

Research Article

Toward a Theory of Normalizing Function of Interestingness Measure of Binary Association Rules

Armand Armand,¹ André Totohasina², and Daniel Rajaonasy Feno³

¹Lycee Mixte Antsiranana, Madagascar

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

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

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

Received 2 September 2018; Accepted 29 October 2018; Published 14 November 2018

Academic Editor: Theodore E. Simos

Copyright © 2018 Armand 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.

Regarding the existence of more than sixty interestingness measures proposed in the literature since 1993 till today in the topics of association rules mining and facing the importance these last one, the research on normalization probabilistic quality measures of association rules has already led to many tangible results to consolidate the various existing measures in the literature. This article recommends a simple way to perform this normalization. In the interest of a unified presentation, the article offers also a new concept of normalization function as an effective tool for resolution of the problem of normalization measures that have already their own normalization functions.

1. Introduction

1.1. Definitions and Notations. We always put ourselves in the framework of a context of binary data mining (see, for example, [1–3], which illustrate the importance of association rules mining based on choosing some quality measure) $D = (T, I, \mathcal{R})$, where I is a nonempty finite set of attributes or variables, T a finite set of entities or objects, \mathcal{R} a binary relation from T to I , and P discrete uniform probability in the probabilistic space $(T, \mathcal{P}(T))$ [4, 5].

In the next sections, we use the following notation for two itemsets X, Y : $X' = \{e \in T / \forall x \in X; e \mathcal{R} x\}$, i.e., the set of all transactions containing the X pattern that is the dual of a pattern X of I ($X \subseteq I$) [4, 6, 7]. $n = |T|$ represents the size of the total sample size; $n_X = |X'|$ represents the number of transactions satisfying pattern X ; $n_{XY} = |X' \cap Y'|$ represents the number of transactions satisfying both X and Y ; $n_{\overline{X}} = |T| - |T - T_{X'}|$, where \overline{X} is the logical negation of X ; $n_X/n = |T_{X'}|/|T|$ represents the support of the pattern X .

Hereafter, our work is divided into three sections. Section 2 gives the definition of normalization function. Section 3 recommends the raw results of normalizing function of some probabilistic quality measures.

Finally Section 4, [5, 8–10] sets out the conclusion and perspectives.

2. Normalizing Function F_n

2.1. Motivations. The theory and practice of normalization probabilistic quality measures (see, for example, [4, 5, 11], (Totohasina et al. [12]), [6, 7]) have been resolving included in the list of tools for problems concerning the data mining. This is done in the view of regrouping [3, 4, 6, 12–17] different existing measures available from the literature. Let us notice that [4] proves existence of infinity quality measures through the concept of the so-called normalized quality measure under five conditions, but recently [16] still proposes a novel interestingness measure. By opening the door to the possibility of creating definitions of new concepts in the context of data mining, perhaps, this will bring to following a new reflection among researchers in this field. This is the normalizing function. What is meant by a normalizing function? The following section will attempt to answer such question. Remember that this paper is the logical continuation of the paper [18].

2.2. Proposal Approach

Definition 1. We consider a probabilistic measure of interest μ . We call normalizing function of μ the piecewise continuous function that can normalize directly the measure μ and is defined as

$$F_n(x) = \begin{cases} F_n^f(x), & \text{if } X \text{ favors } Y \text{ and } x \in]x_{ind}; x_{imp}[; \\ F_n^d(x), & \text{if } X \text{ disfavors } Y \text{ and } x \in]x_{inc}; x_{ind}[. \end{cases} \quad (1)$$

It takes the following particular values:

$$F_{an}(x) = \begin{cases} x_f x + y_f, & \text{if } X \text{ favors } Y; \\ x_d x + y_d, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (2)$$

$$F_{shnd}(x) = \begin{cases} \frac{x_f}{x+m} + y_f, & \text{if } X \text{ favors } Y; \\ x_d x + y_d, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (3)$$

$$F_{shng}(x) = \begin{cases} x_f x + y_f, & \text{if } X \text{ favors } Y; \\ \frac{x_d}{x+m} + y_d, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (4)$$

$$F_{hn}(x) = \begin{cases} \frac{x_f}{x+m} + y_f, & \text{if } X \text{ favors } Y; \\ \frac{x_d}{x+m} + y_d, & \text{if } X \text{ disfavors } Y, \end{cases} \quad (5)$$

where x is value of μ ; x_{imp} is value of μ at logical implication; x_{ind} is value of μ at independence; x_{inc} is value of μ at incompatibility; F_n^f is normalizing function, if X favors Y , that is to say, in case where $P(Y'/X') > P(Y')$; F_n^d is normalizing function, if X disfavors Y , that is to say, in case where $P(Y'/X') < P(Y')$; F_{an} is normalizing function of measure μ , that is, affine normalizable; F_{hn} is normalizing function of measure μ , that is, homographic normalizable; F_{shnd} is normalizing function of measure μ , that is, semihomographic-normalizable to right; F_{shng} is μ normalizing function of measure μ , that is, semihomographic-normalizable to left; x_f , y_f , x_d , and y_d are μ of normalization coefficient x ; m is real and $\overline{\mathbb{R}} = [-\infty; +\infty]$.

The definition of normalizing function is thus determined according to the quality measures. Note that this is a function as any other, with respect to the variable x ; it is a numerical function of a real variable; only, this variable x is as follows:

$x = f_1(n, P(X'), P(Y'), P(Y' \cap X'))$. The four coefficients of normalization of probabilistic measures quality have to meet the following conditions:

$x_f = f_2(n, P(X'), P(Y'))$, $y_f = f_3(n, P(X'), P(Y'))$, $x_d = f_4(n, P(X'), P(Y'))$, and $y_d = f_5(n, P(X'), P(Y'))$, where f_1, f_2, f_3, f_4 , and f_5 are the five real functions. We note in passing that there is also a group of measures with the same normalization coefficients. The normalizing function is also one of the means that allow to give an interpretation of a normalized measurement after providing the respective

values of normalization coefficients x_f , y_f , x_d , and y_d . It also expresses an opportunity of value for all normalized measures. According to what we have just written, any normalization of function reflecting the objective in normalization must necessarily have the following properties.

Property 2 (necessary conditions). P_1 : F_n^f is a continuous function, positive and strictly increasing on the interval $]x_{ind}; x_{imp}[$ and realizes a bijection on the interval $]x_{ind}; x_{imp}[\rightarrow]0; 1[$; that is to say, F_n^f must have the following: a limit 1 to the point x_{imp} ; that is to say, $\lim_{x \rightarrow x_{imp}} F_n^f(x) = 1$; a limit 0 to the point x_{ind} ; that is to say, $\lim_{x \rightarrow x_{ind}} F_n^f(x) = 0$.

P_2 : F_n^d is a continuous function, negative, and strictly increasing on the interval $]x_{inc}; x_{ind}[$ and realizes a bijection on the interval $]x_{inc}; x_{ind}[\rightarrow]-1; 0[$; that is to say, F_n^d must have the following: a limit 0 to the point x_{ind} ; that is to say, $\lim_{x \rightarrow x_{ind}} F_n^d(x) = 0$; a limit -1 to the point x_{inc} ; that is to say, $\lim_{x \rightarrow x_{inc}} F_n^d(x) = -1$. As a recap, we have the following.

P_3 : F_n is a continuous function at the point x_{ind} and increasing on the interval $]x_{inc}; x_{imp}[$ and realizes a bijection on the interval $]x_{inc}; x_{imp}[\rightarrow]-1; 1[$; that is to say, F_n must have the following: a limit 1 to the point x_{imp} ; that is to say, $\lim_{x \rightarrow x_{imp}} F_n(x) = 1$; a limit 0 to the point x_{ind} ; that is to say, $\lim_{x \rightarrow x_{ind}} F_n(x) = 0$ and $F_n(x_{ind}) = 0$; limit -1 to the point x_{ind} ; that is to say, $\lim_{x \rightarrow x_{inc}} F_n(x) = -1$.

3. Applications

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

(1) **Cost multiplying:** $x = P(X' \cap Y')P(\overline{Y'})/P(X' \cap \overline{Y'})P(Y') = (1 - P(Y'))P(Y'/X')/(1 - P(Y'/X'))P(Y')$ which is such that $x_{imp} = +\infty$, $x_{ind} = 1$, $x_{inc} = 0$. Using research theories of normalization coefficients x_f , y_f , x_d , and y_d in [18] we have $x_f = -1$, $y_f = 1$, $x_d = 1$, and $y_d = -1$. As a result, by replacing x , x_f , y_f , x_d , and y_d by their values in the expression (3) its normalizing function is such that

$$F_{shnd}(x)$$

$$= \begin{cases} -\frac{1}{x} + 1, & \text{if } X \text{ favors } Y \text{ and } x \in [1; +\infty[\\ x - 1, & \text{if } X \text{ disfavors } Y \text{ and } x \in [0; 1]. \end{cases} \quad (6)$$

It is easy to see that this function is continuous piecewise, in particular point $x_{ind} = 1$.

Finally, it represents a function that exposed the necessary and sufficient conditions for the normalized and continuous measures.

Its tables of variation are respectively presented in Figures 1 and 2.

The graphic illustration and the variation table of F_n of the measure Cost Multiplying reveal that F_n^f is strictly increasing, is positive, and realizes a bijection on

TABLE 1: Probabilistic quality measures of expression.

N^o	Measures μ	Expressions of $x = \mu(X \rightarrow Y)$
1	Cost multiplying	$x = \frac{P(X' \cap Y') P(\bar{Y}')}{P(X' \cap \bar{Y}') P(Y')}$
2	Example counter-example	$x = \frac{P(X' \cap Y') - P(X' \cap \bar{Y}')}{P(X' \cap Y')}$
3	Informal gain	$x = \log \frac{P(X' \cap Y')}{P(X') P(Y')}$
4	Odd-Ratio	$x = \frac{P(X' \cap Y') P(\bar{X}' \cap \bar{Y}')}{{P(\bar{X}' \cap Y')} P(X' \cap \bar{Y}')}}$
5	Conviction	$x = \frac{P(X') P(\bar{Y}')}{P(X' \cap \bar{Y}')}}$
6	Sebag	$x = \frac{P(X' \cap Y')}{P(X') P(\bar{Y}')}$
7	M_{GK}	$x = \begin{cases} \frac{P(Y'/X') - P(Y')}{1 - P(Y')} \\ \frac{P(Y'/X') - P(Y')}{P(Y')} \end{cases}$
8	Rucel and Rao Index	$x = P(X' \cap Y')$
9	Confidence or precision	$x = P\left(\frac{Y'}{X'}\right)$
10	Recall	$x = P\left(\frac{X'}{Y'}\right)$
11	Interest or Lift	$x = \frac{P(Y'/X')}{P(Y')}$
12	Leverage	$x = P\left(\frac{Y'}{X'}\right)$
13	Centered confidence	$x = P\left(\frac{Y'}{X'}\right) - P(Y')$
14	Confirmed confidence	$x = 1 - 2P\left(\frac{\bar{Y}'}{X'}\right)$
15	Certainty factor	$x = \frac{P(Y'/X') - P(Y')}{1 - P(Y')}$
16	Gras implication	$x = \sqrt{n} \frac{P(X' \cap Y') - P(X') P(Y')}{\sqrt{P(X') P(Y')}}$
17	Piatesky-Shapiro	$x = n(P(X' \cap Y'))$
18	Cosinus	$x = \frac{P(X' \cap Y')}{\sqrt{P(X') P(Y')}}$
19	Loevinger	$x = 1 - \frac{P(X' \cap \bar{Y}')}{P(X') P(\bar{Y}')}$
20	Cohen ou Kappa	$x = 2 \frac{P(X' \cap Y') - P(X') P(Y')}{P(X') + P(Y') - 2P(X') P(Y')}$
21	Addiction	$x = \left P\left(\frac{Y'}{X'}\right) - P\left(\frac{\bar{Y}'}{X'}\right) \right $
22	Novelty	$x = P(X'Y') - P(X') P(Y')$
23	Czekanowski-Dice	$x = \frac{2P(X' \cap Y')}{P(X' \cap Y') + 1 - P(\bar{X}' \cap \bar{Y}')}$
24	Relative risk	$x = \frac{P(X'/Y')}{P(X'/\bar{Y}')}$
25	Negative reliability	$x = P\left(\frac{\bar{X}'}{\bar{Y}'}\right)$

x	0	1	$+\infty$
$F'_{shnd}(x)$	+		+
$F_{shnd}(x)$	Repulsion zone -1	0	Attraction zone 1

FIGURE 1: Change in measure normalizing function “Cost Multiplying”.

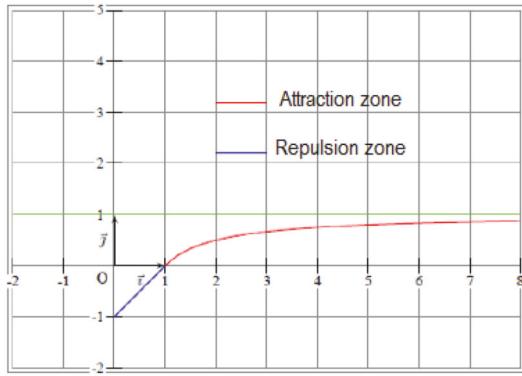


FIGURE 2: Geometric interpretation of the normalization function of the “Cost Multiplying” measure.

$[1; +\infty[\rightarrow [0; 1]$ and even f_d is strictly increasing and negative and realizes a bijection on $[0; 1] \rightarrow [-1; 0]$. These results show that F_n is strictly increasing and it realizes a bijection on $[0; +\infty[\rightarrow [-1; 1]$.

Note that, in the following, for the search of the four normalization coefficients, the same principle is used with the measure “dimension multiplier” for the other following measures because they are already expressed in Table 1.

- (2) **Example counter-example: its normalizing function is such that** using the expression (4) for $m = 0$ with $x_f = P(Y')/(1 - P(Y'))$, $y_f = (1 - 2P(Y'))/(1 - P(Y'))$, $x_d = (2P(Y') - 1)/P(Y')$ and $y_d = -1$ so

$$F_{shng}(x)$$

$$= \begin{cases} \frac{P(Y')}{1 - P(Y')} x + \frac{1 - 2P(Y')}{1 - P(Y')}, & \text{if } X \text{ favors } Y \\ \frac{2P(Y') - 1}{P(Y')} \frac{1}{x} - 1, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (7)$$

- (3) **Informal gain: its normalizing function is such that** using expression (4) for $m = 1$ with $x_f = -1/\log P(Y')$, $y_f = 0$, $x_d = 1$, and $y_d = -1$ so

$$F_{shng}(x) = \begin{cases} \frac{-1}{\log P(Y')} x, & \text{if } X \text{ favors } Y \\ \frac{1}{x + 1} - 1, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (8)$$

- (4) **Odd-ratio: its normalizing function is such that** using expression (3) for $m = 0$ with $x_f = -1$, $y_f = 1$, $x_d = 1$, and $y_d = -1$ so

$$F_{shnd}(x) = \begin{cases} \frac{1}{x} + 1, & \text{if } X \text{ favors } Y \text{ and } x \in [1; +\infty[\\ x - 1, & \text{if } X \text{ disfavors } Y \text{ and } x \in [0; 1]. \end{cases} \quad (9)$$

- (5) **Conviction: its normalizing function is such that** using expression (5) for $m = 0$ with $x_f = -1$, $y_f = 1$, $x_d = -(1 - P(Y'))/P(Y')$, and $y_d = (1 - P(Y'))/P(Y')$ so

$$F_{hn}(x)$$

$$= \begin{cases} \frac{-1}{x} + 1, & \text{if } X \text{ favors } Y \\ -\frac{1 - P(Y')}{P(Y')} \frac{1}{x} + \frac{1 - P(Y')}{P(Y')}, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (10)$$

- (6) **Sebag: its normalizing function is such that** using expression (3) for $m = 0$ with $x_f = -P(Y')/(1 - P(Y'))$, $y_f = 1$, $x_d = (1 - P(Y'))/P(Y')$, and $y_d = -1$ so

$$F_{shnd}(x) = \begin{cases} -\frac{P(Y')}{1 - P(Y')} \frac{1}{x} + 1, & \text{if } X \text{ favors } Y \\ \frac{1 - P(Y')}{P(Y')} x - 1, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (11)$$

- (7) **M_{GK} : its normalizing function is such that** using expression (2) with $x_f = 1$, $y_f = 0$, $x_d = 1$, and $y_d = 0$ so

$$F_{an}(x) = x \quad \text{avec } x = \begin{cases} \frac{P(Y'/X') - P(Y')}{1 - P(Y')}, & \text{if } X \text{ favors } Y \text{ and } x \in [0; 1] \\ \frac{P(Y'/X') - P(Y')}{P(Y')}, & \text{if } X \text{ disfavors } Y \text{ and } x \in [-1; 0]. \end{cases} \quad (12)$$

- (8) **Support; its normalizing function is such that** using expression (2) with $x_f = 1/P(X')(1 - P(Y'))$, $y_f = -P(Y')/(1 - P(Y'))$, $x_d = 1/P(X')P(Y')$, and $y_d = -1$ so

$$F_{an}(x) = \begin{cases} \frac{1}{P(X')(1 - P(Y'))}x - \frac{P(Y')}{1 - P(Y')}, & \text{if } X \text{ favors } Y \\ \frac{1}{P(X')P(Y')}x - 1, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (13)$$

- (9) **Confidence: its normalizing function is such that** using expression (2) with $x_f = 1/(1 - P(Y'))$, $y_f = -P(Y')/(1 - P(Y'))$, $x_d = 1/P(Y')$, and $y_d = -1$ so

$$F_{an}(x) = \begin{cases} \frac{1}{1 - P(Y')}x - \frac{P(Y')}{1 - P(Y')}, & \text{if } X \text{ favors } Y \\ \frac{1}{P(Y')}x - 1, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (14)$$

- (10) **Recall: its normalizing function is such that** using expression (2) with $x_f = P(Y')/P(X')(1 - P(Y'))$, $y_f = -P(Y')/(1 - P(Y'))$, $x_d = 1/P(X')$, and $y_d = -1$ so

$$F_{an}(x) = \begin{cases} \frac{P(Y')}{P(X')(1 - P(Y'))}x - \frac{P(Y')}{1 - P(Y')}, & \text{if } X \text{ favors } Y \\ \frac{1}{P(X')}x - 1, & \text{if } X \text{ disfavors } Y \end{cases} \quad (15)$$

- (11) **Lift: its normalizing function is such that** using expression (2) with $x_f = P(Y')/(1 - P(Y'))$, $y_f = -P(Y')/(1 - P(Y'))$, $x_d = 1$, and $y_d = -1$ so

$$F_{an}(x) = \begin{cases} \frac{P(Y')}{1 - P(Y')}x - \frac{P(Y')}{1 - P(Y')}, & \text{if } X \text{ favors } Y \\ x - 1, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (16)$$

- (12) **Leverage: its normalizing function is such that** using expression (2) with $x_f = P(Y')/(1 - P(Y'))$, $y_f = -P(Y')(P(X') - 1)/(1 - P(Y'))$, $x_d = 1/P(Y')$, and $y_d = -P(Y')(P(X') - 1)/P(Y')$ so

$$F_{an}(x) = \begin{cases} \frac{P(Y')}{1 - P(Y')}x - \frac{P(Y')(P(X') - 1)}{1 - P(Y')}, & \text{if } X \text{ favors } Y \\ \frac{1}{P(Y')}x - \frac{P(Y')(P(X') - 1)}{P(Y')}, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (17)$$

- (13) **Centered confidence: its normalizing function is such that** using expression (2) with $x_f = 1/(1 - P(Y'))$, $y_f = 0$, $x_d = 1/P(Y')$, and $y_d = 0$ so

$$F_{an}(x) = \begin{cases} \frac{1}{1 - P(Y')}x, & \text{if } X \text{ favors } Y \\ \frac{1}{P(Y')}x, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (18)$$

- (14) **Featured confirmed confidence: its normalizing function is such that** using expression (2) with $x_f = 1/2(1 - P(Y'))$, $y_f = (1 - 2P(Y'))/2P(Y')$, $x_d = 1/2P(Y')$, and $y_d = (1 - 2P(Y'))/2P(Y')$ so

$$F_{an}(x) = \begin{cases} \frac{1}{2(1 - P(Y'))}x + \frac{1 - 2P(Y')}{2(1 - P(Y'))}, & \text{if } X \text{ favors } Y \\ \frac{1}{2P(Y')}x + \frac{1 - 2P(Y')}{2P(Y')}, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (19)$$

- (15) **Certainty factor: its normalizing function is such that** using expression (2) with $x_f = 1$, $y_f = 0$, $x_d = (1 - P(X'))/P(Y')$, and $y_d = 0$ so

$$F_{an}(x) = \begin{cases} x, & \text{if } X \text{ favors } Y \\ \frac{1 - P(X')}{P(Y')}x, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (20)$$

- (16) **Gras implication index: its normalizing function is such that** using expression (2) with $x_f = \sqrt{P(X')P(Y')}/\sqrt{n}P(X')(1 - P(Y'))$, $y_f = 0$, $x_d = (\sqrt{P(X')P(Y')}/\sqrt{n}P(X')P(Y'))x$, and $y_d = 0$ so

$$F_{an}(x) = \begin{cases} \frac{\sqrt{P(X')P(Y')}}{\sqrt{n}P(X')(1 - P(Y'))}x, & \text{if } X \text{ favors } Y \\ \frac{\sqrt{P(X')P(Y')}}{\sqrt{n}P(X')P(Y')}x, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (21)$$

- (17) **Piatetsky-Shapiro: its normalizing function is such that** using expression (2) with $x_f = 1/nP(X')(1 - P(Y'))$, $y_f = 0$, $x_d = 1/nP(X')P(Y')$, and $y_d = 0$ so

$$F_{an}(x) = \begin{cases} \frac{1}{nP(X')(1 - P(Y'))}x, & \text{if } X \text{ favors } Y \\ \frac{1}{nP(X')P(Y')}x, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (22)$$

- (18) **Cosinus: its normalizing function is such that** using expression (2) with $x_f = \sqrt{P(X')P(Y')}/P(X')(1 -$

$$P(Y')), \quad y_f = -P(X')/(1 - P(Y')), \quad x_d = -2P(X')P(Y')/2P(X')(1 - (P(Y'))), \quad x_d = (P(X') + P(Y'))/2P(X')P(Y'), \text{ and } y_d = -1 \text{ so}$$

$F_{an}(x)$

$$= \begin{cases} \frac{\sqrt{P(X')P(Y')}}{P(X')(1-P(Y'))}x - \frac{P(X')}{1-P(Y')}, & \text{if } X \text{ favors } Y \\ \frac{\sqrt{P(X')P(Y')}}{P(X')P(Y')}x - 1, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (23)$$

- (19) **Loevinger: its normalizing function is such that** using expression (2) with $x_f = 1$, $y_f = 0$, $x_d = (1 - P(Y'))/P(Y')$, and $y_d = 0$ so

$$F_{an}(x) = \begin{cases} x, & \text{if } X \text{ favors } Y \\ \frac{1 - P(Y')}{P(Y')}x, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (24)$$

- (20) **Cohen ou Kappa: its normalizing function is such that** using expression (2) with $x_f = (P(X') + P(Y') - 2P(X')P(Y'))/2P(X')(1 - P(Y'))$, $y_f = 0$, $x_d = (P(X') + P(Y') - 2P(X')P(Y'))/2P(X')P(Y')$, and $y_d = 0$ so

$$F_{an}(x) = \begin{cases} \frac{P(X') + P(Y') - 2P(X')P(Y')}{2P(X')(1 - P(Y'))}x, & \text{if } X \text{ favors } Y \\ \frac{P(X') + P(Y') - 2P(X')P(Y')}{2P(X')P(Y')}x, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (25)$$

- (21) **Addiction: its normalizing function is such that** using expression (2) with $x_f = 1/(1 - P(Y'))$, $y_f = 0$, $x_d = 1/P(Y')$, and $y_d = -1$ so

$$F_{an}(x) = \begin{cases} \frac{1}{1 - P(Y')}x, \\ \frac{1}{P(Y')}x - 1. \end{cases} \quad (26)$$

- (22) **Novelty: its normalizing function is such that** using expression (2) with $x_f = 1/P(X')(1 - P(Y'))$, $y_f = 0$, $x_d = 1/P(X')P(Y')$, and $y_d = 0$ so

$$F_{an}(x) = \begin{cases} \frac{1}{P(X')(1 - P(Y'))}x \\ \frac{1}{P(X')P(Y')}x. \end{cases} \quad (27)$$

- (23) **Czekanowski-Dice ou F-measure: its normalizing function is such that** using expression (2) with $x_f = (P(X') + P(Y'))/2P(X')(1 - P(Y'))$, $y_f =$

$F_{an}(x)$

$$= \begin{cases} \frac{P(X') + P(Y')}{2P(X')(1 - P(Y'))}x - \frac{2P(X')P(Y')}{2P(X')(1 - (P(Y')))} \\ \frac{P(X') + P(Y')}{2P(X')P(Y')}x - 1. \end{cases} \quad (28)$$

- (24) **Relative risk: its normalizing function is such that** using expression (2) with $x_f = (P(Y') - P(X'))/(1 - P(Y'))$, $y_f = -(P(Y') - P(X'))/(1 - P(Y'))$, $x_d = 1$, and $y_d = -1$ so

$$F_{an}(x) = \begin{cases} \frac{P(Y') - P(X')}{1 - P(Y')}x - \frac{P(Y') - P(X')}{1 - P(Y')} \\ x - 1. \end{cases} \quad (29)$$

- (25) **Negative reliability: its normalizing function is such that** using expression (2) with $x_f = 1/P(X')$, $y_f = -(1 - P(X') - P(Y') + P(Y')P(X'))/P(X')(1 - P(Y'))$, $x_d = (1 - P(Y'))/P(X')(Y')$, and $y_d = -(1 - P(X') - P(Y') + P(Y')P(X'))/P(X')P(Y')$ so

$F_{an}(x)$

$$= \begin{cases} \frac{1}{P(X')}x - \frac{1 - P(X') - P(Y') + P(Y')P(X')}{P(X')(1 - P(Y'))} \\ \frac{1 - P(Y')}{P(X')(Y')}x - \frac{1 - P(X') - P(Y') + P(Y')P(X')}{P(X')P(Y')}. \end{cases} \quad (30)$$

Researches out on the expressions of measurement normalizing functions prove that certain measures have identical normalizing functions, leading to what state the theorem below.

Theorem 3. (i) All measures having the same baseline (x_{imp} , x_{ind} , x_{inc}) have the same normalizing function.

(ii) All measures affine normalizable are homographic normalizable, but the converse is false.

Proof. (i) Suppose the four possible forms of F_n .

(a) μ is affine normalizable:

$$F_{an}(x) = \begin{cases} x_f x + y_f, & \text{if } X \text{ favors } Y \\ x_d x + y_d, & \text{if } X \text{ disfavors } Y \end{cases} \quad (31)$$

with

$$x_f = \frac{1}{x_{imp} - x_{ind}},$$

$$y_f = -\frac{1}{x_{imp} - x_{ind}}x_{ind}, \quad (32)$$

$$x_d = \frac{1}{x_{ind} - x_{inc}}$$

and

$$y_d = -\frac{1}{x_{ind} - x_{inc}} x_{ind} \quad (33)$$

so we have

$$\begin{aligned} F_{an}(x) &= \begin{cases} \frac{1}{x_{imp} - x_{ind}} x - \frac{1}{x_{imp} - x_{ind}} x_{ind}, & \text{if } X \text{ favors } Y \\ \frac{1}{x_{ind} - x_{inc}} x - \frac{1}{x_{ind} - x_{inc}} x_{ind}, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (34) \end{aligned}$$

(b) μ is to right semihomographic normalizable:

in the case where $x_{imp} = +\infty$, that is to say, $\lim_{x \rightarrow +\infty} (x_f / (x + m)) = 0$, the expression is used:

$$F_{shnd}(x) = \begin{cases} \frac{x_f}{x + m} + y_f, & \text{if } X \text{ favors } Y \\ x_d x + y_d, & \text{if } X \text{ disfavors } Y \end{cases} \quad (35)$$

with

$$\begin{aligned} x_f &= -(x + m), \\ y_f &= 1, \\ x_d &= \frac{1}{x_{ind} - x_{inc}} \end{aligned} \quad (36)$$

and

$$y_d = -\frac{1}{x_{ind} - x_{inc}} x_{ind} \quad (37)$$

so

$$\begin{aligned} F_{shnd}(x) &= \begin{cases} -\frac{x_{ind} + m}{x + m} + 1, & \text{if } X \text{ favors } Y \\ \frac{1}{x_{ind} - x_{inc}} x + \frac{1}{x_{ind} - x_{inc}} x_{inc}, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (38) \end{aligned}$$

(c) μ is to left semihomographic normalizable: in the case where $x_{inc} = -\infty$; that is to say, $\lim_{x \rightarrow -\infty} (x_f / (x_{inc} + m)) = 0$. This time we use the following expression:

$$F_{shng}(x) = \begin{cases} x_f x + y_f, & \text{if } X \text{ favors } Y \\ \frac{x_d}{x + m} + y_d, & \text{if } X \text{ disfavors } Y \end{cases} \quad (39)$$

such that

$$\begin{aligned} x_f &= \frac{1}{x_{imp} - x_{ind}}, \\ y_f &= -\frac{1}{x_{imp} - x_{ind}} x_{ind}, \\ x_d &= x_{ind} + m \end{aligned} \quad (40)$$

and

$$y_d = -1 \quad (41)$$

so

$$\begin{aligned} F_{shnd}(x) &= \begin{cases} \frac{1}{x_{imp} - x_{ind}} x - \frac{1}{x_{imp} - x_{ind}} x_{ind}, & \text{if } X \text{ favors } Y \\ \frac{x_{ind} + m}{x + m} - 1, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (42) \end{aligned}$$

(d) μ is homographic normalizable:

in the case where $x_{inc} = -\infty$, $x_{imp} = +\infty$ and $x_{ind} \in \mathbb{R}^*$; that is to say,

$$\lim_{x \rightarrow -\infty} \frac{x_f}{x_{inc} + m} = 0 \quad (43)$$

and

$$\lim_{x \rightarrow -\infty} \frac{x_f}{x_{inc} + m} = 0. \quad (44)$$

Here we use the following expression:

$$F_{hn}(x) = \begin{cases} \frac{x_f}{x + m} + y_f, & \text{if } X \text{ favors } Y \\ \frac{x_d}{x + m} + y_d, & \text{if } X \text{ disfavors } Y \end{cases} \quad (45)$$

where

$$\begin{aligned} x_f &= -(x_{ind} + m), \\ y_f &= 1, \\ x_d &= (x_{ind} + m) \end{aligned} \quad (46)$$

and

$$y_d = -1, \quad (47)$$

so

$$F_{hn}(x) = \begin{cases} -\frac{x_{ind} + m}{x + m} + 1, & \text{if } X \text{ favors } Y \\ \frac{x_{ind} + m}{x + m} - 1, & \text{if } X \text{ disfavors } Y. \end{cases} \quad (48)$$

(ii) (a) It is seen that if $x \in \mathbb{R}$ and $(x + m) \in \mathbb{R}^*$, then the four terms of the function of normalizations are well defined; therefore, there is no problem for calculating the normalization coefficients.

(b) We see that if $x = \infty$, then we can always get a projective application $\overline{\mathbb{R}}$ in \mathbb{R} for the functions F_{shn} , F_{shnd} , and F_{shng} and therefore the four coefficients of normalization are always calculable; by constraint, if $x = \infty$, then you can never get a projective application over the interval $\overline{\mathbb{R}}$ in \mathbb{R} for the function F_{an} ; therefore, we can not calculate these four coefficients. The theorem is stated. \square

Table 1 recalls the respective definitions of the various measures.

4. Conclusion and Perspectives

This study showed that normalization of probabilistic quality measures with a homographic homeomorphism is more powerful than the normalization homeomorphism refines initiated by André Totohasina. Indeed, we showed that any measure affine-normalizable is homographic normalizable, while the converse is false. Besides, this work has explained the process of normalization by homographic function and combination with an affine function by trying to sweep the present main measures in the literature with the aim of a presentation easier to understand. The database has several branches; the purpose of this research is the normalization of quality measures. We always say, in the context of the database, the study on association rules knows an important development, added to the measures called interest; yet probabilistic quality measures have an important place in the context of data mining. Thereafter, the probabilistic measure of quality and its normalization must be complement. As shown by research on normalization probabilistic quality measures realizing the normalization operation requires passing through a relatively complex theory. We can consider several possible ways to carry out its standardization process. In our opinion, the use of normalizing function seems the simplest way.

In future work, we understand the positive impact of consideration of these normalizing functions in the development of the bases of the rules in search of binary data.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

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, 1993.
- [2] M. Kamber and R. Shinghal, "Proposed interestingness measure for characteristic rules," *Review*, vol. 20, no. 1, p. 21, 1996.
- [3] D. Grissa, *Etude comportementale des mesures d'intérêt d'extraction de connaissance [Ph.D. thesis]*, Université Clermont-Ferrand II, 2013.
- [4] A. Totohasina, "Towards a theory unifying implicative interestingness measures and critical values consideration in MGK," in *Educ. Matem. Pesq.*, vol. 16 no 3, pp. 881–900, São Paulo, 2014.
- [5] G. Yongmei and B. Fuguang, "The Research on Measure Method of Association Rules Mining," *Journal of Database Theory and Application*, vol. 8, no. 2, pp. 245–258, 2015.
- [6] H. F. Rakotomalala, B. B. Ralahady, and A. Totohasina, "A Novel cohesive implicative classification based on MGK and application on diagnostic on informatics literacy of students of higher education in Madagascar," in *Third International Congress on Information and Communication Technology*, vol. 797 of *Advances in Intelligent Systems and Computing*, pp. 161–174, Springer Singapore, Singapore, 2019.
- [7] B. Shekar and R. Natarajan, "A Transaction-Based Neighbourhood-Driven Approach to Quantifying Interestingness of Association Rules," in *Proceedings of the Fourth IEEE International Conference on Data Mining (ICDM'04)*, pp. 194–201, Brighton, UK.
- [8] X. Wu, C. Zhang, and S. Zhang, "Efficient mining of both positive and negative association rules," *ACM Transactions on Information and System Security*, vol. 22, no. 3, pp. 381–405, 2004.
- [9] L. Nguyen, B. Vo, and T.-P. Hong, "CARIM: An efficient algorithm for mining class-association rules with interestingness measures," *International Arab Journal of Information Technolog*, vol. 12, no. 6A, pp. 627–634, 2015.
- [10] P. Tan, V. Kumar, and J. Srivastava, "Selecting the right interestingness measure for association patterns," in *Proceedings of the eighth ACM SIGKDD international conference*, p. 32, Edmonton, Alberta, Canada, July 2002.
- [11] Dr. Niket Bhargava and Manoj. 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 (IJCSITS)*, vol. 6, no. 2, pp. 2249–2249-9555, 2016.
- [12] J. Diatta, H. Ralambondrainy, and A. Totohasina, "Towards a unifying probabilistic implicative normalized quality measure for association rules," *Studies in Computational Intelligence*, vol. 43, pp. 237–250, 2007.
- [13] A. Totohasina, *Contribution à l'étude des mesures de qualité des règles d'associations: normalisation sous cinq contraintes et cas de MGK: propriétés, bases composites des règles et extension en vue d'applications en statistique et en sciences physiques [Ph.D. thesis]*, Université, d'Antsiranana, Madagascar: Mathématiques Informatique, Madagascar, 2008.
- [14] P. Bemariska and A. Totohasina, "Optimized mining of potential positive and negative association rules," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Preface*, vol. 10440, pp. 424–432, 2017.
- [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, Cameroon, December 2005.
- [16] M. Hahsler and K. Hornik Vienna, *New Probabilistic Interest Measures for Association Rules*, University of Economics and Business Administration, Augasse 26, A-1090 Vienna, Austria, 2018.
- [17] L. Lin, M.-L. Shyu, and S.-C. Chen, "Association rule mining with a correlation-based interestingness measure for video semantic concept detection," *International Journal of Information and Decision Sciences*, vol. 4, no. 2-3, pp. 199–216, 2012.
- [18] A. Armand, *Totohasina and Feno Daniel Rajaonasy. An extension of Totohasina's normalization theory of quality measures of association rules*, 2018, In press.

