On normalizing fuzzy coincidence matrices to compare fuzzy and/or possibilistic partitions with the Rand index

Romain Quéré, Hoel Le Capitaine, Noel Fraisseix, Carl Frélicot

On normalizing fuzzy coincidence matrices to compare fuzzy and/or possibilistic partitions with the Rand index

R. Quéré, H. Le Capitaine, N. Fraisseix, C. Frélicot
Mathématiques, Image et Applications
Université de La Rochelle, France
$\{$*romain.quere,hoel.le-capitaine01,noel.fraisseix,carl.frelicot*$\}$*@univ-lr.fr*

***Abstract*—Most already existing indices used to compare two strict partitions with different number of clusters are based on coincidence matrices. To extend such indices to fuzzy partitions, one can define fuzzy coincidence matrices by means of triangular norms. It has been shown this can require some kind of normalization to reinforce the corresponding indices. We propose in this paper a generic solution to perform this normalization considering the generators of the used triangular norms. Although the solution is not index-dependant, we focus on the Rand index and some of its fuzzy counterparts.**

***Keywords*-coincidence matrix; fuzzy partition; triangular norms; Rand index**

I. Introduction

In unsupervised classification, partitioning a set X of n objects into c clusters is a common task. It results in a partition matrix U, whose general term u_{ik} represents the membership degree of the k^{th} object $\mathbf{x_k}$ to the i^{th} cluster. A partition matrix can take values in numerous spaces. We will distinguish the main ones [1], called the sets of *possibilistic*, *fuzzy* and *crisp (or hard)* partition matrices, respectively : $\mathcal{M}_{pcn} = \{U \in \mathbb{R}^{cn} : u_{ik} \in [0,1]\}$, $\mathcal{M}_{fcn} = \{U \in \mathcal{M}_{pcn} : \sum_{i=1}^{c} u_{ik} = 1\}$, $\mathcal{M}_{hcn} = \{U \in \mathcal{M}_{fcn} : u_{ik}$ $u_{ik} \in \{0,1\}\}$. For sake of simplicity we identify a *partition* with the corresponding *partition matrix*. Choosing the best partition, according to the feature space of objects, the number of clusters or the parameters of a specific algorithm is a problem of great interest. A solution consists in computing a concordance measure between a partition and the data called *internal index* [2]. When one wants to evaluate the agreement with an expert-assessed and supposed ground-truth reference partition, an *external index* will be preferred [3]. This last approach is generalized to the comparison of two partitions obtained, for instance, with two different algorithms or the same algorithm with two different parameterizations and has lead to the definition of numerous *relational indices*, topic of this paper.

Historically defined for crisp partitions, numerous indices were recently proposed for the comparison of fuzzy partitions, in order to take into account the non-exclusive membership of objects to clusters. We will distinguish indices relying on a direct approach from those extending crisp indices. Among the first ones, let us cite those constructed from a measure of similarity based on the comparison of the whole set of α-cuts of the two partitions [4], a pseudo-distance between rows of partitions matrices viewed as fuzzy sets [5], or a simple and intuitive extension of the transfer distance to the fuzzy case [6]. Most of the crisp indices are based on the contingency matrix of pairs of objects belonging or not to the same cluster in both partitions to be compared, defined from the coincidence matrices of the two partitions. Most of fuzzy indices of the second type aim at extending the definition of the crisp coincidence matrices with fuzzy set theory tools that allow to model set-theoretic operations [3]. A majority of authors has used triangular norms for this purpose [7]. Unfortunately, this kind of construction can lead to counterintuitive values for the coincidence matrix elements, so that they have to be normalized. In this paper, we propose a generic solution to perform this normalization considering either the additive or the multiplicative generator of the used triangular norm.

The remaining part of this paper is organized as follows. In Section II, we first recall the basics of relative indices based on contingence matrices for crisp partitions and give the definition of one of them, specifically the Rand index. Then, two particular extensions to fuzzy partitions are reviewed, both based on a fuzzy contingency matrix defined by means of triangular norms. Section III is concerned by the normalization problem of fuzzy coincidence matrices. The generic solution we propose is given, the derivation for most of triangular norms is provided, and the behavior of the resulting Rand indices is discussed considering simple examples of strict, fuzzy and possibilistic partitions. In Section IV, we report experimental results which show the pertinence of the proposed solution. The conclusion and some perspectives for future work are drawn in Section V.

II. Contingency Indices

A. Strict Indices

The $(n \times n)$ coincidence matrix Ψ_U of $(\mathbf{x}_k, \mathbf{x}_l)$ pairs of objects of general term $\psi_{U,kl}$, associated with a crisp partition $U \in \mathcal{M}_{hc_u n}$, is defined by:

$$\Psi_U = {}^tU\,U. \tag{1}$$

Given another crisp partition $V \in \mathcal{M}_{hc_v n}$ associated with Ψ_V, the contingency matrix C of $q = \frac{n\,(n-1)}{2}$ different pairs of objects, crossing U and V, is defined by:

$$C(U,V) = \begin{pmatrix} n_{11} & n_{10} \\ n_{01} & n_{00} \end{pmatrix} \tag{2}$$

where $n_{\alpha\,\beta}$ $(\alpha, \beta = 0, 1)$ represent the number of pairs (k, l), $k < l$ such that $\psi_{U,kl} = \alpha$ AND $\psi_{V,kl} = \beta$. Is is easy to show that $n_{00} + n_{01} + n_{11} + n_{10} = q$.
From the contingency matrix, numerous indices have been defined to compare two strict partitions, *e.g.* in [8], [9]. The most known and probably most controversial one is the *Rand Index*, taking values in [0,1] and defined by:

$$RI(U, V) = \frac{n_{00} + n_{11}}{n_{00} + n_{01} + n_{11} + n_{10}}. \tag{3}$$

which is maximum when $U = V$ since $n_{11} = q$.

Example 1: Consider two partitions $U_h = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $V_h = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$ in $\mathcal{M}_{h23}$. We have $C(U_h, U_h) = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ and $C(U_h, V_h) = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ which result in Rand index values of $RI(U_h, U_h) = 1$ and $RI(U_h, V_h) = 0.33$ as expected.

The main problem of RI is that its expected value $E(RI)$ is not null when both partitions are drawn at random, a behavior obviously due to a certain number of agreements brought by chance. For crisp partitions, the *Adjusted Rand Index* [10] overcomes this drawback and is generally preferred. However, it is based on $E(RI)$ under an assumption which only holds for integer values, making it formally valid for crisp partitions comparison but not valid for fuzzy or possibilistic partitions, so we will use the classical RI instead. Many other indices exist, *e.g.* in [7], [11], [12].

B. Extended Indices by Means of Triangular Norms

A triangular norm (*t-norm*) is a commutative, associative and monotonic function $\top : [0, 1]^2 \rightarrow [0, 1]$, satisfying $\top(a, 1) = a$. Some extensions also use the triangular conorm (*t-conorm*) which is the dual operator $\perp$ of a t-norm with respect to the usual fuzzy complement. Basic and main parametrized families of couples $(\top_\lambda, \perp_\lambda)$ are given in Table I, see [13] for an extensive review. Parametric couples allow to control the way the values are aggregated and special values of λ make the couples correspond to some basic ones. For instance, taking $\lambda = 1$, both $\top_{AA_1}$ and $\top_{H_1}$ t-norms reduce to $\top_P$, and both $\top_{SS_1}$ and $\top_{Y_1}$ t-norms reduce to $\top_L$. For a fuzzy partition $U \in \mathcal{M}_{fc_u n}$, the general term $\psi_{U,kl}$ of the associated coincidence matrix (1) is no more a sum of binary products with output in $\{0, 1\}$ but in [0,1]. Since the product is a particular t-norm ($\top_P$), the most common approach consists in extending the product to any t-norm $\top$. Following Borgelt's notations [3], one can replace Ψ_U by $\Psi_U^\top$, of general term

$$\psi_{U,kl}^\top = \sum_{i=1}^{c_u} \top(u_{ik}, u_{il}). \tag{4}$$

Thus, all strict comparison indices own their fuzzy extension, computed from a fuzzy contingency matrix $C_\top(U, V)$, whose elements $n_{\alpha\,\beta}^\top$ $(\alpha, \beta = 0, 1)$ are no more integers but fuzzy cardinalities defined by:

Table I
BASIC TRIANGULAR NORMS AND MAIN PARAMETRIZED FAMILIES.

Standard	$\top_M(a, b) = \min(a, b)$
Produit	$\top_P(a, b) = a\,b$
Łukasiewicz	$\top_L(a, b) = \max(a + b - 1, 0)$

Aczel-Alsina $\lambda \in \mathbb{R}_+^\star$	$\top_{AA_\lambda}(a, b) = e^{-\left((-\ln a)^\lambda + (-\ln b)^\lambda\right)^{1/\lambda}}$
Dombi $\lambda \in \mathbb{R}_+^\star$	$\top_{D_\lambda}(a, b) = \left(1 + \left(\left(\frac{1-a}{a}\right)^\lambda + \left(\frac{1-b}{b}\right)^\lambda\right)^{1/\lambda}\right)^{-1}$
Frank $\lambda \in \mathbb{R}^{+\star}\backslash[1]$	$\top_{F_\lambda}(a, b) = \ln\left(1 + \frac{(\lambda^a - 1)\,(\lambda^b - 1)}{\lambda - 1}\right) / \ln \lambda$
Hamacher $\lambda \in \mathbb{R}_+$	$\top_{H_\lambda}(a, b) = a\,b / \,(\lambda + (1 - \lambda)\,(a + b - a\,b))$
Schweizer-Sklar $\lambda \in \mathbb{R}^\star$	$\top_{SS_\lambda}(a, b) = \left(\max\left(a^\lambda + b^\lambda - 1, 0\right)\right)^{1/\lambda}$
Sugeno-Weber $\lambda \in]-1, +\infty[$	$\top_{SW_\lambda}(a, b) = \max\left(\frac{a+b-1+\lambda a b}{1+\lambda}, 0\right)$
Yager $\lambda \in \mathbb{R}_+^\star$	$\top_{Y_\lambda}(a, b) = \max\left(1 - ((1 - a)^\lambda + (1 - b)^\lambda)^{1/\lambda}, 0\right)$

$$n_{\alpha\,\beta}^\top(\Psi_U^\top, \Psi_V^\top) = \sum_{k=2}^{n} \sum_{l=1}^{k-1} \top\big((1 - \alpha) + (2\,\alpha - 1)\ \psi_{U,kl}^\top, \\ (1 - \beta) + (2\,\beta - 1)\ \psi_{V,kl}^\top\big) \tag{5}$$

Note that (4) does not restrict to fuzzy partitions but also holds to any partition in $\mathcal{M}_{pc_u n}$ or $\mathcal{M}_{hc_u n}$. If U and V are crisp partitions, the Ψ_U and Ψ_V are binary, $C_\top(U, V) = C(U, V)$, and the resulting indices are exactly the strict ones, *i.e.* $RI^\top(U, V) = RI(U, V)$. For comparison purpose, we need to briefly recall the set-theoretic based extension of the contingency matrix $C(U, V)$ proposed by Campello [7]. Since $n_{\alpha\,\beta}$ are the cardinalities of sets $U_\alpha \cup V_\beta$, where U_1 (U_0) is the set of pairs of objects belonging to the same cluster (different clusters) in U, and V_1 (V_0) the counterpart for V, the author proposes to model the belongingness to clusters, union of sets and cardinality of sets by use of a t-norm, a t-conorm and the sigma count operator. Thus, $C(U, V)$ is extended to $C^f(U, V)$ by:

$$n_{\alpha\,\beta}^f(U, V) = \sum_{k=2}^{n} \sum_{l=1}^{k-1} \top\Big(\perp^{c_U}_{\substack{i,\,j\,=\,1;\\ \alpha\,=\,1,\,j\,=\,i;\\ \alpha\,=\,0,\,j\,\neq\,i}} U_\alpha^{ij}(k, l), \\ \perp^{c_V}_{\substack{i,\,j\,=\,1;\\ \beta\,=\,1,\,j\,=\,i;\\ \beta\,=\,0,\,j\,\neq\,i}} V_\beta^{ij}(k, l)\Big) \tag{6}$$

where U_1^{ii} and V_1^{ii} are the fuzzy sets of pairs (k, l) of objects belonging to the i^{th} cluster in U and V respectively, and U_0^{ij} and V_0^{ij} the fuzzy sets of pairs (k, l) of objects belonging to the i^{th} and j^{th} clusters (with $i \neq j$, such as they are in different clusters) in U and V respectively, see details in [7]. As for the previous construction, all strict comparison indices own their fuzzy counterpart, in particular the Rand Index to which the author refers as *Fuzzy Rand Index*. We will denote it $RI_f^\top$ for writing convenience. Note that $RI_f^\top(U, U) = 1$ only if $U \in \mathcal{M}_{hc_u n}$, so it is recommended to use this index to compare a fuzzy partition to a crisp reference one. Recently, Campello's construction has been strongly criticized, in particular because the underlying topological relationships existing in partitions are not taken into account [5]. Regardless of the pertinence of this fact, we will use it as a reference, because it is the most different (by construction) but comparable approach (contingency matrix) to Borgelt's and our ones, while the alternative in [5] is not. To conclude this section, one should notice that nothing prohibits the use of different (families of) t-norms and t-conorms for the computations of the intermediate terms involved in both Borgelt's and Campello's constructions leading respectively to the $RI_\top$ and $RI_\top^f$ indices, either to counterpart/reinforce the behavior of each operator or to underline some situation. This could be the topic of a whole study, and since it would result in an important number of combinations, we restrict ourselves to a single t-norm (and the corresponding dual t-conorm if needed) at a time.

III. A Solution to the Coincidence Matrix Normalization Problem

A. Motivation

Extending comparison indices as shown in the previous section may produce undesirable results. Let us consider the fuzzy coincidence matrix as defind by (4). Whatever the t-norm, diagonal terms $\psi_{U,kk}^\top$ representing the degree with which each $\mathbf{x}_k$ is as in the same class as itself, are no more equal to 1. In [9], where the extension of indices to fuzzy partitions is formalized with matrices and consequently only the product is used, Brouwer proposes to replace the inner product of rows of U with the *cosine-correlation*. It is easy to see that it means replacing $\psi_{U,kl}^\top$ with $\top_P$ in the computation of Ψ_U by:

$$\phi_{U,kl}^{\top_P} = \frac{\psi_{U,kl}^{\top_P}}{\sqrt{\psi_{U,kk}^{\top_P}}\sqrt{\psi_{U,ll}^{\top_P}}}. \quad (7)$$

We denote the transformed coincidence matrix by $\Phi_U^{\top_P}$. As the author notices it, this transformation does not affect the resulting comparison indices if the partitions are crisp, simply because it does not modify the coincidence matrices.

Example 2: Consider $U_f = \begin{pmatrix} 0.7 & 0.9 & 0.1 \\ 0.3 & 0.1 & 0.9 \end{pmatrix}$ in $\mathcal{M}_{f23}$; $\Psi_{U_f}^{\top_P} = \begin{pmatrix} 0.58 & 0.66 & 0.34 \\ 0.66 & 0.82 & 0.18 \\ 0.34 & 0.18 & 0.82 \end{pmatrix}$ and $\Phi_{U_f}^{\top_P} = \begin{pmatrix} 1 & 0.96 & 0.49 \\ 0.96 & 1 & 0.22 \\ 0.49 & 0.22 & 1 \end{pmatrix}$.

Not only the diagonal terms change because of interactions. Our proposition is to generalize this idea to the coincidence matrices extended by any t-norm.

B. Solution and Properties

In order to obtain, from a fuzzy coincidence matrix $\Psi_U^\top$, a matrix $\Phi_U^\top$, normalized in the sense that all its diagonal terms $\phi_{U,kk}^\top$ are equal to 1, for any t-norm, one only has to find a function $K_\top(a) \ : \ [0,1] \to [0,1]$ such that:

$$\begin{aligned} \frac{a}{\top\big(K_\top(a), K_\top(a)\big)} &= 1 \\ \Leftrightarrow \top\big(K_\top(a), K_\top(a)\big) &= a. \end{aligned} \quad (8)$$

Given $(\top, K_\top)$, (7) is then easily generalized by:

$$\phi_{U,kl}^\top = \frac{\psi_{U,kl}^\top}{\top\left(K_\top\left(\psi_{U,kk}^\top\right), K_\top\left(\phi_{U,ll}^\top\right)\right)}. \quad (9)$$

This transformation does not affect the indices if partitions are crisp. Indeed, if $U \in \mathcal{M}_{hc_un}$, then $\psi_{U,kk}^\top = \psi_{U,ll}^\top = 1$ and consequently, the denominator equals 1 by (8).

Proposition 1: For the basic *standard*, *product* and *Łukasiewicz* t-norms, the normalizing functions $K_\top$ are : $K_{\top_M}(a) = a$, $K_{\top_P}(a) = \sqrt{a}$ and $K_{\top_L}(a) = \frac{a+1}{2}$.

Proposition 2: Given an archimedean[1] t-norm $\top$, with additive generator $f_\top$ or multiplicative generator $g_\top$, the normalizing function $K_\top$ such that (8) is

$$K_\top(a) = \begin{cases} f_\top^{-1}\left(\frac{f_\top(a)}{2}\right) \\ g_\top^{-1}\left(\sqrt{g_\top(a)}\right) \end{cases} \quad (10)$$

where $f_\top^{-1}$ and $g_\top^{-1}$ are the pseudo-inverses of $f_\top$ and $g_\top$.

The normalizing functions of parametrized families of t-norms of Table I, obtained with their additive generator f are given in Table II, proofs being left to a long forthcoming paper. In the case where the values of λ make the t-norm to be equal (or tend) to a basic one or another, provided it is archimedean ($\top_P$, $\top_L$), the corresponding normalizing function is recognized, *e.g.*: $K_{\top_{AA_1}}(a) = a^{\frac{1}{2}} = K_{\top_P}(a)$, $K_{\top_{H_1}}(a) = \sqrt{a} = K_{\top_P}(a)$, $K_{\top_{SS_1}}(a) = \frac{a+1}{2} = K_{\top_L}(a)$, $K_{\top_{Y_1}}(a) = 1 - \frac{1-a}{2} = K_{\top_L}(a)$. For values of λ out of the ranges specified in Table I, one must use the corresponding generator or compute the limits. We illustrate the first alternative by the Sugeno-Weber t-norm, which leads to the Łukasiewicz one if $\lambda = 0$. The $\top_{SW_0}$ additive generator and its pseudo-inverse are defined by $f(a) = f^{(-1)}(a) = 1 - a$ and (10) gives: $K_{\top_{SW_0}}(a) = 1 - \frac{1-a}{2}$, actually equals $K_{\top_L}(a)$. As an example of limits computation, consider the Frank t-norm which is equal to the standard, the product and the Łukasiewicz basic t-norms when λ approaches 0, 1 and $+\infty$ respectively. By using the Taylor series if needed, it is easy to prove that: $K_{\top_{F_0}}(a) \to a = K_{\top_M}(a)$ if $\lambda \to 0$, $K_{\top_{F_1}}(a) \to \sqrt{a} = K_{\top_P}(a)$ if $\lambda \to 1$, $K_{\top_{F_{+\infty}}}(a) \to \frac{a+1}{2} = K_{\top_L}(a)$ if $\lambda \to +\infty$. The same property holds for relations between parametrized families of t-norms. For example, Hamacher and Dombi t-norms are equals if their respective parameter value is $\lambda = 0$ and $\lambda = 1$. From the generator and the pseudo-inverse of $\top_{H_0}$, defined by $f(a) = \frac{1-a}{a}$ and $f^{-1}(a) = \frac{1}{1+a}$, we obtain by (10): $K_{\top_{H_0}}(a) = \frac{1}{1+\frac{1-a}{2a}}$, which is equal to $K_{\top_{D_1}}(a)$.

Example 3: Consider U_f of Ex. 2. With $\top_{H_0}$, we have $\Psi_{U_f}^{\top_H} = \begin{pmatrix} 0.71 & 0.73 & 0.39 \\ 0.73 & 0.87 & 0.20 \\ 0.39 & 0.20 & 0.87 \end{pmatrix}$ and $\Phi_{U_f}^{\top_H} = \begin{pmatrix} 1 & 0.93 & 0.49 \\ 0.93 & 1 & 0.23 \\ 0.49 & 0.23 & 1 \end{pmatrix}$.

[1] a t-norm $\top$ is archimedean if $\top(a,a) < a$ for all $a \in [0,1]$

Table II
ADDITIVE GENERATORS AND PSEUDO-INVERSES OF PARAMETRIZED FAMILIES OF ARCHIMEDEAN T-NORMS, AND NORMALIZING FUNCTIONS

t-norm	$f(a)$	$f^{-1}(a)$	$K_\top(a)$
$\top_{AA}$	$(-\ln a)^\lambda$	$e^{-a^{1/\lambda}}$	$a^{\left(\frac{1}{2}\right)^{1/\lambda}}$
$\top_D$	$\left(\frac{1-a}{a}\right)^\lambda$	$\left(1+a^{1/\lambda}\right)^{-1}$	$\left(1+\left(\frac{1}{2}\right)^{1/\lambda}\ \frac{1-a}{a}\right)^{-1}$
$\top_F$	$\ln\left(\frac{\lambda-1}{\lambda^a-1}\right)$	$\frac{\ln\left(1+(\lambda-1)\,e^{-a}\right)}{\ln\lambda}$	$\frac{\ln\left((\lambda-1)\sqrt{\frac{\lambda^a-1}{\lambda-1}}+1\right)}{\ln\lambda}$
$\top_H$	$\ln\left(\frac{\lambda+(1-\lambda)a}{a}\right)$	$\frac{\lambda}{e^a+\lambda-1}$	$\frac{\lambda\sqrt{a}}{\sqrt{\lambda+(1-\lambda)\,a}+(\lambda-1)\sqrt{a}}$
$\top_{SS}$	$\frac{1-a^\lambda}{\lambda}$	$(1-\lambda\,a)^{1/\lambda}$	$\left(\frac{a^\lambda+1}{2}\right)^{1/\lambda}$
$\top_{SW}$	$1-\frac{\ln(1+\lambda\,a)}{\ln(1+\lambda)}$	$\frac{(1+\lambda)^{1-a}-1}{\lambda}$	$\frac{\sqrt{(1+\lambda)(1+\lambda\,a)}-1}{\lambda}$
$\top_Y$	$(1-a)^\lambda$	$1-a^{1/\lambda}$	$1-\frac{1-a}{2^{1/\lambda}}$

The proposed generic normalization allows to transform the fuzzy cardinalities $n^\top_{\alpha\,\beta}(\Psi^\top_U,\Psi^\top_V)$ given by (5) by $n^\top_{\alpha\,\beta}(\Phi^\top_U,\Phi^\top_V)$. For any $(\top, K_\top)$-combination, one can derive new versions of every existing comparison index, provided it is based on a contingency matrix. We denote $RI^\top_K$ the so derived Rand index.

C. Numerical Comparisons of Crisp, Fuzzy and Possibilistic Partitions

We discuss here the influence of the normalization on the Rand index by comparing the resulting $RI^\top_K$ to the non normalized $RI^\top$ presented in Section II. Five compatible partitions of different kind are chosen for this purpose:

- $U_1 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \in \mathcal{M}_{h34}$,
- $U_2 = \begin{pmatrix} 0.1 & 0.2 & 0.1 & 0.7 \\ 0.9 & 0.8 & 0 & 0.3 \\ 0 & 0 & 0.9 & 0 \end{pmatrix} \in \mathcal{M}_{f34}$,
- $U_3 = \begin{pmatrix} 0.28 & 0.3 & 0.28 & 0.43 \\ 0.47 & 0.45 & 0.25 & 0.32 \\ 0.25 & 0.25 & 0.47 & 0.25 \end{pmatrix} \in \mathcal{M}_{f34}$,
- $U_4 = \begin{pmatrix} 0.05 & 0.1 & 0.05 & 0.35 \\ 0.45 & 0.4 & 0 & 0.15 \\ 0 & 0 & 0.45 & 0 \end{pmatrix} \in \mathcal{M}_{p34}$,
- $U_5 = \begin{pmatrix} 0.140 & 0.150 & 0.140 & 0.215 \\ 0.235 & 0.225 & 0.135 & 0.160 \\ 0.125 & 0.125 & 0.235 & 0.125 \end{pmatrix} \in \mathcal{M}_{p34}$.

These matrices are chosen such that:

- U_1 is the closest crisp partition w.r.t. the others,
- between the two fuzzy partitions, U_2 is closer to U_1 than U_3 because it is less fuzzy,
- between the two possibilistic partitions, U_4 is closer to U_1 than U_5 because of a bigger gap between its membership degrees,
- the possibilistic partition U_4 is closer to U_1 than the fuzzy one U_3 for the same reason,
- due to their construction, the fuzzy partitions U_2 and U_3 are quite close to the possibilistic ones $U_4 = \frac{U_2}{2}$ and $U_5 = \frac{U_3}{2}$, the closeness between U_3 and U_5 being more significant because the differences between their values are even smaller.

The results obtained on all partition couples (U_i, U_j) are reported in Table III, for two basic t-norms ($\top_M$, $\top_P$) and two parametric ones ($\top_H$, $\top_F$) with two values of λ. In each cell of each symmetric subtable, the upper (resp. lower) value refers to $RI^\top$ (resp. $RI^\top_K$) computed from $(\Psi^\top_{U_i}, \Psi^\top_{U_j})$) (resp. $(\Phi^\top_{U_i}, \Phi^\top_{U_j})$ normalized by the $K_\top$ functions). On can observe several specific situations:
1) increased values, sometimes in a significant way (depending on the t-norm) when comparing fuzzy partitions with themselves, *e.g.* $RI^\top_K(U_3, U_3) = 0.89\ vs\ RI^\top(U_3, U_3) = 0.51$ with $\top_{H_0}$,
2) either the same behavior or a slight but not significant decrease (high magnitudes), when comparing possibilistic partitions with themselves, *e.g.* $RI^\top_K(U_5, U_5) = 0.76\ vs\ RI^\top(U_5, U_5) = 0.54$ with $\top_M$, and $RI^\top_K(U_4, U_4) = 0.79\ vs\ RI^\top(U_4, U_4) = 0.89$ with $\top_P$,
3) the same behavior is met when comparing close partitions of a different type, a fuzzy and a possibilistic one, *e.g.* $RI^\top_K(U_3, U_5) = 0.90\ vs\ RI^\top(U_3, U_5) = 0.47$ with $\top_{H_0}$, and $RI^\top_K(U_3, U_5) = 0.85\ vs\ RI^\top(U_3, U_5) = 0.87$ with $\top_{H_5}$,
4) more or less decreased values when comparing the crisp partition U_1 with the fuzzy and possibilistic ones, an extremely weakness arising for the closest partitions (U_2, U_4) and becoming greater for the less close partitions (U_3, U_5) for any t-norm, *e.g.* $RI^\top_K(U_1, U_2) = 0.79\ vs\ RI^\top(U_1, U_2) = 0.82$ with $\top_P$, and $RI^\top_K(U_1, U_5) = 0.26\ vs\ RI^\top(U_1, U_5) = 0.81$ with $\top_{F_5}$.
The above examples clearly show the interest of the normalization, which allows to rise and lower significantly the indices with pertinence, for some low cost weakenings. With $\top_M$, the values of $RI^\top_K$ do not differ from those obtained with $RI^\top$ since U_4 and U_5 are not involved. This result is explained logically by the fact that for crisp and fuzzy partitions, we have $\sum_{i=1}^{c_u} u_{ik} = 1$ and thus $\psi^{\top_M}_{U,kk} = 1$ and finally $\phi^{\top_M}_{U,kl} = \psi^{\top_M}_{U,kl}$. In the case of possibilistic partitions, the reinforcement or the weakening of $RI^\top_K$ compared to $RI^\top$ depends directly on the membership degrees, and thus on the elements of the partitions. For $\top_M$ and $\top_P$, one can also establish that $RI^\top_K(U_i, U_4) = RI^\top_K(U_i, U_2)$ and $RI^\top_K(U_i, U_5) = RI^\top_K(U_i, U_3)$. This result comes from a trivial property of both t-norms, whose (trivial) proof is left to the reader : $RI^\top_K(U_i, U_j) = RI^\top_K(\alpha\,U_i, U_j)$. Since $\top_M$ is the largest t-norm, the values of the induced $RI^\top$ are greater than those obtained with any other t-norm. By making possible pertinent reinforcements, normalization yields to the loss of this property for $RI^\top_K$. As a final result, let us consider, given a parametric t-norm of Table III, the indices as a function of λ, and let $D = \sum |RI^\top(\lambda_1) - RI^\top(\lambda_2)|$ and $D_K = \sum |RI^\top_K(\lambda_1) - RI^\top_K(\lambda_2)|$. For both Hamacher and Frank t-norms, D is about four times D_K. This shows that, for parametrized families of t-norms, the proposed $RI^\top_K$ is less sensitive to the choice of λ than $RI^\top$.

Table III
RAND INDEX $RI^{\top}$ VALUES (UP) AND NORMALIZED ONES $RI_K^{\top}$ (DOWN) FOR THE COMPARISON OF FIVE PARTITIONS WITH VARIOUS T-NORMS

$\top_M$	U_1	U_2	U_3	U_4	U_5
U_1	1.00 1.00	0.78 0.78	0.32 0.32	0.81 0.78	0.58 0.32
U_2		0.70 0.70	0.42 0.42	0.67 0.70	0.55 0.42
U_3			0.76 0.76	0.29 0.42	0.44 0.76
U_4				0.74 0.70	0.56 0.42
U_5					0.54 0.76

$\top_{H_0}$	U_1	U_2	U_3	U_4	U_5
U_1	1.00 1.00	0.79 0.77	0.46 0.22	0.80 0.73	0.66 0.21
U_2		0.70 0.73	0.48 0.43	0.68 0.72	0.58 0.42
U_3			0.51 0.89	0.45 0.46	0.47 0.90
U_4				0.78 0.72	0.67 0.45
U_5					0.61 0.91

$\top_{F_{0.1}}$	U_1	U_2	U_3	U_4	U_5
U_1	1.00 1.00	0.80 0.78	0.52 0.23	0.82 0.77	0.74 0.23
U_2		0.71 0.75	0.52 0.42	0.70 0.76	0.64 0.43
U_3			0.50 0.88	0.52 0.43	0.51 0.88
U_4				0.83 0.76	0.77 0.43
U_5					0.73 0.88

$\top_P$	U_1	U_2	U_3	U_4	U_5
U_1	1.00 1.00	0.82 0.79	0.62 0.25	0.83 0.79	0.78 0.25
U_2		0.75 0.79	0.59 0.42	0.74 0.79	0.70 0.42
U_3			0.56 0.86	0.65 0.42	0.64 0.86
U_4				0.89 0.79	0.86 0.42
U_5					0.85 0.86

$\top_{H_5}$	U_1	U_2	U_3	U_4	U_5
U_1	1.00 1.00	0.86 0.84	0.76 0.30	0.84 0.83	0.82 0.27
U_2		0.85 0.88	0.76 0.42	0.82 0.87	0.80 0.40
U_3			0.83 0.83	0.87 0.42	0.87 0.85
U_4				0.96 0.87	0.96 0.40
U_5					0.96 0.86

$\top_{F_5}$	U_1	U_2	U_3	U_4	U_5
U_1	1.00 1.00	0.84 0.81	0.69 0.27	0.83 0.82	0.81 0.26
U_2		0.79 0.82	0.66 0.43	0.77 0.82	0.74 0.42
U_3			0.67 0.83	0.76 0.42	0.76 0.83
U_4				0.93 0.82	0.92 0.41
U_5					0.92 0.84

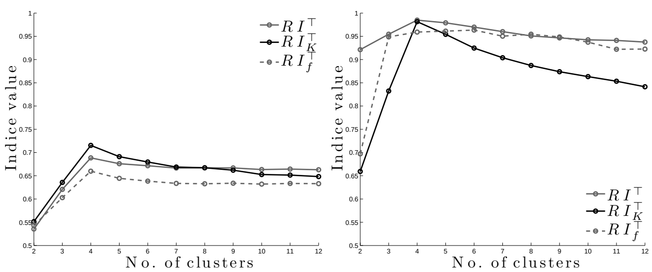

Figure 1. $RI^{\top}(R, Q_c)$, $RI_K^{\top}(R, Q_c)$ and $RI_f^{\top}(R, Q_c)$ for the data set (A), for $c = 2, 12$, with $\top_{H_0}$ (*left*) and $\top_{H_{100}}$ (*right*)

IV. EXPERIMENTAL RESULTS

A. Comparing two Fuzzy Partitions Provided by FCM

Following the work by Campello [6], we first consider a 2-dimensional data set (A) composed of 400 points drawn from a mixture of 4 overlapping normal distributions of 100 points each with means $(1, 1)$, $(1, 3)$, $(3, 1)$, and $(3, 3)$ and equal covariance matrices $\Sigma = \frac{1}{2}I$. The *Fuzzy C-Means algorithm* (FCM) [1] is used to produce a reference partition R and a collection of fuzzy $c-$partitions Q_c to be compared to R, c varying from 2 to 12. The known true centers of the clusters are used to initialize FCM in order to obtain R. FCM is run 10 times to obtain 10 partitions Q_c, for each value of c. Six t-norms are used to compute index values: $\top_{H_0}$, $\top_{H_5}$, $\top_{H_{100}}$ $\top_{F_{0.1}}$, $\top_{F_1}$ and $\top_{F_{100}}$. Values of $RI^{\top}(R, Q_c)$, $RI_f^{\top}(R, Q_c)$ and $RI_K^{\top}(R, Q_c)$ are computed for all resulting Q_c, and the maximum value of each index out of 10 runs for each value of c is stored. As expected, all indices exhibit a local maximum at $c = 4$. Moreover, for all the tested t-norms, the values of $RI_K^{\top}$ overcome the values of $RI^{\top}$ and $RI_f^{\top}$. For the other values of c, relative positions of each index curve clearly depend on both the t-norm and λ. For instance, for $c = 2$, the differences between the magnitudes of the indices increase with λ in favour of $RI_K^{\top}$, in particular with the Hamacher family, as illustrated in Fig. 1. Both $RI^{\top}$ and $RI_f^{\top}$ curves have a similar shape, and the larger λ, the smaller the difference. Moreover, they share a tendency to drop slowly or not to drop since c becomes greater than the true number of clusters, whereas the proposed $RI_K^{\top}$ keeps a good dynamic. This behavior agrees with and strengthens what we observed in section III-C concerning its reinforcement ability.

B. Comparing a Fuzzy and a Possibilistic Partition Provided by FCM and PCM

Let us consider another data set (B), constructed similarly to data set (A), with less separated means: $(1, 1)$, $(1, 2.5)$, $(2.5, 1)$ and $(2.5, 2.5)$ and same Σ. The same procedure is used to produce the reference partition R. As well, the *Possibilistic C-Means* algorithm [1] is run to generate 10 possibilistic $c = 4-$partitions Q_p. The indices $RI^{\top}(R, Q_p)$, $RI_f^{\top}(R, Q_p)$ and $RI_K^{\top}(R, Q_p)$ are computed using the same 6 t-norms, and their maximum value out of the 10 partitions Q_p for each t-norm $\top$ is reported in Table IV. For each index, the average value over each t-norm, as well as the standard deviation are also reported. The average value of $RI_K^{\top}$ is significantly smaller than the others ones (0.42 vs 0.64 and 0.70). This is explained by the 10 possibilistic partitions Q_p whose values are small as compared to the values of R, so the clusters are smoother fuzzy sets. This can be connected to a previous discussion, see section III-C (case 4). Moreover, the standard deviation values show that $RI_K^{\top}$ (0.04) is less sensitive to the t-norm choice than $RI^{\top}$ (0.18) and $RI_f^{\top}$ (0.07).

C. Sensitivity to the Fuzzifier Exponent of FCM

Let us consider the well known Fisher Iris data set composed of three classes of 50 flowers each described by 4 physical attibutes. Two classes have a substantial overlap

Table IV
RAND INDEX $RI^\top$, $RI_f^\top$ AND $RI_K^\top$ VALUES WITH DIFFERENT T-NORMS

t-norm	$\top_{H_0}$	$\top_{H_5}$	$\top_{H_{100}}$	$\top_{F_{0.1}} \to \top_M$	$\top_{F_1} = \top_P$	$\top_{F_{100}} \to \top_L$	$av.$	std
$RI^\top$	0.40	0.60	0.94	0.55	0.60	0.73	0.64	0.18
$RI_f^\top$	0.54	0.63	0.73	0.55	0.57	0.62	0.70	0.07
$RI_K^\top$	0.46	0.41	0.34	0.43	0.44	0.42	0.42	0.04

in the feature space. The expert-assessed crisp partition is chosen as the reference one (R). Although it has been shown in [14] that FCM provides best results for m lying in $[1.5, 2.5]$, we run it to produce fuzzy partitions Q_m, one for each $m \in \{2, ..., 11\}$. Again, the resulting indices values are computed using the same 6 t-norms. Comments on the shape and the relative positions of indices curves remain valid, see section IV-A. In particular, the larger m, the fuzzier partitions Q_m and the larger the slope of $RI_K^\top$, whereas $RI^\top$ and $RI_f^\top$ approach a quite high asymptotic value, as shown in Fig. 2. As opposed to the first experiment, λ does not have so much influence on the indices magnitudes, and almost no influence on $RI_K^\top$. Nevertheless, it modifies the shape of the curves. The larger λ, the smoother the drop.

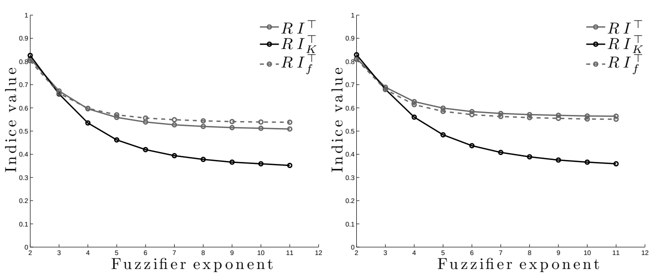

Figure 2. $RI^\top(R, Q_m)$, $RI_K^\top(R, Q_m)$ and $RI_f^\top(R, Q_m)$ for the data set (B), with different values of the fuzzifier exponent $m = 2, 11$, with $\top_{F_{0.1}} \to \top_M$ (*left*) and $\top_{F_1} = \top_P$ (*right*)

V. CONCLUSION

In this article, we propose a generic solution to perform some kind of normalization of a fuzzy coincidence matrix. It is based on the additive or the multiplicative generator of the triangular norm used to define the coincidence matrix. We derive the normalizing functions of the basic triangular norms and the main parametrized families of triangular norms. It is shown that the approach allows to correctly fullfil the demand to overcome some disagreements encountered when using non normalized coincidence matrices. The proposition enables to define new versions of any existing relative indices that are based on a contingency matrix, such as the Rand index. Results obtained on several synthetic examples and data sets show a better behavior of the so derived Rand index compared to its non normalized versions. Moreover, this statement holds for the comparison of partitions of different kind (crisp, fuzzy, possibilistic). When parametric triangular norms are used, the proposed index appears to be less sensitive to the parameter, whose adjustment can be really subtle in practice and can require a preliminary learning process.

We plan to study the empirical density of the resulting indices, according to the methodology developed in [15]. Replacing triangular norms by better adapted functions is another perspective.

REFERENCES

[1] J. Bezdek, J. Keller, R. Krishnapuram, and N. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Kluwer Academic, 1999.

[2] W. Wang and Y. Zhang, "On fuzzy cluster validity indices," *Fuzzy Sets and Syst.*, vol. 158, no. 19, pp. 2095–2117, 2007.

[3] C. Borgelt, "Finding the number of fuzzy clusters by resampling," in *16th IEEE Int. Conf. on Fuzzy Syst.*, 2006, pp. 48–54.

[4] S. Bodjanova, "Comparison of fuzzy partitions based on their α-cuts," *Fuzzy Sets and Syst.*, vol. 105, no. 1, pp. 99–112, 1999.

[5] E. Hüllermeier and M. Rifqi, "A fuzzy variant of the rand index for comparing clustering structures," in *Int. Fuzzy Syst. Association World Congress*, 2009, pp. 1294–1298.

[6] R. J. G. B. Campello, "Generalized external indexes for comparing data partitions with overlapping categories," *Pattern Recognition Letters*, vol. 31, no. 9, pp. 966–975, 2010.

[7] ——, "A fuzzy extension of the rand index and other related indexes for clustering and classification assessment," *Pattern Recognition Letters*, vol. 28, no. 7, pp. 833–841, 2007.

[8] M. Brun, C. Sima, J. Hua, J. Lowey, B. Carroll, E. Suh, and E. R. Dougherty, "Model-based evaluation of clustering validation measures," *Pattern Recognition*, vol. 40, no. 3, pp. 807–824, 2007.

[9] R. K. Brouwer, "Extending the rand, adjusted rand and jaccard indices to fuzzy partitions," *J. of Intelligent Information Syst.*, vol. 32, no. 3, pp. 213–235, 2009.

[10] L. Hubert and P. Arabie, "Comparing partitions," *J. of Classification*, vol. 2, no. 1, pp. 193–198, 1985.

[11] G. Acciani, G. Fornarelli, and L. Litturi, "Comparing fuzzy data sets by means of graph matching technique," in *Lect. Notes in Computer Science 2714*, 2003, pp. 367–374.

[12] A. G. Di Nuovo and V. Catania, "On external measures for validation of fuzzy partitions," in *Lect. Notes in Computer Science 4529*, 2007, pp. 491–501.

[13] E. Klement and R. Mesiar, *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*. Elsevier, 2005.

[14] J. C. Bezdek and N. R. Pal, "Some new indexes of cluster validity," *IEEE Trans. on Syst., Man and Cybernetics*, vol. 23, no. 3, pp. 301–315, 1998.

[15] G. Youness and G. Saporta, "Comparing partitions of two sets of units based on the same variables," *Advances in Data Analysis and Classification*, vol. 4, no. 1, pp. 53–64, 2009.