ab}+1}{n_a+2}$	[Bayardo and Agrawal, 1999] [Clark and Boswell, 1991]
Sebag and Schoenauer index	$\frac{n_{ab}}{n_{ab^*}}$	[Sebag and Schoenauer, 1988]
example and counter-example ratio	$\frac{n_{ab}-n_{ab^*}}{n_{ab}}$	[Huynh et al., 2006]
Ganascia index	$\frac{n_{ab}-n_{ab^*}}{n_a}$	[Ganascia, 1991]
least-contradiction	$\frac{n_{ab}-n_{ab^*}}{n_b}$	[Aze and Kodratoff, 2002]
inclusion index	$\sqrt[4]{I^2_{b/a=1} I^2_{a^*/b=0}}$	[Blanchard et al., 2003]
Loevinger index	$1 - \frac{n n_{ab^*}}{n_a n_{b^*}}$	[Loevinger, 1947]
correlation coefficient	$\frac{n n_{ab}-n_a n_b}{\sqrt{n_a n_b n_{a^*} n_{b^*}}}$	[Pearson, 1896]
rule-interest	$n_{ab} - \frac{n_a n_b}{n}$	[Piatetsky-Shapiro, 1991]
novelty (leverage)	$\frac{n_{ab}}{n} - \frac{n_a n_b}{n^2}$	[Lavrac et al., 1999]
lift (interest)	$\frac{n n_{ab}}{n_a n_b}$	[Brin et al., 1997a]
conviction	$\frac{n_a n_{b^*}}{n n_{ab^*}}$	[Brin et al., 1997b]
collective strength	$\frac{n_{ab}+n_{a^*b^*}}{n_a n_b + n_{a^*} n_{b^*}} \frac{n^2 - n_a n_b - n_{a^*} n_{b^*}}{n - n_{ab} - n_{a^*b^*}}$	[Aggarwal and Yu, 2001]
Yule index	$\frac{n_{ab} n_{a^*b^*} - n_{ab^*} n_{a^*b}}{n_{ab} n_{a^*b^*} + n_{ab^*} n_{a^*b}}$	[Yule, 1900]
odds ratio	$\frac{n_{ab} n_{a^*b^*}}{n_{ab^*} n_{a^*b}}$	[Mosteller, 1968]
Bayes factor	$\frac{n_{ab} n_{b^*}}{n_{ab^*} n_b}$	[Jeffreys, 1935]
κ	$\frac{n n_{ab} + n n_{a^*b^*} - n_a n_b - n_{a^*} n_{b^*}}{n^2 - n_a n_b - n_{a^*} n_{b^*}}$	[Cohen, 1960]
implication intensity	$\mathrm{P}(Poisson(\frac{n_a n_{b^*}}{n}) > n_{ab^*})$	[Gras and Kuntz, 2008]
likelihood linkage index	$\mathrm{P}(Poisson(\frac{n_a n_b}{n}) < n_{ab})$	[Lerman, 1993]
implication index	$-\frac{n_{ab^*} - \frac{n_a n_{b^*}}{n}}{\sqrt{\frac{n_a n_{b^*}}{n}}}$	[Gras and Kuntz, 2008]
directed contribution to χ^2	$\frac{n_{ab} - \frac{n_a n_b}{n}}{\sqrt{\frac{n_a n_b}{n}}}$	[Lerman, 1993]
support (Russel-Rao index)	$\frac{n_{ab}}{n}$	[Agrawal et al., 1993] [Russel and Rao, 1940]
causal support (Sokal-Michener index)	$\frac{n_{ab}+n_{a^*b^*}}{n}$	[Kodratoff, 2001] [Sokal and Michener, 1958]
Rogers-Tanimoto index	$\frac{n - n_{ab^*} - n_{a^*b}}{n + n_{ab^*} + n_{a^*b}}$	[Rogers and Tanimoto, 1960]
Jaccard index	$\frac{n_{ab}}{n - n_{a^*b^*}}$	[Jaccard, 1901]
Dice index	$\frac{n_{ab}}{n_{ab} + \frac{1}{2}(n_{ab^*} + n_{a^*b})}$	[Dice, 1945]
Ochiai index (cosine)	$\frac{n_{ab}}{\sqrt{n_a n_b}}$	[Ochiai, 1957]
Kulczynski index	$\frac{1}{2}(\frac{n_{ab}}{n_a} + \frac{n_{ab}}{n_b})$	[Kulczynski, 1927]

definition because of their symmetries). Information-theoretic measures are not studied in this chapter because they are generally association measures between multivalued variables. The reader can refer to (Blanchard et al., 2005b) for a specific survey on this kind of measures.

The variations of a RIM with regard to n_{ab} and n_a were originally mentioned by Piatetsky-Shapiro (1991) as desirable features of a measure. Here we consider them as the foundations of the concept of RIM. Piatetsky-Shapiro considers that a good measure has to decrease with n_b too, but this requirement is too restricting to appear in a general definition of RIM. More precisely, with regard to n_b or n, RIMs have no particular behavior: some increase, others decrease, and others do not depend on these parameters.

Comparison to similarity measures

Similarity measures are indexes used in data analysis to study objects described by binary variables. They allow one to assess the likeness between two objects or two variables. Lerman gives the following definition for similarity measures (Lerman, 1981).

Definition 3. We note S the following subset of $\aleph^4$: $\mathsf{S}=\{ (n_{ab}, n_{ab*}, n_{a*b}, n) \mid n_{ab}+n_{ab*}+n_{a*b} \leq n \}$. A **similarity measure** is a function $S(n_{ab},n_{ab*},n_{a*b},n)$ from S to $\Re$ which is positive, symmetrical with n_{ab*} and n_{a*b}, increases with n_{ab} and decreases with n_{ab*} when the other parameters are fixed. The variations are strict.

Within the Table 3, the similarity measures are the indexes of Russel and Rao (support), Sokal and Michener (causal support), Rogers and Tanimoto, Jaccard, Dice, Ochiai, and Kulczynski. Below we prove that a similarity measure is a RIM.

Proof 1. Let S be a similarity measure. Given $(n_{ab}, n_a, n_b, n) \in \mathsf{R}$, we have $(n_{ab}, n_a-n_{ab}, n_b-n_{ab}, n) \in \mathsf{S}$. Thus we can define the following function I from R to $\Re$:

$$\forall (n_{ab},n_a,n_b,n) \in \mathsf{R},\ I(n_{ab},n_a,n_b,n)=S(n_{ab},n_a-n_{ab},n_b-n_{ab},n)$$

The function I is a RIM if it increases with n_{ab} and decreases with n_a when the other parameters are fixed. Let us make n_{ab} increase while n_a, n_b, and n are fixed. n_a-n_{ab} and n_b-n_{ab} decrease. As S increases with its first parameter and decreases with its second and third parameters, we conclude that I increases. Let us make n_a increase while n_{ab}, n_b, and n are fixed. S decreases with its second parameter, so I decreases. □

On the other hand, a RIM is not systematically a similarity measure, even if it is positive and symmetrical with a and b. For example, the lift can decrease when n_{ab} increases while n_{ab*}, n_{a*b}, and n are fixed.

CLASSIFICATION OF INTERESTINGNESS MEASURES

In this section, we present an original classification of RIMs according to three criteria: the subject, the scope, and the nature of the measure. In brief, the subject is the notion measured by the index, the scope is the entity concerned by the result of the measure, and the nature is the descriptive or statistical feature of the index. These criteria seem to us essential to grasp the meaning of the measures, and therefore to help the user to choose the ones (s)he wants to apply.

Subject of a RIM

An rule is better when it has many examples and few counter-examples (Definition 1). Thus, given the cardinalities n_a, n_b, and n, the interestingness of $a \rightarrow b$ is maximal when $n_{ab}=min(n_a, n_b)$ and minimal when $n_{ab}=max(0, n_a+n_b-n)$. Between these extreme situations, there exist two significant configurations in which the rules appear non-directed relations and therefore can be considered as neutral or non-existing: the independence and the equilibrium. If a rule is in this configuration, then it must be discarded.

Independence

Table 4. Cardinalities at independence between a and b.

a \ b	1	0	
1	$\frac{n_a \times n_b}{n}$	$\frac{n_a \times n_{b^*}}{n}$	n_a
0	$\frac{n_{a^*} \times n_b}{n}$	$\frac{n_{a^*} \times n_{b^*}}{n}$	n_{a^*}
	n_b	n_{b^*}	n

The binary variables a and b are independent iff $P(a \cap b)=P(a) \times P(b)$, i.e. $\boldsymbol{n.n_{ab}=n_a n_b}$. In this case, each variable gives no information about the other, since knowing the value taken by one of the variables does not alter the probability distribution of the other variable: $P(b \setminus a)=P(b \setminus a^*)=P(b)$ and $P(b^* \setminus a)=P(b^* \setminus a^*)=P(b^*)$ (same for the probabilities of a and a^* given b or b^*). In other words, knowing the value taken by a variable lets our uncertainty about the other variable intact.

Given two variables a and b, there exists only one independence situation, common to the eight rules $a \rightarrow b$, $a \rightarrow b^*$, $a^* \rightarrow b$, $a^* \rightarrow b^*$, $b \rightarrow a$, $b \rightarrow a^*$, $b^* \rightarrow a$, and $b^* \rightarrow a^*$. The contingency table of two independent variables is given in Table 4. There are two ways of deviating from the independence:

- either the variables a and b are positively correlated ($P(a \cap b) > P(a) \times P(b)$), and the four rules $a \rightarrow b$, $a^* \rightarrow b^*$, $b \rightarrow a$, and $b^* \rightarrow a^*$ appear in data (Table 5);
- or they are negatively correlated ($P(a \cap b) < P(a) \times P(b)$), and the opposite four rules appear in data: $a \rightarrow b^*$, $a^* \rightarrow b$, $b \rightarrow a^*$, and $b^* \rightarrow a$ (Table 6).

This dichotomy between the rules and their opposites is due to the fact that two opposite rules are contravariant, since the examples of the one are the counter-examples of the other and vice versa.

Table 5. Positive correlation between a and b with a degree of freedom Δ>0.

a \ b	1	0	
1	$\frac{n_a \times n_b}{n} + \Delta$	$\frac{n_a \times n_{b^*}}{n} - \Delta$	n_a
0	$\frac{n_{a^*} \times n_b}{n} - \Delta$	$\frac{n_{a^*} \times n_{b^*}}{n} + \Delta$	n_{a^*}
	n_b	n_{b^*}	n

Table 6. Negative correlation between a and b with a degree of freedom Δ>0.

a \ b	1	0	
1	$\frac{n_a \times n_b}{n} - \Delta$	$\frac{n_a \times n_{b^*}}{n} + \Delta$	n_a
0	$\frac{n_{a^*} \times n_b}{n} + \Delta$	$\frac{n_{a^*} \times n_{b^*}}{n} - \Delta$	n_{a^*}
	n_b	n_{b^*}	n

Equilibrium

We define the equilibrium of a rule $a \rightarrow b$ as the situation where examples and counter-examples are equal in numbers: $n_{ab}=n_{ab^*}=n_a/2$ (Blanchard et al., 2005a). This corresponds to the maximum uncertainty of *b* given that *a* is true. In this situation, the event *a=1* is as concomitant with *b=1* as with *b=0* in the data. So a rule $a \rightarrow b$ at equilibrium is as directed towards *b* as towards *b**.

Table 7. Cardinalities at the equilibrium of b with regard to a=1.

a \ b	1	0	
1	$\frac{n_a}{2}$	$\frac{n_a}{2}$	n_a
	n_{ab}	n_{ab^*}	n_a

Table 8. Imbalance of b in favor of b=1 with regard to a=1 (degree of freedom Δ>0).

a \ b	1	0	
1	$\frac{n_a}{2}+\Delta$	$\frac{n_a}{2}-\Delta$	n_a
	n_{ab}	n_{ab^*}	n_a

Table 9. Imbalance of b in favor of b=0 with regard to a=1 (degree of freedom Δ>0).

a \ b	1	0	
1	$\frac{n_a}{2}-\Delta$	$\frac{n_a}{2}+\Delta$	n_a
	n_{ab}	n_{ab^*}	n_a

The equilibrium $n_{ab}=n_{ab^*}$ is not defined for the two variables *a* and *b* but for the variable *b* with regard to the literal *a=1*. Thus, with two variables, there exist four different equilibriums, each being common to two opposite rules. Table 7 gives the cardinalities for the equilibrium of *b* with regard to *a=1* under the form of a contingency half-table. There are two ways of deviating from the equilibrium:

- either *a=1* is more concomitant with *b=1* than with *b=0*, and the rule $a \rightarrow b$ appears in data (Table 8);
- or *a=1* is more concomitant with *b=0* than with *b=1*, and the opposite rule $a \rightarrow b^*$ appears in data (Table 9).

Deviations from independence and equilibrium

The fact that there exist two different notions of neutrality for the rules proves that rule interestingness must be assessed from (at least) two complementary points of view: the deviation from independence and the deviation from equilibrium (Blanchard, 2005). These deviations are directed in favor of examples and in disfavor of counter-examples.

Definition 4. The subject of a RIM *M* is the **deviation from independence** iff the measure has a fixed value at independence:

$$M(n_a n_b/n, n_a, n_b, n) = constant$$

Definition 5. The subject of a RIM *M* is the **deviation from equilibrium** iff the measure has a fixed value at equilibrium:

$$M(n_a/2, n_a, n_b, n) = constant$$

Any rule such as $M(a \rightarrow b) \leq constant$ has to be discarded. Also the constant values can be used as reference for setting thresholds to filter the rules. Relevant thresholds are above the constant value.

Regarding the deviation from independence, a rule $a \rightarrow b$ with a good deviation means:

"When a is true, then b is more often true."
(more than usual, i.e. more than without any information about a)

On the other hand, regarding the deviation from equilibrium, a rule $a \rightarrow b$ with a good deviation means:

"When a is true, then b is very often true."

Deviation from independence is a comparison relatively to an expected situation (characterized by n_b), whereas deviation from equilibrium is an absolute statement. From a general point of view, the measures of deviation from independence are useful to discover relations between a and b (do the truth of a influence the truth of b?), while the measures of deviation from equilibrium are useful to take decisions or make predictions about b (knowing or assuming that a is true, is b true or false?).

Example. Let us consider a rule *smoking* → *cancer* assessed by means of confidence and lift. Its deviation from equilibrium is measured by a 30% confidence, which means that 30% of smokers have cancer; its deviation from independence is measured by a lift of 10, which means that smoking increases the risk of cancer by a factor of 10. A smoker that wants to know whether (s)he could have cancer is more interested in deviation from equilibrium. On the contrary, somebody that does not smoke but hesitates to start is more interested in deviation from independence. □

Independence is defined by means of the four parameters n_{ab}, n_a, n_b and n, whereas equilibrium is defined only by means of the two parameters n_{ab} and n_a. Thus, all the measures of deviation from independence depend on the four parameters, whereas there is no reason for any measure of deviation from equilibrium to depend on n_b or n. To the best of our knowledge, the only exceptions to this principle are the inclusion index (Blanchard et al., 2003) and least-contradiction (Aze & Kodratoff, 2002):

- The inclusion index depends on n_b and n because it combines two measures. One is for the direct rule (function of n_{ab} and n_a), and the other is for the contrapositive rule (function of n_{a*b*} and n_{b*}, i.e. n-n_a-n_b+n_{ab} and n-n_b). This is the contribution of the contrapositive which introduces a dependency regarding n_b and n.
- The least-contradiction depends on n_b. This is an hybrid measure which has a fixed value at equilibrium thanks to its numerator –as the measures of deviation from equilibrium– but decreases with n_b thanks to its denominator –as the measures of deviation from independence.

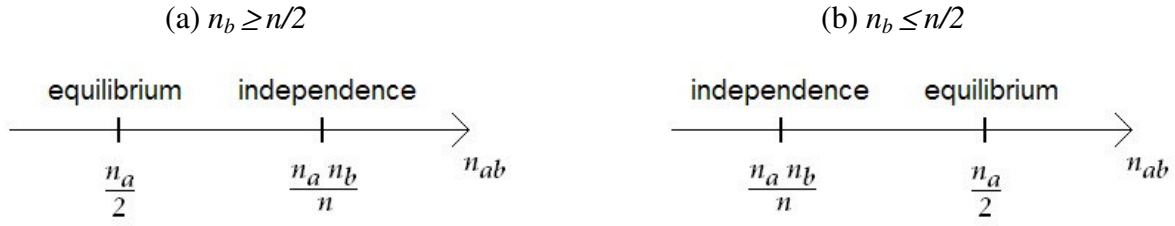
Some RIMs evaluate neither the deviation from equilibrium, nor the deviation from independence. These measures are the similarity indexes of Russel and Rao (support), Sokal and Michener (causal support), Rogers and Tanimoto, Jaccard, Dice, Ochiai, and Kulczynski. They generally have a fixed value only for the rules with no counter-examples ($n_{ab*}=0$) or for the rules with no examples ($n_{ab}=0$). In our classification, we have to create a third class for the similarity measures. They can have various meanings:

- support evaluates the generality/specificity of the rule;
- causal support can also be considered as a measure of generality/specificity, but for a rule and its contrapositive;
- Ochiai index is the geometric mean of the confidences of $a \rightarrow b$ and $b \rightarrow a$;
- Kulczynski index is the arithmetic mean of the confidences of $a \rightarrow b$ and $b \rightarrow a$.

The signification of the other similarity measures cannot be easily expressed in terms of rules.

Preorder comparison

Figure 2. Two possible cases for equilibrium and independence.



Let M_{idp} and M_{eql} be two RIMs which measure the deviations from independence and equilibrium respectively. The fixed values at independence and equilibrium are noted v_{idp} and v_{eql}:

$$M_{idp}(n_a n_b/n, n_a, n_b, n) = v_{idp} \qquad (1)$$

$$M_{eql}(n_a/2, n_a, n_b, n) = v_{eql} \qquad (2)$$

Here we want to exhibit two rules r_1 and r_2 which are differently ordered by M_{idp} and M_{eql}, i.e. $M_{idp}(r_1) \leq M_{idp}(r_2)$ and $M_{eql}(r_1) \geq M_{eql}(r_2)$. To do so, we present below two categories of rules which are always ordered differently.

Let us consider a rule (n_{ab}, n_a, n_b, n). By varying n_{ab} with fixed n_a, n_b, and n, one can distinguish two different cases (see Figure 2):

- If $n_b \geq n/2$ (case 1), then $n_a n_b/n \geq n_a/2$, so the rule goes through the equilibrium before going through the independence when n_{ab} increases.
- If $n_b \leq n/2$ (case 2), then $n_a n_b/n \leq n_a/2$, so the rule goes through the independence before going through the equilibrium when n_{ab} increases.

Let us assume that n_{ab} is between $n_a/2$ and $n_a n_b/n$. The rule is between equilibrium and independence. More precisely:

- In case 1, we have $n_a/2 \leq n_{ab} \leq n_a n_b/n$. Since a RIM increases with n_{ab} when the other parameters are fixed, from (1) and (2) we get that

  $$M_{idp}(n_{ab}, n_a, n_b, n) \leq v_{idp} \text{ and } M_{eql}(n_{ab}, n_a, n_b, n) \geq v_{eql} \qquad (3)$$

  The rule is to be discarded according to its deviation from independence, but it is acceptable according to its deviation from equilibrium.
- In case 2, we have $n_a n_b/n \leq n_{ab} \leq n_a/2$. In the same way, from (1) and (2) we get that

  $$M_{idp}(n_{ab}, n_a, n_b, n) \geq v_{idp} \text{ and } M_{eql}(n_{ab}, n_a, n_b, n) \leq v_{eql} \qquad (4)$$

  The rule is to be discarded according to its deviation from equilibrium, but it is acceptable according to its deviation from independence.

Thus, to exhibit two rules r_1 and r_2 which are ordered differently by M_{idp} and M_{eql}, one only needs to choose r_1 between equilibrium and independence in case 1 and r_2 between equilibrium and independence in case 2, i.e.:

$$r_1 = (n_{ab1}, n_{a1}, n_{b1}, n_1) \text{ with } n_{a1}/2 \leq n_{ab1} \leq n_{a1} n_{b1}/n_1 \text{ and } n_b \geq n_1/2$$

$$r_2 = (n_{ab2}, n_{a2}, n_{b2}, n_2) \text{ with } n_{a2} n_{b2}/n_2 \leq n_{ab2} \leq n_{a2}/2 \text{ and } n_{b2} \leq n_2/2$$

($n_1 = n_2$ if one wants to choose two rules coming from the same dataset)

The inequalities (3) and (4) applied to r_1 and r_2 respectively lead to:

$$M_{idp}(r_1) \leq v_{idp} \leq M_{idp}(r_2) \quad \text{and} \quad M_{eql}(r_2) \leq v_{eql} \leq M_{eql}(r_1)$$

Example. Let us consider the rules r_1=(800, 1000, 4500, 5000) and r_2=(400, 1000, 1000, 5000) assessed by means of confidence (measure of deviation from equilibrium) and lift (measure of deviation from independence). Confidence is worth 0.5 at equilibrium and lift is worth 1 at independence. We obtain

$$confidence(r_1)=0.8 \quad \text{and} \quad lift(r_1)=0.9$$

$$confidence(r_2)=0.4 \quad \text{and} \quad lift(r_2)=2.$$

The rule r_1 is good according to confidence but bad according to lift, whereas the rule r_2 is good according to lift but bad according to confidence. We do have that $confidence(r_1) \geq confidence(r_2)$ and $lift(r_1) \leq lift(r_2)$. □

A measure of deviation from independence and a measure of deviation from equilibrium do not create the same preorder on R. This demonstrates that a RIM (unless being constant) cannot measure both the deviations from independence and equilibrium. This also confirms that the two deviations are two different aspects of rule interestingness. A rule can have a good deviation from equilibrium with a bad deviation from independence, and vice versa. Surprisingly, even if this idea underlies various works about association rules, it seems that it has never been claimed clearly that rule objective interestingness lies in the two deviations.

Comparison of the filtering capacities

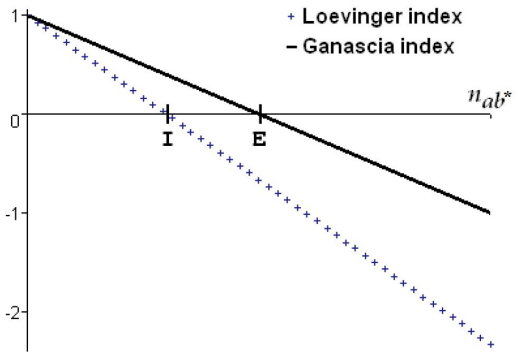
RIMs can be used to filter rules by discarding those that do not satisfy a minimal threshold. We now compare the filtering capacities of a measure of deviation from equilibrium M_{eql} and of a measure of deviation from independence M_{idp}. For the comparison to be fair, we assume that the two measures have similar behaviors: same value for zero counter-examples, same value for equilibrium/independence, same decrease speed with regard to the counter-examples. For example, M_{eql} and M_{idp} can be the Ganascia and Loevinger indexes (cf. the definitions in Table 3).

Let us consider the cases 1 and 2 introduced in the previous section. As shown in Figure 3, M_{idp} is more filtering than M_{eql} in case 1, whereas M_{eql} is more filtering than M_{idp} in case 2. In other words, in case 1, it is M_{idp} which contributes to rejecting the bad rules, while in case 2 it is M_{eql}. This shows that the measures of deviations from equilibrium and from independence have to be regarded as complementary, the second ones not being systematically "better" than the first ones[4]. In particular, the measures of deviation from equilibrium must not be neglected when the realizations of the studied variables are rare. Indeed, in this situation, should the user not take an interest in the rules having non-realizations (which is confirmed in practice), case 2 is more frequent than case 1.

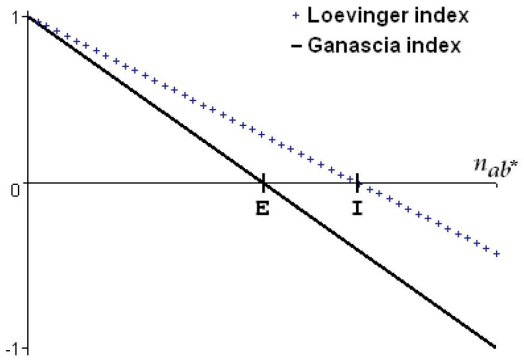
[4] Numerous authors consider that a good interestingness measure must vanish at independence (principle P1 originally proposed in (Piatetsky-Shapiro, 1991), see the "Related Works" section). This principle totally denies the deviation from equilibrium. It amounts to say that measures of deviation from independence are better.

Figure 3. Comparison of Ganascia and Loevinger indexes. (E: equilibrium, I: independence)

(a) case 1 ($n_b \geq n/2$)



(b) case 2 ($n_b \leq n/2$)



Scope of a RIM

Quasi-implication

At first sight, one can think that a rule is an approximation of the logical implication (also called material implication) which accepts counter-examples. However, rules and implications are actually not so similar. This can be seen by comparing the Tables 10.(a) and (b), which are the contingency table of the rule $a \rightarrow b$ and the truth table of the logical implication $a \supset b$.

Table 10. Comparison of a rule to the logical implication.

(a) Contingency table of the rule $a \rightarrow b$.

a \ b	**1**	**0**	
1	examples n_{ab}	counter-examples n_{ab^*}	n_a
0	n_{a^*b}	$n_{a^*b^*}$	n_{a^*}
	n_b	n_{b^*}	n

(b) Truth table of the logical implication $a \supset b$.

a \ b	**1**	**0**
1	1	0
0	1	1

The cases with the same role for the rule and the implication are the cases (*a=1* and *b=1*) and (*a=1* and *b=0*): the first ones satisfy the rule and the implication, while the second ones contradict them. On the other hand, the cases (*a=0* and *b=1*) and (*a=0* and *b=0*) do not play the same role for $a \rightarrow b$ and $a \supset b$: they satisfy the implication but are not examples of the rule. Actually, a rule only conveys the tendency of the consequent to be true when the antecedent is true. The fact that cases (*a=0* and *b=1*) and (*a=0* and *b=0*) satisfy the implication is often considered as a paradox of logical implication (a false antecedent can imply any consequent). Paradoxes like this one have motivated the development of nonclassical logics aiming at

representing the common sense "logic" more faithfully. In these logics, the implication does not give rise (or gives less rise) to counter-intuitive statements. These are the modal logics, the conditional logics, and the relevant logics (Dunn, 1986).

A logical implication $a \supset b$ is equivalent to its contrapositive $b^* \supset a^*$. Thus, the following deductions are possible:

- either by affirming the antecedent a (*Modus ponens*) –the direct form of the implication is used,
- or by denying the consequent b (*Modus tollens*) –the contrapositive of the implication is used.

On the other hand, in the general case, a rule $a \rightarrow b$ is not equivalent to its contrapositive $b^* \rightarrow a^*$: the tendency of the consequent to be true when the antecedent is true is not systematically identical to the tendency of the antecedent to be false when the consequent is false. In particular, some RIMs can measure very different interestingness for a rule and its contrapositive[5]. The user can nevertheless interpret that the meaning of the rule lies both in the direct and contrapositive forms, as for a logical implication. In this case, one can legitimately assess the relation discovered in data not as a rule stricto sensu. More precisely, behind the notation $a \rightarrow b$, Kodratoff (Kodratoff, 2000) distinguishes two types of relations with implicative meaning that can be discovered in data:

- Some associations noted $a \rightarrow b$ express a rule for the user, such as "crows are black" (*crows* $\rightarrow$ *black*). They must be invalidated each time that ($a=1$ and $b=0$) is observed, and validated each time that ($a=1$ and $b=1$) is observed. In accordance with Hempel's paradox[6], the cases ($a=0$ and $b=0$) which validate the contrapositive are not taken into account.
- Some associations noted $a \rightarrow b$ express a quasi-implication for the user, such as "smoking causes cancer" (*smoking* $\rightarrow$ *cancer*). Their meaning is more causal. They must be invalidated each time that ($a=1$ and $b=0$) is observed, and validated each time that ($a=1$ and $b=1$) or ($a=0$ and $b=0$) is observed.

We note the quasi-implications $a \Rightarrow b$. By considering the cases ($a=0$ and $b=0$) as examples, a quasi-implication is not a rule stricto sensu since it conveys both the rule $a \rightarrow b$ and the contrapositive $b^* \rightarrow a^*$ (similarly to logical implication).

Definition 6. A **quasi-implication** is a couple of boolean variables (a, b) noted $a \Rightarrow b$. The examples of the quasi-implication are the objects in $A \cap B$ and $A^* \cap B^*$, while the counter-examples are the objects in $A \cap B^*$ (see Table 11). So $a \Rightarrow b$ is equivalent to its contrapositive $b^* \Rightarrow a^*$. A quasi-implication is better when it has many examples and few counter-examples.

Since quasi-implications rather express a causality for the user, it must be possible to use them to do "quasi-deductions" in the direct way or in the contrapositive way. So the measures M used to assess quasi-implications must render the interestingness of the two rules: $M(a \Rightarrow b)=M(a \rightarrow b)=M(b^* \rightarrow a^*)$.

---

[5] For example with $(n_{ab},n_a,n_b,n)=(750; 800; 900; 1000)$, one have a *confidence*$(a \rightarrow b)=93\%$ and *confidence*$(b^* \rightarrow a^*)=50\%$.

[6] Hempel's paradox lies in the fact that a statement such as "All the crows are black" (logically equivalent to "Something not black is not a crow") is validated by the observation of anything which is not black: a white shoe, a traffic jam...

Table 11. Contingency table of the quasi-implication $a \Rightarrow b$.

a \ b	1	0	
1	examples n_{ab}	counter-examples n_{ab^*}	n_a
0	n_{a^*b}	examples $n_{a^*b^*}$	n_{a^*}
	n_b	n_{b^*}	n

Definition 7. The scope of a RIM M is **quasi-implication** iff $M(a \rightarrow b)=M(b^* \rightarrow a^*)$, i.e. $M(n_{ab}, n_a, n_b, n)=M(n-n_a-n_b+n_{ab}, n-n_b, n-n_a, n)$.

In practice, if the user is sure that contrapositive rules are not relevant for the current application, then (s)he should not apply RIMs whose scope is quasi-implication since the resulting scores are interfered by the interestingness of the contrapositive.

Quasi-conjunction

Table 12. Comparison of a quasi-conjunction to the logical conjunction.

(a) Contingency table of the quasi-conjunction $a \leftrightarrow b$.

a \ b	1	0	
1	examples n_{ab}	counter-examples n_{ab^*}	n_a
0	counter-examples n_{a^*b}	$n_{a^*b^*}$	n_{a^*}
	n_b	n_{b^*}	n

(b) Truth table of the logical conjunction $a \wedge b$.

a \ b	1	0
1	1	0
0	0	0

Following the quasi-implication which approximates the logical implication, we define the quasi-conjunction which approximates the logical conjunction[7].

Definition 8. A **quasi-conjunction** is a couple of boolean variables (a, b) noted $a \leftrightarrow b$. The examples of the quasi-conjunction are the objects in $A \cap B$, while the counter-examples are the objects in $A \cap B^*$ and $A^* \cap B$. So $a \leftrightarrow b$ is equivalent to its converse $b \leftrightarrow a$. A quasi-conjunction is better when it has many examples and few counter-examples.

[7] One can think that the counterpart of the implication is not the conjunction but the equivalence. However, it must be recalled that the implication $a \Rightarrow b$ is the disjunction $a^* \vee b$.

The contingency table of the quasi-conjunction $a \leftrightarrow b$ and the truth table of the logical conjunction $a \wedge b$ are given in Tables 12.(a) and (b). As for quasi-implication and its logical counterpart, 75% of the cases play the same role:

- the cases ($a=1$ and $b=1$) satisfy the quasi-conjunction and the conjunction,
- the cases ($a=1$ and $b=0$) and ($a=0$ and $b=1$) contradict the quasi-conjunction and the conjunction,
- only cases ($a=0$ and $b=0$) are considered differently: they contradict the conjunction but have no precise role for the quasi-conjunction.

A quasi-conjunction $a \leftrightarrow b$ conveys both the rule $a \rightarrow b$ and its converse $b \rightarrow a$. Similarly to quasi-implication, we propose to assess quasi-conjunctions with RIMs that render the interestingness of both $a \rightarrow b$ and $b \rightarrow a$.

Definition 9. The scope of a RIM M is **quasi-conjunction** iff $M(a \rightarrow b)=M(b \rightarrow a)$, i.e. $M(n_{ab}, n_a, n_b, n)=M(n_{ab}, n_b, n_a, n)$.

If the user intuitively gives sense to the rule $b \rightarrow a$ when (s)he reads a rule $a \rightarrow b$, then RIMs whose scope is quasi-conjunction should be applied.

The scope of a similarity measure is quasi-conjunction, since a similarity measure is a RIM (see Proof 1) and is symmetrical with a and b. On the other hand, a RIM whose scope is quasi-conjunction (even a positive RIM such as lift) is not a similarity measure according to the Definition 3.

Quasi-equivalence

Table 13. Comparison of a quasi-equivalence to the logical equivalence.

(a) Contingency table of the quasi-equivalence $a \Leftrightarrow b$.

a \ b	**1**	**0**	
1	examples n_{ab}	counter-examples n_{ab^*}	n_a
0	counter-examples n_{a^*b}	examples $n_{a^*b^*}$	n_{a^*}
	n_b	n_{b^*}	n

(b) Truth table of the logical equivalence $a \equiv b$.

a \ b	**1**	**0**
1	1	0
0	0	1

Definition 10. A **quasi-equivalence** is a couple of boolean variables (a, b) noted $a \Leftrightarrow b$. The examples of the quasi-equivalence are the objects in $A \cap B$ and $A^* \cap B^*$, while the counter-examples are the objects in $A \cap B^*$ and $A^* \cap B$. So $a \Leftrightarrow b$ is equivalent to its contrapositive $b^* \Leftrightarrow a^*$ and to its converse $b \Leftrightarrow a$. A quasi-equivalence is better when it has many examples and few counter-examples.

The contingency table of the quasi-equivalence $a \Leftrightarrow b$ and the truth table of the logical equivalence $a \equiv b$ are given in Tables 13.(a) and (b) (see also (Zembowicz & Zytkow, 1996)). The analogy is strong since all the cases play the same role:

- the cases ($a=1$ and $b=1$) and ($a=0$ and $b=0$) satisfy the quasi-equivalence and the equivalence,
- the cases ($a=1$ and $b=0$) and ($a=0$ and $b=1$) contradict the quasi-equivalence and the equivalence.

A quasi-equivalence lies on the four rules $a \rightarrow b$, $b \rightarrow a$, $b^* \rightarrow a^*$, and $a^* \rightarrow b^*$. We propose to assess quasi-equivalence with RIMs that render the interestingness of the four rules.

Definition 11. The scope of a RIM M is **quasi-equivalence** iff $M(a \rightarrow b)=M(b \rightarrow a)=M(b^* \rightarrow a^*)=M(a^* \rightarrow b^*)$, i.e. $M(n_{ab}, n_a, n_b, n)=M(n_{ab}, n_b, n_a, n)=M(n-n_a-n_b+n_{ab}, n-n_b, n-n_a, n)$ $=M(n-n_a-n_b+n_{ab}, n-n_a, n-n_b, n)$

If the scope of a RIM is quasi-equivalence, then it evaluates both a quasi-implication and a quasi-conjunction. The scope of a similarity measure is not necessarily quasi-equivalence. Among the measures of Table 3, only the Sokal-Michener and Rogers-Tanimoto indexes evaluate quasi-equivalence.

Nature of a RIM

Our last classification criterion is the descriptive or statistical nature of RIMs.

Descriptive measures

Definition 12. The nature of a RIM M is **descriptive** (or frequency-based) iff the measure does not vary with cardinality expansion (when all the data cardinalities are increased or decreased in equal proportion), i.e.

$$\forall \alpha > 0,\ M(n_{ab}, n_a, n_b, n)=M(\alpha.n_{ab}, \alpha.n_a, \alpha.n_b, \alpha.n)$$

Descriptive measures take the data contingencies into account only in a relative way (by means of the probabilities $P(a)$, $P(b)$, $P(a \cap b)$) an not in an absolute way (by means of the cardinalities n_a, n_b, n_{ab}).

Statistical measures

Definition 13. The nature of a RIM M is **statistical** iff the measure varies with cardinality expansion.

Statistical measures take into account the size of the phenomena studied. Indeed, a rule is statistically all the more reliable since it is assessed on a large amount of data.

Among the statistical measures, there are some measures based on a probabilistic model. They compare the observed data distribution to an expected distribution. Two probabilistic measures come from the LLA method (likelihood linkage analysis), a method developed by Lerman in 1981 for the hierarchical clustering of variables (Lerman, 1981):

- the likelihood linkage index of Lerman $P(N_{ab} < n_{ab})$ (Lerman, 1993),
- the implication intensity of Gras $P(N_{ab^*} > n_{ab^*})$ (Gras, 1996; Gras & Kuntz, 2008),

where N_{ab} and N_{ab^*} are the random variables for the numbers of examples and counter-examples under the hypothesis H_0 of independence between a and b. These measures respectively quantify the unlikelihood of the greatness of the number of examples n_{ab} and the unlikelihood of the smallness of the number of counter-examples n_{ab^*}, with respect to the hypothesis H_0. Although

they can be seen as the complement to 1 of the p-value of a hypothesis test, the aim is not testing the hypothesis H_0 but actually using it as a reference to evaluate and sort the rules.

Lerman (1981) proposed three possible random models for H_0. According to the model chosen, the random variables N_{ab} and N_{ab*} can be hypergeometric, binomial, or poissonian. To analyze rules, the most appropriate model is the poissonian one which is the most asymmetric. Then the likelihood linkage index *LLI* and implication intensity *II* are:

$$LLI(a \rightarrow b)=P(Poisson(n_a n_b/n) < n_{ab})$$

$$II(a \rightarrow b)=P(Poisson(n_a n_{b*}/n) > n_{ab*})$$

With the poissonian model, *LLI* evaluates the quasi-conjunction $a \leftrightarrow b$ while *II* evaluates the quasi-implication $a \Rightarrow b$. Of course they both measure deviation from independence. As the two measures are probabilities, they have the advantage of referring to an intelligible scale of values (scale of probabilities). This is not the case for many RIMs. Also, *LLI* and *II* facilitate the choice of a threshold for filtering the rules, since the complement to 1 of the threshold has the meaning of the significance level of a hypothesis test (generally in a test, one chooses $\alpha \in \{0.1\%, 1\%, 5\%\}$).

As they are statistical, the probabilistic measures take into account the size of the phenomena studied. However, this is also their main limit: the probabilistic measures have a low discriminating power when the size of the phenomena is large (beyond around 10^4) (Elder & Pregibon, 1996). Indeed, with regard to large cardinalities, even minor deviations can be statistically significant.

Classification

The classification of RIMs according to subject, scope, and nature is given in Table 14 (Blanchard, 2005). Some cells are empty. First, as a similarity index is symmetrical with *a* and *b*, it can evaluate neither a single rule, nor a quasi-implication. Then, there exists no measure of quasi-conjunction or quasi-equivalence for deviation from equilibrium. Such RIMs could be developed, but they would require to combine rules whose equilibriums are not the same. Contrary to independence, the equilibrium of a rule $a \rightarrow b$ is indeed neither the equilibrium of $b \rightarrow a$, nor the one of $b^* \rightarrow a^*$, nor the one of $a^* \rightarrow b^*$. The only RIM which combines different equilibriums (rule and contrapositive) is the inclusion index.

Whereas one generally considers that RIMs are very numerous, or even too numerous, the classification shows that there are actually few measures whose scope is a single rule. In particular, the only measure of deviation from independence whose scope is a rule stricto sensu is Bayes factor (Jeffreys, 1935). Also the classification shows that there is no statistical RIM measuring the deviation from equilibrium.

Table 14. Classification of RIMs.

Scope / Subject	**Rule**	**Quasi-implication**	**Quasi-conjunction**	**Quasi-equivalence**
Deviation from equilibrium	– confidence, –Sebag and Schoenauer index, – example and counter-example ratio, – Laplace[8], – Ganascia index, – least-contradiction	– inclusion index		
Deviation from independence	– Bayes factor	– Loevinger index, – conviction – **implication intensity**, – **implication index**	– lift – **likelihood linkage index**, – **directed contribution to χ^2**	– correlation coefficient, – novelty, – collective strength, – κ, – Yule index, – odds ratio – **rule-interest**
Similarity			– support (Russel and Rao), – Jaccard index, – Dice index, – Ochiai index, – Kulczynski index	– causal support (Sokal and Michener), – Rogers and Tanimoto index

The **nature** of RIMs is given by the font: the measures in **bold** are statistical, while the others are descriptive.

[8] Laplace varies only slightly with cardinality expansion. This is the reason why we classify it among the descriptive measures.

Two original RIMs

Using the previous classification, we were able to identify gaps and propose two novel RIMs with unique features in (Blanchard et al., 2005a) and (Blanchard et al., 2005b).

Probabilistic measure of deviation from equilibrium *IPEE* (Blanchard et al., 2005a)

IPEE evaluates the deviation from equilibrium while having a statistical nature, which is a unique feature for a RIM according to our classification. More precisely, *IPEE* is based on a probabilistic model and measures the statistical significance of the deviation from equilibrium (whereas implication intensity or likelihood linkage index, for example, measure the statistical significance of the deviation from independence). The measure has the advantage of taking into account the size of the phenomena studied, contrary to the other measures of deviation from equilibrium. Experimental studies show that *IPEE* is efficient even to assess rules with no counter-examples, and well adapted to the search for specific rules ("nuggets").

Directed Information Ratio *DIR* (Blanchard et al., 2005b)

Information-theoretic measures are particularly useful to assess rules since they can be interpreted in terms of information. More precisely, as pointed out by Smyth and Goodman (1992), there is an interesting parallel to draw between the use of information theory in communication systems and the use of information theory to evaluate rules. In communication systems, a channel has a high capacity if it can carry a great deal of information from the source to the receiver. As for a rule, the relation is interesting when the antecedent provides a great deal of information about the consequent. The main drawback of information-theoretic measures is that they do not respect the value-based semantics of association rules, i.e. they systematically give the same value to $a \rightarrow b$ and to its opposite $a \rightarrow b^*$. However, if $a \rightarrow b$ is strong, then intuitively $a \rightarrow b^*$ should be weak.

In (Blanchard et al., 2005b), we presented a new RIM based on information theory which respects the value-based semantics of association rules. This new measure named *DIR* is the only RIM which rejects both independence and equilibrium. In other words, with only one fixed threshold *DIR* discards both the rules whose antecedent and consequent are negatively correlated, and the rules which have more counter-examples than examples[9]. Formal and empirical experiments show that *DIR* is a very filtering measure, which is useful for association rule post-processing.

COMPARISON WITH RELATED WORKS

In 1991, Piatetsky-Shapiro has proposed three principles for an objectively "good" RIM *M* (Piatetsky-Shapiro, 1991):

- P1. $M(a \rightarrow b)=0$ if a and b are statistically independent;

[9] According to the classification presented in this chapter, *DIR* is a mesure of deviation from independence (it vanishes at independence). Nevertheless, as *DIR* is always negative or zero at equilibrium, any strictly positive threshold is enough to reject both independence and equilibrium.

- P2. M monotically increases with n_{ab} when all other parameters remain the same;
- P3. M monotically increases with n_a or n_b when all other parameters remain the same[10].

We think that the principles P1 and P3 (concerning n_b) are too restricting. They lead to discard a wide range of RIMs whereas practical experiment can show that these measures are useful for certain applications. For example, it is well-known that confidence is an appropriate RIM to analyze market basket data, and more generally sparse data. Nevertheless, these principles have been used in many later works (Brin et al., 1997a; Freitas, 1999; Yao & Zhong, 1999; Tan & Kumar, 2000; McGarry, 2005).

More recently, more general works have been conducted to study RIM properties with formal or empirical experiments. Bayardo (1999) showed that several RIMs are redundant when n_b is fixed. In this case, using support and confidence is enough to discover the best rules. Tan et al. (2004) compared 20 symmetrical or symmetrized[11] RIMs according to different formal criteria and on synthetic rule sets. The measures show to be sometimes redundant, sometimes conflicting, and none is significantly better than all the others. Closer to our work is the paper of Lenca et al. (2007) who study 20 RIMs according to the eight following formal criteria:

- C1. symmetry with regard to a and b;
- C2. decrease with n_b;
- C3. value at independence;
- C4. value for rules with no counter-example;
- C5. linearity with n_{ab*} around 0^+;
- C6. sensitivity to n;
- C7. easiness to fix a threshold;
- C8. intelligibility.

The criteria C1 and C3 can be retrieved from our classification. In our opinion, the other criteria, although interesting, do not seem closely related to the meaning of the measures, i.e. they do not help to understand what is measured. With these criteria, Lenca et al. partition the 20 RIM in two ways: first they build a formal partition from the matrix measures×criteria, then they build an experimental partition by comparing the measure behaviors on real rule sets. It is interesting to notice that the resulting clusters tend to greatly confirm our classification, especially regarding subject and scope:

- measures with different subjects belong to different clusters;
- measures of quasi-implication and measures of quasi-conjunction/quasi-equivalence belong to different clusters.

Another approach close to our work is the experimental study of Huynh et al. (2006) performed on 36 RIMs. From two datasets, they mine the association rules and then partition the RIMs using their correlations. If the datasets are considered singly, the resulting clusters differ from our classification. On the other hand, if the partitions are merged by intersecting clusters, then the clusters tend to confirm our classification, again especially regarding subject and scope. The comparison of our semantics-based classification to data-based partitions needs to be explored further with numerous datasets. Indeed, the results of data-based approaches depend on

---

[10] This statement is not precise enough since the modeling parameters are not given. For example, we could use (n_{ab}, n_a, n_b, n) as well as $(n_{ab}, n_a, n_b, n_{a*b*})$ or even $(n_{ab}, n_a, n_b, n_{b*})$, which alters the principles. As explained in the section "Rule Modeling", the choice of the modeling parameters is generally an implicit assumption in the literature about rule interestingness.

[11] The measures do not assess a rule stricto sensu since they estimate a rule and its converse identically: $M(a \rightarrow b) = M(b \rightarrow a)$.

the data and on the biases induced by the parameters of the association rule mining algorithms (support threshold, confidence threshold, maximum number of items, considering of item negations). Actually, a formal classification like ours can be seen as a data-based analysis performed on a rule set that would be the unbiased theoretical set R.

CONCLUSION

By defining the notions of *rule* and *rule interestingness measure*, this chapter provides a formal framework to study rules. Within this framework, we are able to compare the rules to closely related concepts such as similarities, implications, and equivalences. Also we make a novel and useful classification of interestingness measures according to three criteria: the subject, the scope, and the nature of the measure.

- The subject is the notion measured by the index. It can be either the deviation from equilibrium, or the deviation from independence, or a similarity. Deviations from equilibrium and from independence are two different but complementary aspects of rule interestingness.
- The scope is the entity concerned by the result of the measure. It can be either a single rule, or a rule and its contrapositive (quasi-implication), or a rule and its converse (quasi-conjunction), or a rule and its contrapositive and converse (quasi-equivalence).
- The nature is the descriptive or statistical feature of the index.

Finally, the classification shows that some interesting combinations of the criteria are not satisfied by any index. Hence we provide two innovative measures specifically developed to complement the classification: the probabilistic measure of deviation from equilibrium *IPEE*, and the directed information ratio *DIR* which rejects both equilibrium and independence.

The subject, scope, and nature seem to us essential to grasp the meaning of rule interestingness measures. Thus, the classification can help the user to choose the measures (s)he wants to apply for a given application. For example, the classification leads to wonder whether the user is interested only in single rules, or whether the contrapositive and converse can make sense. Also it is relevant to question whether the user wants to measure deviations from equilibrium or deviations from independence, or both. Without information from the user, we think that a judicious solution is using together a descriptive measure of deviation from equilibrium, a statistical measure of deviation from equilibrium, a descriptive measure of deviation from independence, and a statistical measure of deviation from independence. According to us, such a quadruplet of indexes allows one to measure four strongly "orthogonal" aspects of rule interestingness.

REFERENCES

Aggarwal, C. C., & Yu, P. S. (2001). Mining associations with the collective strength approach. *IEEE Transactions on Knowledge and Data Engineering*, 13(6):863–873.

Agrawal, R., Imielienski, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 SIGMOD international conference on management of data*, pages 207–216. ACM Press.

Aze, J., & Kodratoff, Y. (2002). A study of the effect of noisy data in rule extraction systems. In Rappl, R. (Ed.), *Proceedings of the Sixteenth European Meeting on Cybernetics and Systems Research (EMCSR'02)*, volume 2, pages 781–786.

Bayardo, R. J., & Agrawal, R. (1999). Mining the most interesting rules. In *Proceedings of ACM KDD'1999*, pages 145–154. ACM Press.

Blanchard, J. (2005). *Un système de visualisation pour l'extraction, l'évaluation, et l'exploration interactives des règles d'association*. PhD thesis, University of Nantes.

Blanchard, J., Guillet, F., Briand, H., & Gras, R. (2005a). Assessing rule interestingness with a probabilistic measure of deviation from equilibrium. In *Proceedings of the 11th international symposium onApplied Stochastic Models and Data Analysis ASMDA-2005*, pages 191–200.

Blanchard, J., Guillet, F., Gras, R., & Briand, H. (2005b). Using information-theoretic measures to assess association rule interestingness. In *Proceedings of the fifth IEEE international conference on data mining ICDM'05*, pages 66–73. IEEE Computer Society.

Blanchard, J., Kuntz, P., Guillet, F., & Gras, R. (2003). Implication intensity: from the basic statistical definition to the entropic version. In *Statistical Data Mining and Knowledge Discovery*, pages 473–485. Chapman & Hall. Chapter 28.

Brin, S., Motwani, R., & Silverstein, C. (1997a). Beyond market baskets: generalizing association rules to correlations. *SIGMOD Record*, 26(2):265–276.

Brin, S., Motwani, R., Ullman, J. D., & Tsur, S. (1997b). Dynamic itemset counting and implication rules for market basket data. *SIGMOD Record*, 26(2):255–264.

Clark, P., & Boswell, R. (1991). Rule induction with CN2: Some recent improvements. In *EWSL'91: Proceedings of the European Working Session on Machine Learning*, pages 151–163. Springer.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, (20):37–46.

Dice, L. (1945). Measures of the amount of ecologic association between species. *Ecology*, (26):297–302.

Dunn, J. M. (1986). Relevance logic and entailment. In Gabbay, D. and Guenthner, F. (Ed.), *Handbook of Philosophical Logic*, volume 3, pages 117–224. Kluwer Academic Publishers.

Elder, J. F., & Pregibon, D. (1996). A statistical perspective on knowledge discovery in databases. In Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. (Ed.), *Advances in knowledge*
discovery and data mining, pages 83–113. AAAI/MIT Press.

Freitas, A. (1999). On rule interestingness measures. *Knowledge-Based Systems Journal*, 12(5-6):309–315.

Ganascia, J. G. (1991). Deriving the learning bias from rule properties. In *Machine intelligence 12: towards an automated logic of human thought*, pages 151–167. Clarendon Press.

Geng, L., & Hamilton, H. J. (2007). Choosing the right lens: Finding what is interesting in data mining. In Guillet, F. and Hamilton, H. J. (Ed.), *Quality Measures in Data Mining*, volume 43 of *Studies in Computational Intelligence*, pages 3–24. Springer.

Gras, R. (1996). *L'implication statistique : nouvelle méthode exploratoire de données*. La Pensée Sauvage Editions.

Gras, R., & Kuntz, P. (2008). An overview of the statistical implicative analysis development. In Gras, R., Suzuki, E., Guillet, F., and Spagnolo, F. (Ed.), *Statistical Implicative Analysis: Theory and Applications*, volume 127 of *Studies in Computational Intelligence*, pages 21–52. Springer.

Huynh, X.-H., Guillet, F., & Briand, H. (2006). Evaluating interestingness measures with linear correlation graph. In *Proceedings of the 19th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (IEA/AIE)*, volume 4031 of *Lecture Notes in Computer Science*, pages 312–321. Springer.

Jaccard, P. (1901). Etude comparative de la distribution florale dans une portion des Alpes et du Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, (37):547–579.

Jaroszewicz, S., & Simovici, D. A. (2001). A general measure of rule interestingness. In *Proceedings of PKDD'2001*, pages 253–265. Springer.

Jeffreys, H. (1935). Some tests of significance treated by the theory of probability. In *Proceedings of the Cambridge Philosophical Society*, pages 203–222.

Kodratoff, Y. (2000). Extraction de connaissances à partir des données et des textes. In *Actes des journées sur la fouille dans les données par la méthode d'analyse statistique implicative*, pages 151–165. Presses de l'Université de Rennes 1.

Kodratoff, Y. (2001). Comparing machine learning and knowledge discovery in databases: an application to knowledge discovery in texts. In Paliouras, G., Karkaletsis, V., and Spyropoulos, C. (Ed.), *Machine Learning and Its Applications*, volume 2049 of *Lecture Notes in Artificial Intelligence*, pages 1–21. Springer.

Kulczynski, S. (1927). Die pflanzenassoziationen der pieninen. *Bulletin International de l'Académie Polonaise des Sciences et des Lettres. Classe des Sciences Mathématiques et Naturelles*, (suppl. II):57–203. série B.

Lavrac, N., Flach, P. A., & Zupan, B. (1999). Rule evaluation measures: a unifying view. In *ILP'99: Proceedings of the ninth International Workshop on Inductive Logic Programming*, pages 174–185. Springer-Verlag.

Lenca, P., Vaillant, B., Meyer, P., & Lallich, S. (2007). Association rule interestingness measures: Experimental and theoretical studies. In Guillet, F. and Hamilton, H. J. (Ed.), *Quality Measures in Data*
Mining, volume 43 of *Studies in Computational Intelligence*, pages 51–76. Springer.

Lerman, I. C. (1981). *Classification et analyse ordinale des données*. Dunod.

Lerman, I. C. (1993). Likelihood linkage analysis (LLA) classification method: an example treated by hand. *Biochimie*, 75(5):379–397.

Liu, B., Hsu, W., Chen, S., & Ma, Y. (2000). Analyzing the subjective interestingness of association rules. *IEEE Intelligent Systems*, 15(5):47–55.

Loevinger, J. (1947). A systematic approach to the construction and evaluation of tests of ability. *Psychological Monographs*, 61(4).

McGarry, K. (2005). A survey of interestingness measures for knowledge discovery. *The Knowledge Engineering Review*, 20(1):39–61.

Mosteller, F. (1968). Association and estimation in contingency tables. *Journal of the American Statistical Association*, 63(321):1–28.

Ochiai, A. (1957). Zoogeographic studies on the soleoid fishes found in japan and its neighbouring regions. *Bulletin of the Japanese Society of Scientific Fisheries*, (22):526–530.

Padmanabhan, B., & Tuzhilin, A. (1999). Unexpectedness as a measure of interestingness in knowledge discovery. *Decision Support Systems*, 27(3):303–318.

Pearson, K. (1896). Mathematical contributions to the theory of evolution: regression, heredity and panmixia. *Philosophical Transactions of the Royal Society Of London*, series A(187):253–318.

Piatetsky-Shapiro, G. (1991). Discovery, analysis, and presentation of strong rules. In *Knowledge Discovery in Databases*, pages 229–248. AAAI/MIT Press.

Quinlan, J. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann.

Rogers, D., & Tanimoto, T. (1960). A computer program for classifying plants. *Science*, (132):1115–1118.

Russel, P., & Rao, T. (1940). On habitat and association of species of anopheline larvae in south-eastern madras. *Journal of the Malaria Institute of India*, (3):153–178.

Sebag, M., & Schoenauer, M. (1988). Generation of rules with certainty and confidence factors from incomplete and incoherent learning bases. In *Proceedings of EKAW88*, pages 28.1–28.20.

Silberschatz, A., & Tuzhilin, A. (1996). User-assisted knowledge discovery: how much should the user be involved. In *Proceedings of the 1996 SIGMOD workshop on research issues on data mining and knowledge discovery (DMKD)*.

Smyth, P., & Goodman, R. M. (1992). An information theoretic approach to rule induction from databases. *IEEE Transactions on Knowledge and Data Engineering*, 4(4):301–316.

Sokal, R., & Michener, C. (1958). A statistical method for evaluating systematic relationships. *University of Kansas Science Bulletin*, (38):1409–1438.

Tan, P.-N., & Kumar, V. (2000). Interestingness measures for association patterns: a perspective. In *Proceedings of the KDD-2000 workshop on postprocessing in machine learning and data mining*.

Tan, P.-N., Kumar, V., & Srivastava, J. (2004). Selecting the right objective measure for association analysis. *Information Systems*, 29(4):293–313.

Yao, Y. Y., & Zhong, N. (1999). An analysis of quantitative measures associated with rules. In *PAKDD'99: Proceedings of the Third Pacific-Asia Conference on Methodologies for Knowledge Discovery and Data Mining*, pages 479–488. Springer.

Yule, G. (1900). On the association of attributes in statistics. *Philosophical Transactions of the Royal Society of London*, series A(194):257–319.

Zembowicz, R., & Zytkow, J. M. (1996). From contingency tables to various forms of knowledge in databases. In *Advances in knowledge discovery and data mining*, pages 328–349. American Association for Artificial Intelligence.