

Research Article

An Extension of Totohasina's Normalization Theory of Quality Measures of Association Rules

Armand,¹ André Totohasina ,² and Daniel Rajaonasy Feno³

¹Lycée Mixte Antsiranana, Antsiranana, Madagascar

²Department of Mathematics and Informatics Application, University of Antsiranana, Madagascar

³Department of Mathematics and Informatics Application, University of Toamasina, Madagascar

Correspondence should be addressed to André Totohasina; andre.totohasina@gmail.com

Received 1 September 2018; Accepted 6 January 2019; Published 29 January 2019

Academic Editor: Theodore E. Simos

Copyright © 2019 Armand et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the context of binary data mining, for unifying view on probabilistic quality measures of association rules, Totohasina's theory of normalization of quality measures of association rules primarily based on affine homeomorphism presents some drawbacks. Indeed, it cannot normalize some interestingness measures which are explained below. This paper presents an extension of it, as a new normalization method based on proper homographic homeomorphism that appears most consequent.

1. Introduction

In the context of the implicative statistical analysis (ISA) [1], you can never leave aside the probabilistic notions of the measures of quality which assess the degree of implicative link between two patterns of association rule. In light of very rich number of measures in the literature of the binary database, researchers are working [2–7] parallelly to find the more general relationships allowing to classify partially or entirely these various measures of interest. Hence the creation of highly founded concept called “normalization under five constraints quality measures” in the context of data mining appeared in 2003 [2]. This definition of normalized measures is acquired. It thus turns out that all M_{GK} normalizable measures by affine homeomorphism become comparable [2, 8]. Although this author has already well-treated this subject, the problem of normalization of probabilistic quality measures remains not fully resolved. Indeed, the tool called affine homeomorphism he used may have some weakness, because it cannot normalize a measure intended to the infinity at least one of reference situations more or less intuitive such as *the incompatibility, the statistical independence, and logical implication* [1, 9–12], for example, the measures *Cost Multiplying, Sebag, Conviction, Odd-Ratio, Informal Gain, and Ratio of Example counter-example*. Therefore, as already announced by the author ([13] paragraph

3.2 page 65) “the problem remains open concerning the transformation allowing normalization of other measures in a way still to be specified”. We are interested in this issue. This article proposes a way to partly solve this problem and discussion. First, let us recall below our proposed definition of a normalized quality measure which is evolving with three intuitive events as incompatibility, stochastic independence, and implication.

Definition 1 (cf. [6, 13–15]). Let X and Y be itemset of a context of binary data mining $D = (T, I, \mathcal{R})$, P the uniform discrete probability on probabilizable space $(T, \mathcal{P}(T))$ [2], T all sets of transactions, I the set of attributes called items or patterns, \mathcal{R} the binary relation from T to I , μ a probabilistic quality measure, and $X \rightarrow Y$ an association rule, with $X \cap Y = \emptyset$. $\forall x \in I$, $X' = \{e \in T / \forall x \in X; e \mathcal{R} x\}$ denoted the extension of the itemset X : X' is hence an event of the discrete probability $(T, \mathcal{P}(T), P)$. A quality measure μ of an association rule is called to be normalized if it verifies the five conditions below:

- (i) $\mu(X \rightarrow Y) = -1$, if $P(Y'/X') = 0$; i.e., X' and Y' are two incompatible events: this means the two patterns X and Y are incompatible then;
- (ii) $-1 < \mu(X \rightarrow Y) < 0$, if $0 < P(Y'/X') < P(Y')$; i.e., X' and Y' are two negatively dependent events,

and as such both X and Y patterns are then negatively dependent;

- (iii) $\mu(X \rightarrow Y) = 0$, if $P(Y'/X') = P(Y')$; i.e., the two events X' and Y' are independent and this means both X and Y are independent itemsets;
- (iv) $0 < \mu(X \rightarrow Y) < 1$, and $0 \neq P(Y'/X') > P(Y')$; i.e., if X' and Y' are positively dependent, therefore the two itemsets X and Y are positively dependent;
- (v) $\mu(X \rightarrow Y) = 1$, and $P(Y'/X') = 1$ or if $X' \subseteq Y'$: the itemset X is then completely included in Y . $X' = \{e \in T / \forall x \in X; e \mathcal{R} x\}$; i.e., the set of all transactions contains the pattern X [6].

Example 2. According to the definition mentioned above, it is easy to show that the M_{GK} measure defined by $M_{GK}(X \rightarrow$

$$Y) = \begin{cases} M_{GK}^f(X \rightarrow Y) \\ M_{GK}^d(X \rightarrow Y) \end{cases}$$

where $M_{GK}^f(X \rightarrow Y) = (P(Y'/X') - P(Y')) / (1 - P(Y'))$, if X favors Y , and $M_{GK}^d(X \rightarrow Y) = (P(Y'/X') - P(Y')) / P(Y')$, if X disfavors Y , is normalized.

Note that this measure was discovered independently by three authors from three continents: by S. Guillaume in France in her thesis [15], by A. Totohasina in [2] in Madagascar at his research on normalization, at that time he appointed ION "Implication Oriented Normalised", and by X. Wu, C. Zhang, and S. Zhang in [5] in USA under the name of CPIR "Conditional Probability Incrementation Ratio". Its rich and interesting mathematical properties are studied in [6, 15]. In addition, it is historically interesting to notice that this measure was partially discovered as *Certainty Factor (CF)* by Edward H. Shortcliffe and Bruce G. Buchanan in USA at 1975 [12]. Fernando Berzal et al. [16] established some statistical properties of *CF* and its relation with some common interestingness measures as Confidence and Conviction.

Hereafter, our work is divided into five sections. Section 2 recalls some properties of a homographic homeomorphism that will be the main subject of our contribution. Section 3 offers a new normalization process based on homographies, in order to solve the aforementioned problem. Section 4 recommends the raw results of each of the normalized measures. Section 5 raises a conclusion.

2. About Homographic Function Processing Tool

2.1. Definition and Reminders. In mathematics, a homographic function is a function which can be represented as the form of quotient of two linear functions. It is bijective and its inverse function is a particular homographic function. In the commutative field \mathbb{R} a homographic function f on \mathbb{R} is a function in itself defined by $f(x) = (ax + b)/(cx + d)$, where a, b, c , and d are real numbers such as $ad - cb \neq 0$. Prohibit $ad - bc$ to be zero to avoid a constant function. Sometimes the condition "c not zero" is added, as the case $c = 0$ corresponds to linear functions, but then we lose the group structure of

the set of homographic functions with the composition of applications.

We will retain that a homographic function is a homeomorphism of the form $\mathbb{R} \setminus \{-d/c\} \rightarrow \mathbb{R} : x \mapsto (ax + b)/(cx + d)$, where $(ad - bc \neq 0)$. This function determines a bijection from $\mathbb{R} \setminus \{-d/c\}$ to $\mathbb{R} \setminus \{a/c\}$ whose inverse bijection $f^{-1}(x) = -(dx - b)/(cx - a)$, which has the same determinant $(ad - bc \neq 0)$. Note that f^{-1} is a homography of the same type as f and the graphs of homography f and f^{-1} are hyperbolas.

It is seen that if we extend f by $f(-d/c) = \infty$ and $f(\infty) = a/c$, we obtain a projective application and let us denote $\overline{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$.

Derivative and Variation. In the real homography case f , its derivative is $f'(x) = \left| \begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right| / (cx + d)^2$, where $\left| \begin{smallmatrix} a & b \\ c & d \end{smallmatrix} \right| = ad - bc$ is the determinant of the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and so is called the determinant of the homography f and denoted $\det(f)$. For this reason, here are the variations in the homographic function: if $\det(f)$ is positive, then f is increasing on its two definition intervals; if $\det(f)$ is negative, then f is strictly decreasing on both definition intervals.

2.2. Canonical Form. Let f be a homography such that $f(x) = (ax + b)/(cx + d)$, with $(ad - bc \neq 0)$.

In case c is not zero, the canonical form (also called reduced form) of a homographic function is $f(x) = \alpha/(x + \beta) + \gamma$, where $\alpha = (bc - ad)/c^2$, $\beta = d/c$, and $\gamma = a/c$. By making a change in reference by taking a point S , the set projective applications, of coordinates (β, γ) as a new origin, the expression of the homographic function becomes $f(x) = \alpha/x$ which corresponds to the inverse function multiplied by the scalar $\alpha = (bc - ad)/c^2$.

Morality. Any own homographic function nonzero determinant can thus be reduced to a homographic function type as $x \mapsto k/(x + m)$, with $k \neq 0$.

From now, we are interested in all homographies of type as $x \mapsto k/(x + m)$, with $k \neq 0$, that $\lim_{x \rightarrow \infty} (k/(x + m)) = 0 \in \mathbb{R}$: own homography thus returns infinity to a real finite. Knowing that, this time, our main objective is to "make the infinity to be finite", we thereafter consider the measures that have infinity value among the three conditions such as *logical implication, stochastic independence, and incompatibility*; we use the homographic function mentioned above. Then, for any situation not leading to infinity, it is relevant to use the theory in [13] which is based on the use of an affine homeomorphism. Taking advantage of the fact that affine applications are part of the great family of homographies and appear as degenerate homographies returning infinity to infinity, we will enhance the whole nonconstant homographies. In the current theory we can combine the theory in [2] and the one we have just proposed. It is thus a natural extension of such approach of [14].

2.3. Notations. For convenience let us denote by x the value of the probabilistic quality measure $\mu(X \rightarrow Y)$, by x_{imp} the value of $\mu(X \rightarrow Y)$, at logical implication, x_{ind} that of $\mu(X \rightarrow Y)$ at independence, and x_{inc} the value of $\mu(X \rightarrow Y)$ at incompatibility, where $(x_{imp}, x_{ind}, x_{inc}) \in \overline{\mathbb{R}}^3$.

3. Normalization Process by Homography

Let μ_{in} be the homography normalization of quality measure μ , μ_{shn} the semihomographic normalized of μ , μ_{shnd} at right semihomographic normalized of μ , and μ_{shng} at left semihomographic normalized of μ .

As announced in [2], the main objective of normalization of quality measure μ is to bring its values in $[-1; 1]$ under the three conditions that μ_n takes the value -1 at incompatibility, 0 at independence, and 1 at logical implication in order to compare two normalizable measures. Remember always that if these three values are finite and pairwise different, so the research carried out by [14] has already taken the approach to solve this kind of problem (problem of normalization of probabilistic quality measure), i.e., the use of the expression of the normalized μ_n of μ :

$$\mu_n(X \longrightarrow Y) = \begin{cases} x_f x + y_f, & \text{if } X \text{ favors } Y; \\ x_d x + y_d, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (1)$$

where these four coefficients, called normalization coefficient, are determined by passing unilateral limits in reference situations (incompatibility, independence, and logical implication) due to the continuity of evolution in both zones: attraction (positive dependence) and repulsion (negative dependence) [2, 6, 17].

If one or two of the three values x_{imp} , x_{ind} , and x_{inc} are infinite and in case we have two infinite values, it is necessary that x_{ind} is excluded, which leads us to use one of the three following expressions to find the four real coefficients, x_f , y_f , x_d , and y_d :

$$\begin{aligned} \mu_{shnd}(X \longrightarrow Y) &= \begin{cases} \frac{x_f}{x+m} + y_f, & \text{if } X \text{ favors } Y; \\ x_d x + y_d, & \text{if } X \text{ disfavors } Y. \end{cases} \\ \mu_{shng}(X \longrightarrow Y) &= \begin{cases} x_f x + y_f, & \text{if } X \text{ favors } Y; \\ \frac{x_d}{x+m} + y_d, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (2) \\ \mu_{in}(X \longrightarrow Y) &= \begin{cases} \frac{x_f}{x+m} + y_f, & \text{if } X \text{ favors } Y; \\ \frac{x_d}{x+m} + y_d, & \text{if } X \text{ disfavors } Y. \end{cases} \end{aligned}$$

These four coefficients are still determined by passing unilateral limits in situations of reference (incompatibility, independence, and logic implication) by taking into account the continuity evolution in both zones: attraction (positive dependence) and repulsion (negative dependence). In case where X favors Y , x_{imp} can be infinite, $x_{ind} \neq x_{inc}$, $(x_{ind}, x_{inc}) \in \mathbb{R}^2$ and $x_{ind} \in \mathbb{R}^*$; then we obtain the following system of equations:

$$\begin{aligned} \frac{x_f}{x_{imp}} + y_f &= 1, & (\text{logical implication}) \\ \frac{x_f}{x_{ind}} + y_f &= 0, & (\text{independence at right}) \end{aligned} \quad (3)$$

As $(x_{ind}, x_{inc}) \in \mathbb{R}^2$ and $x_{ind} \in \mathbb{R}^*$, so you only use the theories in [2] for the left normalization. We can write the system of four nonlinear equations, with the following four unknowns:

$$\begin{aligned} \frac{x_f}{x_{imp}} + y_f &= 1 \\ \frac{x_f}{x_{ind}} + y_f &= 0 \\ x_d x_{ind} + y_d &= 0 \\ x_d x_{ind} + y_d &= -1 \end{aligned} \quad (4)$$

We are just here to take $m = 0$ and we have four equations with four unknowns, with this particularity that the coefficient x_{imp} can be infinite. Hence we have the following proposition.

Proposition 3. (i) If $(x_{imp}, x_{ind}) \in (\mathbb{R}^*)^2$ with x_{imp}, x_{ind} , and x_{inc} pairwise distinct, then the system of equations (4) admits four real solutions; (ii) if $x_{imp} = \infty$ and $x_{ind} \in \mathbb{R}^*$, then the system of equations (4) has four real solutions such that $y_f = 1$, $x_f = -x_{ind}$, $x_d = 1/(x_{ind} - x_{inc})$, and $y_d = -x_d x_{ind}$; (iii) otherwise, the system of equations has no solution.

Proof. (i) The system of equations (4) is equivalent to the system of equations (5):

$$\begin{aligned} x_{imp} y_f + x_f &= x_{imp} \\ x_{ind} y_f + x_f &= 0 \\ x_d x_{ind} + y_d &= 0 \\ x_d x_{inc} + y_d &= -1 \end{aligned} \quad (5)$$

It became a system of linear equations. The matrix writing system of equations (5) is given by the vector equation (6):

$$\begin{pmatrix} x_{imp} & 1 & 0 & 0 \\ x_{ind} & 1 & 0 & 0 \\ 0 & 0 & x_{ind} & 1 \\ 0 & 0 & x_{inc} & 1 \end{pmatrix} \begin{pmatrix} y_f \\ x_f \\ x_d \\ y_d \end{pmatrix} = \begin{pmatrix} x_{imp} \\ 0 \\ 0 \\ -1 \end{pmatrix} \quad (6)$$

Let us call $M_1 = \begin{pmatrix} x_{imp} & 1 & 0 & 0 \\ x_{ind} & 1 & 0 & 0 \\ 0 & 0 & x_{ind} & 1 \\ 0 & 0 & x_{inc} & 1 \end{pmatrix}$. So we must have $\det M_1 = (x_{imp} - x_{ind})(x_{ind} - x_{inc}) \neq 0$; so (i); for (ii), just take $\lim_{x_{imp} \rightarrow \infty} (x_f/x_{imp}) = 0$; for (iii), if $x_{ind} = \infty$, the last two equations do not make sense. \square

The following system of equations presents the common features of equation (5) with the only difference that x_{inc} can be infinite, $x_{imp} \neq x_{ind}$ and $x_{ind} \in \mathbb{R}^*$. This gives the system of equations (7).

$$\begin{aligned} x_f x_{imp} + y_f &= 1 \\ x_f x_{ind} + y_f &= 0 \end{aligned}$$

$$\begin{aligned} \frac{x_d}{x_{ind}} + y_d &= 0 \\ \frac{x_d}{x_{inc}} + y_d &= -1 \end{aligned} \tag{7}$$

Hence we obtain the following proposition.

Proposition 4. (i) If $(x_{ind}, x_{inc}) \in (\mathbb{R}^*)^2$, with x_{imp} , x_{ind} , and x_{inc} pairwise distinct, then the system of equations (7) admits four real solutions; (ii) if $x_{inc} = \infty$ and $x_{ind} \in \mathbb{R}^*$, then the system of equations (7) has four real solutions such that $y_d = -1$, $x_d = x_{ind}$ and $x_f = 1/(x_{imp} - x_{ind})$, $y_f = -x_f x_{ind}$; (iii) otherwise, the system of equations has no solution.

Proof. (i) The system of equations (7) is equivalent to the system of equations (8):

$$\begin{aligned} x_f x_{imp} + y_f &= 1 \\ x_f x_{ind} + y_f &= 0 \\ y_d x_{ind} + x_d &= 0 \\ y_d x_{inc} + x_d &= -x_{inc} \end{aligned} \tag{8}$$

It became a system of linear equations. The matrix writing the system of equations (8) is given by the vector equation (9):

$$\begin{pmatrix} x_{imp} & 1 & 0 & 0 \\ x_{ind} & 1 & 0 & 0 \\ 0 & 0 & x_{ind} & 1 \\ 0 & 0 & x_{inc} & 1 \end{pmatrix} \begin{pmatrix} x_f \\ y_f \\ y_d \\ x_d \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -x_{inc} \end{pmatrix} \tag{9}$$

Let us call $M_2 = \begin{pmatrix} x_{imp} & 1 & 0 & 0 \\ x_{ind} & 1 & 0 & 0 \\ 0 & 0 & x_{ind} & 1 \\ 0 & 0 & x_{inc} & 1 \end{pmatrix}$. So we must have $\det M_2 = (x_{imp} - x_{ind})(x_{ind} - x_{inc}) \neq 0$; so (i); for (ii), just take $\lim_{x_{inc} \rightarrow \infty} (x_d/x_{inc}) = 0$; for (iii), if $x_{ind} = \infty$, then the last two equations do not make sense. \square

The following system is similar to the previous equation (7), this time with $x_{ind} \in \mathbb{R}$ and $m \in \mathbb{R}^*$, including case where $x_{ind} = 0$; then in this case, m must be nonzero. Take, for example, $m = 1$.

$$\begin{aligned} x_f x_{imp} + y_f &= 1 \\ x_f x_{ind} + y_f &= 0 \\ \frac{x_d}{x_{ind} + 1} + y_d &= 0 \\ \frac{x_d}{x_{inc} + 1} + y_d &= -1 \end{aligned} \tag{10}$$

Then we obtain the following proposition.

Proposition 5. (i) If $(x_{ind}, x_{inc}) \in \mathbb{R}^2$, with $x_{imp} \neq 0$, x_{imp} , x_{ind} , and x_{inc} pairwise distinct, then the system of equations

(10) has four real solutions; (ii) if $x_{inc} = \infty$, $x_{ind} \in \mathbb{R}$, and $x_{imp} \in \mathbb{R}^*$, then the system of equations (10) has four real solutions such that $y_d = -1$, $x_d = x_{ind} + 1$ and $x_f = 1/(x_{imp} - x_{ind})$, $y_f = -x_f x_{ind}$; (iii) then, the system of equations has no solution.

Proof. (i) The system of equations (10) is equivalent to the system of equations (11):

$$\begin{aligned} x_f x_{imp} + y_f &= 1 \\ x_f x_{ind} + y_f &= 0 \\ y_d (x_{ind} + 1) + x_d &= 0 \\ y_d (x_{inc} + 1) + x_d &= -x_{inc} - 1 \end{aligned} \tag{11}$$

It is reduced to a system of linear equations. The matrix writing of the system of equations (11) is given by the vector equation (12):

$$\begin{pmatrix} x_{imp} & 1 & 0 & 0 \\ x_{ind} & 1 & 0 & 0 \\ 0 & 0 & x_{ind} + 1 & 1 \\ 0 & 0 & x_{inc} + 1 & 1 \end{pmatrix} \begin{pmatrix} x_f \\ y_f \\ y_d \\ x_d \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -x_{inc} - 1 \end{pmatrix} \tag{12}$$

Let us call $M_3 = \begin{pmatrix} x_{imp} + 1 & 1 & 0 & 0 \\ x_{ind} + 1 & 1 & 0 & 0 \\ 0 & 0 & x_{ind} + 1 & 1 \\ 0 & 0 & x_{inc} + 1 & 1 \end{pmatrix}$. Thus we must have $\det M_3 = (x_{imp} - x_{ind})(x_{ind} - x_{inc}) \neq 0$; so (i); for (ii), just take $\lim_{x_{inc} \rightarrow \infty} (x_d/(x_{inc} + 1)) = 0$; for (iii), if $x_{ind} = \infty$, then the last two equations do not make sense. \square

The current form of the studied system was the appearance of the system of equations (13):

$$\begin{aligned} \frac{x_f}{x_{imp} + m} + y_f &= 1 \\ \frac{x_f}{x_{ind} + m} + y_f &= 0 \\ \frac{x_d}{x_{ind} + m} + y_d &= 0 \\ \frac{x_d}{x_{inc} + m} + y_d &= -1 \end{aligned} \tag{13}$$

In (13) x_{imp} and x_{inc} can be infinite, x_{ind} is real, and $(x_{ind} + m) \in \mathbb{R}^*$.

Proposition 6. (i) If $(x_{imp}, x_{ind}, x_{inc}) \in \mathbb{R}^3$, with x_{imp} , x_{ind} , and x_{inc} pairwise distinct, then the system of equations (13) has four real solutions; (ii) if $x_{inc} = \infty$, $x_{ind} \in \mathbb{R}$ and if $x_{imp} = \infty$, then the system of equations (13) has four solutions actual such

that $y_d = -1$, $x_d = (x_{ind} + m)$ and $x_f = 1$, $y_f = -(x_{ind} + m)$;
 (iii) alternatively, the system of equations has no solution.

Proof. (i) The system of equations (13) is equivalent to the system of equations (14):

$$\begin{aligned} y_f (x_{imp} + m) + x_f &= (x_{imp} + m) \\ y_f (x_{ind} + m) + x_f &= 0 \\ y_d (x_{ind} + m) + x_d &= 0 \\ y_d (x_{inc} + m) + x_d &= -(x_{inc} + m) \end{aligned} \tag{14}$$

It therefore comes to a system of linear equations. The matrix writing system of equations (14) is given by the vector equation (15):

$$\begin{pmatrix} x_{imp} + m & 1 & 0 & 0 \\ x_{ind} + m & 1 & 0 & 0 \\ 0 & 0 & x_{ind} + m & 1 \\ 0 & 0 & x_{inc} + m & 1 \end{pmatrix} \begin{pmatrix} y_f \\ x_f \\ y_d \\ x_d \end{pmatrix} = \begin{pmatrix} x_{imp} + m \\ 0 \\ 0 \\ -(x_{inc} + m) \end{pmatrix} \tag{15}$$

Let us call $M_4 = \begin{pmatrix} x_{imp} + m & 1 & 0 & 0 \\ x_{ind} + m & 1 & 0 & 0 \\ 0 & 0 & x_{ind} + m & 1 \\ 0 & 0 & x_{inc} + m & 1 \end{pmatrix}$. Thus, we must always have $\det M_4 = (x_{imp} - x_{ind})(x_{ind} - x_{inc}) \neq 0$; for (ii), just take $\lim_{x_{inc} \rightarrow \infty} (x_d/x_{inc}) = 0$ and $\lim_{x_{imp} \rightarrow \infty} (x_f/x_{imp}) = 0$; for (iii), if $x_{ind} = \infty$, then all the equations do not make sense. \square

Note 7. Note that the different matrices for the four above-mentioned propositions as M_1, M_2, M_3, M_4 have the same determinant; this means that all the measures have the following conditions $x_{imp} \neq x_{ind} \neq x_{inc}$ and $(x_{imp}, x_{ind}, x_{inc}) \in \mathbb{R}^3$ have the same condition of normalizability.

4. Application of These Four Propositions

We recall in Table 1 the respective definitions of the various measures that lead to the results below.

(1) The normalized measure of “Cost Multiplying” is

$$\begin{aligned} &(Cost - Mult)_{shnd} (X \rightarrow Y) \\ &= \begin{cases} \frac{P(Y'/X') - P(Y')}{P(Y'/X')(1 - P(Y'))} & \text{if } X \text{ favors } Y; \\ \frac{P(Y'/X') - P(Y')}{P(Y')(1 - P(Y'/X'))} & \text{if } X \text{ disfavors } Y. \end{cases} \end{aligned}$$

TABLE 1: The probabilistic quality measures.

N^o	Measures μ	$\mu(X \rightarrow Y)$
1	Cost multiplying	$\frac{P(X' \cap Y')P(\bar{Y}')}{P(X' \cap \bar{Y}')P(Y')}$
2	Example counter-example	$\frac{P(X' \cap Y') - P(X' \cap \bar{Y}')}{P(X' \cap Y')}$
3	Informal Gain	$\log \frac{P(X' \cap Y')}{P(X')P(Y')}$
4	Odd-Ratio	$\frac{P(X' \cap Y')P(\bar{X}' \cap \bar{Y}')}{P(\bar{X}' \cap Y')P(X' \cap \bar{Y}')}$
5	Conviction	$\frac{P(X')P(\bar{Y}')}{P(X' \cap \bar{Y}')}$
6	Sebag	$\frac{P(X' \cap Y')}{P(X')P(\bar{Y}')}$

Soit $(Cost - Mult)_{shnd} (X \rightarrow Y)$

$$= \begin{cases} \frac{M_{GK}(X \rightarrow Y)}{P(Y'/X')} & \text{if } X \text{ favors } Y; \\ \frac{M_{GK}(X \rightarrow Y)}{(1 - P(Y'/X'))} & \text{if } X \text{ disfavors } Y. \end{cases} \tag{16}$$

(2) The normalized measure of “Sebag” is

$Sebag_{shnd} (X \rightarrow Y)$

$$= \begin{cases} \frac{P(Y'/X') - P(Y')}{P(Y'/X')(1 - P(Y'))} & \text{if } X \text{ favors } Y; \\ \frac{P(Y'/X') - P(Y')}{P(Y')(1 - P(Y'/X'))} & \text{if } X \text{ disfavors } Y. \end{cases} \tag{17}$$

Soit $Sebag_{shnd} (X \rightarrow Y)$

$$= \begin{cases} \frac{M_{GK}(X \rightarrow Y)}{P(Y'/X')} & \text{if } X \text{ favors } Y; \\ \frac{M_{GK}(X \rightarrow Y)}{(1 - P(Y'/X'))} & \text{if } X \text{ disfavors } Y. \end{cases}$$

(3) The normalized measure of “Example counter-example” is

$(Ex - Cont)_{shnd} (X \rightarrow Y)$

$$= \begin{cases} \frac{P(Y'/X') - P(Y')}{P(Y'/X')(1 - P(Y'))} & \text{if } X \text{ favors } Y; \\ \frac{P(Y'/X') - P(Y')}{P(Y')(1 - 2P(Y'/X'))} & \text{if } X \text{ disfavors } Y. \end{cases} \tag{18}$$

Soit $Ex - Cont_{shng} (X \rightarrow Y)$

$$= \begin{cases} \frac{M_{GK}(X \rightarrow Y)}{P(Y'/X')} & \text{if } X \text{ favors } Y; \\ \frac{M_{GK}(X \rightarrow Y)}{(1 - P(Y'/X'))} & \text{if } X \text{ disfavors } Y. \end{cases}$$

(4) The normalized measure of “Odd-Ratio” is

$$\begin{aligned}
 \text{Odd}R_{shnd}(X \rightarrow Y) &= \begin{cases} \frac{P(Y'/X') - P(Y')}{P(Y'/X')(1 - P(X') - P(Y') + P(Y'/X')P(X'))} & \text{if } X \text{ favors } Y; \\ \frac{P(Y'/X') - P(Y')}{(1 - P(Y'/X'))(P(Y') - P(Y'/X')P(X'))} & \text{if } X \text{ disfavors } Y. \end{cases} \\
 (\text{Odd} - \text{Ratio})_{shnd}(X \rightarrow Y) &= \begin{cases} \frac{(1 - P(X'))M_{GK}(X \rightarrow Y)}{P(Y'/X')(1 - P(X') - P(Y') + P(Y'/X')P(X'))}, & \text{if } X \text{ favors } Y; \\ \frac{P(Y')M_{GK}(X \rightarrow Y)}{(1 - P(Y'/X'))(P(Y') - P(Y'/X')P(X'))}, & \text{if } X \text{ disfavors } Y. \end{cases}
 \end{aligned} \tag{19}$$

(5) The normalized measure of “Conviction” is

$$\begin{aligned}
 \text{Conv}_{in}(X \rightarrow Y) &= \begin{cases} \frac{P(Y'/X') - P(Y')}{1 - P(X')}, & \text{if } X \text{ favors } Y; \\ \frac{P(Y'/X') - P(Y')}{P(Y')}, & \text{if } X \text{ disfavors } Y. \end{cases}
 \end{aligned} \tag{20}$$

Soit $\text{Conv}_{in}(X \rightarrow Y) = M_{GK}(X \rightarrow Y)$.

(6) The normalized measure of “Informal Gain” is

$$\begin{aligned}
 (\text{Gain} - \text{Inf})_{smg} &= \begin{cases} \frac{\log P(Y') - \log P(Y'/X')}{\log P(Y')}, & \text{if } X \text{ favors } Y; \\ \frac{\log P(Y'/X') - \log P(Y')}{\log P(Y'/X') - \log P(Y') + 1}, & \text{if } X \text{ disfavors } Y. \end{cases}
 \end{aligned} \tag{21}$$

Note that the normalized of Gain-Infomal measure has no relationship with M_{GK} , but it is continuing. Table 1 calls the respective definitions of the various measures.

The following theorem supplements in its part (i) the statement in [17] (paragaph 3.2, page 62).

Theorem 8. *A probabilistic quality measure μ is normalizing if and only if, for any association rule $X \rightarrow Y$, the following conditions are met: (i) the following inequalities are satisfied $x_{imp} \neq x_{ind} \neq x_{inc}$, if $(x_{imp}, x_{ind}, x_{inc}) \in \mathbb{R}^3$; (ii) if $(x_{imp}, x_{ind}, x_{inc}) \in \mathbb{R}^3$, then it is necessary that any of the three values $(x_{imp}, x_{ind}, x_{inc})$ must be infinite; or in the case where two values are infinite, x_{ind} must always be finite, and $x_2 + m \in \mathbb{R}^*$.*

Proof. (i) $x_{imp} \neq x_{ind} \neq x_{inc}$ and if $(x_{imp}, x_{ind}, x_{inc}) \in \mathbb{R}^3$, then the determinant of the matrix M_i ($i = 1, 2, 3, 4$) associated with all the equations (6), (9), (12), and (15) is equal to $(x_{imp} - x_{ind})(x_{ind} - x_{inc})$ and is, thus, nonzero; for (ii), if $(x_{imp}, x_{ind}, x_{inc}) \in \mathbb{R}^3$, and one or two of them are infinite, then it is necessary to use Propositions 3, 4, 5, and 6 (ii) above. Hence the theorem is stated. \square

5. Conclusion and Perspectives

This study makes a significant number of quality measures that remain nonaffine normalizable (e.g., *Sebag, Odd-Ratio, Example counter-example, Informal Gain, Cost Multiplying, and Conviction*), through the use of homographic functions. However, we have shown that each measure has its own position in relation to the three intuitive references such as incompatibility, independency, and logical implication. In this case, we have shown that we can use a homographic homeomorphism or even combine a homographic homeomorphism with an affine homeomorphism, when the infinity appears as reference coefficient. Only, we must be able to play with the homographic homeomorphism in accordance with its property and according to the need. It is explained in our work only that any situation is infinite; moreover, it is not yet risen to the position of stochastic independence: only homographic functions of the type $x \mapsto b/(x + m)$, where $m = 0$ and $b \in \mathbb{R}^*$, are sufficient to solve the problem of normalization. A small exception will be noticed on the measurement *Informal Gain*, because it is infinite at incompatibility and equal to zero at the stochastic independence. Therefore, it appears necessary to introduce the two functions $x_{inc} \mapsto b/(x_{inc} + m)$ and $x_{ind} \mapsto b/(x_{ind} + m)$ and $m \neq 0$, for example, $m = 1$. Finally, this study is based on normalization by homography complementarily to those by affine application; this remains around the quality measure M_{GK} with its two components M_{GK}^f and M_{GK}^d multiplying by closely factors. This reinforces the unifying property of M_{GK} relative to all measures existing in literature. Although we have these homographic homeomorphisms and affine homeomorphisms to normalize interestingness measures, the route to be taken is even longer. Indeed, there is a group of measures that resist to the use of these new tools, namely, *Klogsen, Pondered dependency, One way support, Bilateral Support, Coverage, and Prevalence*, because they do not meet the conditions for normalization. So, the problem remains still open with respect to the transformation that allows normalization of those quality measures in a sense still to be specified.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] R. Agrawal, T. Imieliński, and A. Swami, "Mining association rules between sets of items in large databases," in *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '93)*, pp. 207–216, Washington, DC, USA, 1993.
- [2] J. Diatta, H. Ralambondrainy, and A. Totohasina, "Towards a unifying probabilistic implicative normalized quality measure for association rules," in *Quality Measures in Data Mining*, 250, 237, 2007.
- [3] J. Diatta, H. Ralambondrainy, and A. Totohasina, "Towards a unifying probabilistic implicative normalized quality measure for association rules," in *Quality Measures in Data Mining*, chapter 10, pp. 237–237-250, Springer-verlag, 2007.
- [4] M. Hahsler and K. Hornik, "New probabilistic interest measures for association rules," *Intelligent Data Analysis*, vol. 11, no. 5, pp. 437–455, 2007.
- [5] W. Aljandal, W. Hsu, V. Bahirwani, D. Caragea, and T. Weninger, "Validation-based Normalization and Selection of Interestingness Measures for Association Rules," in *Proceedings of the 18th International Conference on Artificial Neural Networks in Engineering*, pp. 1–8, 2008.
- [6] A. Totohasina, "Towards a theory unifying implicative interestingness measures and critical values consideration in MGK," in *Educação Matemática Pesquisa*, vol. 16, pp. 881–900, Special ASI, São Paulo, Brazil, 13 edition, 2014.
- [7] E. M. Nguifo, D. Grissa, and S. Guillume, "Categorisation des mesures d'intérêt pour l'extraction des connaissances," in *Revue des Nouvelles Technologies de L'Information*, D. A. Zighed and G. Venturri, Eds., pp. 117–143, 2012.
- [8] M. Hahsler and K. H. Vienna, *New Probabilistic Interest Measures for Association Rules*, University of Economics and Business Administration, Vienna, Austria, 2018.
- [9] G. Piatetsky-Shapiro, P. et Smyth, and U. M. Fayyad, "Knowledge discovery and data mining: towards a unifying framework," in *Proceedings of the second International Conference on Knowledge Discovery and Data Mining*, pp. 82–88, 1996.
- [10] D. R. Feno, *Mesure de qualité des règles d'association : normalisation et caractérisation des règles d'association des bases [PhD thesis]*, Université de La Réunion spécialité : Mathématiques Informatique, 2007.
- [11] N. Bhargava and M. Shukla, "Survey of Interestingness Measures for Association Rules Mining: Data Mining, Data Science for Business Perspective," *IRACST - International Journal of Computer Science and Information Technology & Security (IJC-SITS)*, vol. 6, no. 2, 2016.
- [12] E. H. Shortliffe and B. G. Buchanan, "A model of inexact reasoning in medicine," *Mathematical Biosciences*, vol. 23, pp. 351–379, 1975.
- [13] A. Totohasina, *Contribution à l'étude des mesures de qualité des règles d'associations: normalisation sous cinq contraintes et cas de M_{GK} : propriétés, bases composites des règles et extension en vue d'applications en statistique et en sciences physiques [PhD thesis]*, Université, d'Antsiranana, Madagascar: Mathématiques Informatique, 2008.
- [14] A. Totohasina, "Normalisation de mesures probabilistes de la qualité des règles," in *Proceedings of the SFDS'03, XXXV ième Journées de Statistiques*, pp. 985–988, Lyon, France, 2003.
- [15] A. Totohasina and H. Ralambondrainy, "ION: A pertinent new measure for mining information from many types of data," in *Proceedings of the 1st IEEE International Conference on Signal-Image Technology and Internet-Based Systems, SITIS 2005*, pp. 202–207, Yaoundé, Cameroon, December 2005.
- [16] F. Berzal, I. Blanco, D. Sanchez, and M.-A. Vila, "Measuring the accuracy and interest of association rules: a new framework," *Intelligent Data Analysis*, vol. 6, no. 3, pp. 221–235, 2002.
- [17] C. C. Aggarwal and P. S. Yu, "A new framework for itemset generation," in *Proceedings of the 98, Symposium on Principles of Database Systems*, pp. 18–24, Seattle, WA, USA, June 1998.

