Comparing Fuzzy, Probabilistic, and Possibilistic Partitions

Derek T. Anderson, *Student Member, IEEE*, James C. Bezdek, *Life Fellow, IEEE*, Mihail Popescu, *Member, IEEE*, and James M. Keller, *Fellow, IEEE*

***Abstract*—When clustering produces more than one candidate to partition a finite set of objects O, there are two approaches to validation (i.e., selection of a "best" partition, and implicitly, a best value for c, which is the number of clusters in O). First, we may use an internal index, which evaluates each partition separately. Second, we may compare pairs of candidates with each other, or with a reference partition that purports to represent the "true" cluster structure in the objects. This paper generalizes many of the classical indices that have been used with outputs of crisp clustering algorithms so that they are applicable for candidate partitions of any type (i.e., crisp or soft, with soft comprising the fuzzy, probabilistic, and possibilistic cases). Space prevents inclusion of all of the possible generalizations that can be realized this way. Here, we concentrate on the Rand index and its modifications. We compare our fuzzy-Rand index with those of Campello, Hullermeier and Rifqi, and Brouwer, and show that our extension of the Rand index is $O(n)$, while the other three are all $O(n^2)$. Numerical examples are given to illustrate various facets of the new indices. In particular, we show that our indices can be used, even when the partitions are probabilistic or possibilistic, and that our method of generalization is valid for *any* index that depends only on the entries of the classical (i.e., four-pair types) contingency table for this problem.**

***Index Terms*—Cluster analysis, cluster validity, external-validation criteria, generalized Rand index, Rand's index.**

I. Introduction

LET $\mathbf{O} = \{o_1, \ldots, o_n\}$ denote n objects (e.g., fish, cigars, motorcycles, beers, etc.). When each object in $\mathbf{O}$ is represented by a (i.e., column) vector $\mathbf{x}$ in $\Re^p$, the set $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \subset \Re^p$ is an *object-data representation* of $\mathbf{O}$. These vectors have a variety of names in the literature, the most common being feature vectors, pattern vectors, or object vectors. The kth component of the ith feature vector ($\mathbf{x}_{ki}$) is the value of the kth feature measurement or attribute (e.g., height, weight, length, color, etc.) of the ith object.

Alternatively, when each *pair* of objects in $\mathbf{O}$ is represented by a relationship between them, then we have *relational data*. There are many names for the relation that begets relational-data values, for example, measures of proximity, agreement, concordance, distance, etc. Let $R = [r_{ij}]$ be the matrix of relational values on $\mathbf{O} \times \mathbf{O}$, r_{ij} being the relation between o_i and o_j. The most common relational data case is the dissimilarity data, say $D = [D_{ij}]$, where D_{ij} is the pairwise dissimilarity (which is usually a distance) $d(o_i, o_j)$ between o_i and o_j, for $1 \leq i, j \leq n$. D can also be a matrix of similarities based on a variety of measures [1]–[3]. Finally, D may be a relation specified by a person observing a process involving pairs of objects. Rectangular relational data are most commonly of this last type. For example, the NetFlix database aggregates values of ratings from viewers of movies; therefore, r_{ij} might correspond to the rating given to movie (i) by reviewer j.

When each object in $o_i \in \mathbf{O}$ (and, when represented by feature vector, $\mathbf{x}_i \in \mathbf{X}$) has a *physical label*, $\mathbf{O}$ is a set of *labeled data*; otherwise, $\mathbf{O}$ is a set of unlabeled data. For example, Anderson's Iris data, which were collected by Anderson [1], and subsequently, made famous by Fisher [2] comprises $n = 150$ feature vectors in four dimensions. Each vector in Iris has one of the three (i.e., crisp) physical labels corresponding to the Iris subspecies to which it belongs, i.e., sestosa, versicolor, or virginica.

In general, there are four class label types: *crisp, fuzzy, probabilistic*, and *possibilistic*. Let integer c be the number of classes, with $1 < c < n$, and define three sets of *label vectors* in $\Re^c$ as follows:

$$N_{pc} = \{\mathbf{p} \in \Re^c : p_i \in [0,1] \quad \forall\, i, \quad p_i > 0 \quad \exists\, i\} \tag{1a}$$

$$N_{fc} = \left\{\mathbf{p} \in N_{pc} : \sum_{i=1}^{c} p_i = 1\right\} \tag{1b}$$

$$N_{hc} = \{\mathbf{p} \in N_{fc} : \; p_i \in \{0,1\} \quad \forall i\}. \tag{1c}$$

Here, N_{hc} is the canonical (i.e., unit vector) basis of $\Re^c$. The ith vertex of N_{hc}, i.e.,

$$\mathbf{e}_i = \left(0, 0, \ldots, \underbrace{1}_{i\text{th place}}, 0, \ldots, 0\right)^T$$

where the 1 occupies the ith place, is the *crisp* label for class i, $1 \leq i \leq c$. The set N_{fc} is a piece of a hyperplane, and is the convex hull of N_{hc}. The vector $\mathbf{p} = (0.1, 0.6, 0.3)^{\mathrm{T}}$ is a label

Manuscript received October 20, 2009; revised February 26, 2010; accepted May 6, 2010. Date of publication June 7, 2010; date of current version September 29, 2010. The work of D. T. Anderson was supported by the National Library of Medicine under Grant T15 LM07089. The work of J. M. Keller was supported in part by the National Science Foundation under Grant IIS-0428420.

D. T. Anderson and J. M. Keller are with the Electrical and Computer Engineering Department, University of Missouri, Columbia, MO 65211 USA (e-mail: dtaxtd@mail.missouri.edu; kellerj@missouri.edu).

J. C. Bezdek, retired, was with the Electrical and Computer Engineering Department, University of Missouri, Columbia, MO 65211 USA (e-mail: jcbezdek@gmail.com).

M. Popescu is with the Informatics Institute, University of Missouri, Columbia, MO 65211 USA (e-mail: popescum@missouri.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2010.2052258

vector in N_{f3}; its entries lie between 0 and 1, and sum equals 1. There are two interpretations for the elements of N_{fc}. If $\mathbf{p}$ comes from a method, such as maximum-likelihood estimation in mixture decomposition, then $\mathbf{p}$ is a (usually posterior) *probabilistic* label, and p_i is interpreted as the probability that, given $\mathbf{x}$, it is in, or came from, class or component i of the mixture [3]. On the other hand, if $\mathbf{p}$ is a label vector for an object (o) generated by, say, the fuzzy-c-means clustering model [4], $\mathbf{p}$ is a *fuzzy* label for (o), and p_i is interpreted as the membership of (o) in class i. An important point for this paper is that N_{fc} has the same mathematical structure for probabilistic and fuzzy labels. Finally, $N_{pc} = [0,1]^c - \{\mathbf{0}\}$ is the unit (hyper)cube in $\Re^c$, *excluding the origin*. Vectors such as $\mathbf{z} = (0.7, 0.2, 0.7)^{\mathrm{T}}$ in N_{p3} are called *possibilistic* label vectors, and in this case, z_i is interpreted as the possibility that $\mathbf{x}$ is in, or came from, class i. Labels in N_{pc} are produced, e.g., by possibilistic clustering algorithms [5]. Evidently $N_{hc} \subset N_{fc} \subset N_{pc}$.

Clustering (i.e., unsupervised learning) in unlabeled data is the assignment of one of the four types of *labels* to the objects in $\mathbf{O}$. There are four types of clustering algorithms corresponding to the four types of labels. If the labels are hard (crisp), we hope they identify c natural subgroups in $\mathbf{O}$. The label vectors in (1) are used as the columns of three types of *c-partitions* of $\mathbf{O}$, which are sets of (cn) values $\{u_{ik}\}$ that can be conveniently arrayed as $(c \times n)$ matrices, say $U = [u_{ik}]$. Let $\mathbf{U}_k$ denote the kth column of U (which is a label vector in $\Re^c$). The three sets are given by

$$M_{pcn} = \left\{ U \in \Re^{cn} : U_k \in N_{pc} \quad \forall\, k, \quad 0 < \sum_{k=1}^{n} u_{ik} \quad \forall i \right\} \tag{2a}$$

$$M_{fcn} = \{ U \in M_{pcn} : U_k \in N_{fc} \quad \forall k \} \tag{2b}$$

$$M_{hcn} = \{ U \in M_{fcn} : U_k \in N_{hc} \quad \forall k \}. \tag{2c}$$

Equations (2a)–(2c) define, respectively, the sets of possibilistic, fuzzy, *or* probabilistic, and crisp *c-partitions* of $\mathbf{O}$. The reason that these matrices are called *partitions* follows from the interpretation of their entries. If U is crisp or fuzzy, u_{ik} is the *membership* of o_k in the ith partitioning crisp or fuzzy subset (cluster) of X. If U in M_{fcn} is probabilistic, u_{ik} is usually the (posterior) probability $p(i|o_k)$ that, given o_k, it came from class i, and if U in M_{pcn} is possibilistic, it has entries between 0 and 1 that do not necessarily sum to 1 over every column. In this case, u_{ik} is interpreted as the possibility that o_k belongs to class i. Observe that $M_{hcn} \subset M_{fcn} \subset M_{pcn}$. It is convenient to have a single name for the set $M_{pcn} - M_{hcn}$. This set excludes the crisp c-partitions and contains the fuzzy, probabilistic, and possibilistic c-partitions of $\mathbf{O}$. We call $M_{pcn} - M_{hcn}$ as the *soft c-partitions* of $\mathbf{O}$.

An alternative characterization of any crisp U in M_{hcn} is in terms of the c crisp subsets that are defined by the rows of U. Specifically, we may write $\mathbf{O} = \bigcup_{j=1}^{c} O_j$, where $O_i \cap O_j = \emptyset$, $i \neq j$. The ith row of U contains a 1 at each column k, where o_k is in class i, and $\sum_{k=1}^{n} u_{ik} = n_i = |O_i|$. $c = 1$ is represented uniquely by the crisp 1-partition, i.e.,

$$\mathbf{1}_n = \underbrace{[\,1\ 1\ \ldots\ 1\,]}_{n \text{ times}}$$

which asserts that all n objects belong to a single cluster. At the other extreme, $c = n$ is represented uniquely by $U = I_n$, which is the $n \times n$ identity matrix, up to a permutation of columns. In this case, each object is in its own singleton cluster. Choosing $c = 1$ or $c = n$ rejects the hypothesis that X contains clusters.

When the data are unlabeled, the following three questions about $\mathbf{O}$ collectively define cluster analysis.

Q1: Does $\mathbf{O}$ *have* cluster substructure at any value of c, $1 < c < n$?
Q2: If $\mathbf{O}$ has substructure, how can we *find* the clusters?
Q3: Once clusters are found, how can we *validate* them?

Q1 is called (i.e., preclustering) *assessment of clustering tendency*; we do not pursue this problem here; for formal and informal treatments, see [6] or [7].

Q2 is called *cluster analysis*. There are *many* models and algorithms for clustering based on crisp, fuzzy, probabilistic, and possibilistic methods [3]–[14]. Our examples will simply use several well-known clustering algorithms to generate candidate partitions for validation studies.

Q3 is called (i.e., postclustering) *cluster validity*; once U is found, do we believe it is the best explanation of substructure in $\mathbf{O}$? Is this U useful? Is there a better one we did not find?, etc. Just as tendency assessment depends on how clusters are defined, validation depends on what we mean by a good partition.

Aside: We have specified Q1–Q3 as if they were straightforward questions; however, they are really pretty vague, because we cannot say unambiguously what is meant by cluster structure. Humans intuitively understand that a cluster is a set of *similar* objects that are somehow "close." The measure of similarity itself may define "closeness"; however, as Rand [15] pointed out: "every definition of 'closer' is natural for some situation." Therefore, any description of clusters *in data* necessitates specification of a (i.e., mathematical) model that encapsulates the ideas of similarity and closeness. *Many* different mathematical properties have been used to define clusters, which invariably result in rather conflicting ideas about what our computers think datasets contain.

Clustering algorithms map $X \subset \Re^p$ *or* $R \subset \Re^{nn} \mapsto M_{pcn}$. Let $\mathrm{CP} = \{U_i : 1 \leq i \leq N\}$ denote N different *candidate partitions* of a fixed dataset $\mathbf{O}$ that may arise as a result of clustering (X or R) with one algorithm at various values of its parameters or, more generally, with different algorithms, each with its own parameters. In the sequel, we concentrate on validation of found partitions. Q2 asks the following: Which $U \in \mathrm{CP}$ best explains and represents the (unknown) structure in $\mathbf{O}$?

If clustering is guided by an objective function to the partitions in CP, at first glance, it seems like values of the objective function should suffice to choose the best one. However, it is well known that even the global extremum of many objective

functions (such as J_1 for hard c-means) can lead to very unrealistic partitions of $\mathbf{O}$; for an example of exactly this behavior, see [4, p. 97]. Moreover, some of the intuitively desirable properties that we may want a partition to have cannot be captured by any functional that is easily optimized. These are arguably the two most-compelling reasons to add the validation step to the clustering process.

The handful of partitions that you can feasibly generate from an unlabeled dataset has as a *least-common denominator*, i.e., the parameter c. Moreover, *c is the most-important* parameter of the partitions in CP, in the sense that the number of clusters sought in the data specifies the solution set being searched: If you are in the wrong set, it will be impossible to find a satisfactory explanation of the data (even though every clustering algorithm will happily supply one!). Thus, any effective validation strategy must find the *best value* for c. There is little guidance in the literature about $c_{\max}$. A rule of thumb that many investigators use is $c_{\max} \leq \sqrt{n}$.

There are two general approaches to choose a best U in CP. First, we may use *validity indices*, i.e., $v(U)$, which are computed at each U in CP. The set of values $\{v(U) : U \in \mathrm{CP}\}$ are subsequently used to identify the best U in CP, say U^*, in one of the following two ways: 1) U^* optimizes v over CP (i.e., min or max); or 2) U^* is the antecedent or successor partition that essentially optimizes the derivative of v. Many call this second method the "big-jump" approach, and is usually confined to sequences of partitions that are hierarchically nested by algorithms such as the single-linkage method [4]. Almost all optimization and big-jump indices are *internal criteria*, i.e., they use only internal information that is generated by the clustering process itself to select U^*.

The second group of methods use *comparison indices*, i.e., $s(U, V)$, that compare pairs of partitions. If U and V are both in CP (and hence, both obtained by algorithmic means), $s(U, V)$ is an *internal criterion*, and otherwise, we call it an external criterion. There are various ways to use such indices. Rand [15] lists the following five applications.

A1 [s(U,V)]: Here, V is a *reference partition* that purports to represent the "true cluster structure" in $\mathbf{O}$. $s(U, V)$ measures the extent to which U's in CP recover or retrieve the true clusters in $\mathbf{O}$, and hence, the sizes of U and V are equal. Retrieval rates are called *resubstitution error rates* when U and V are crisp, $r = c$, and s is a measure that counts matches between the columns in U and V.

A2 [$s(U, U')$]: Here, U is a partition of $\mathbf{O}$ found with clustering algorithm A. Perturb each $\mathbf{x}_k$ or r_{ij} with zero mean, unit-variance noise, and obtain U', which is a clustering of the perturbed data. $s(U, U')$ measures the sensitivity of A to perturbations of the data.

A3 [$s(U, U')$]: Here, U is a partition of $\mathbf{O}$ found with clustering algorithm A. Delete some $\mathbf{x}_k$'s or r_{ij}'s from the data and obtain U', which is a clustering of the reduced data. $s(U, U')$ measures the sensitivity of A to missing data.

A4 [$s(U, U')$]: Here, U and U' are partitions of $\mathbf{O}$ found with clustering algorithms A and A'. $s(U, U')$ measures the agreement between the two algorithms about clusters in $\mathbf{O}$.

A5 [$s(U, U')$]: Here, U and U' are partitions of $\mathbf{O}$ found by successive iterations of clustering algorithms. $s(U, U')$ measures the change in similarity due to the next iteration, and hence, can be used as a stopping criterion for iterative methods.

To this list, Hubert and Arabie [16] add the following application.

A6 [$s(U, V)$]: This is the prediction of V given U, or U given V (regression).

We *never* have an external reference partition in a real-clustering situation, which, by definition, involves unlabeled data. Therefore, why have case A1? Well, the only way you can evaluate *any* clustering algorithm before using it in a real-world situation is to see how well it recovers "true but unknown" reference partitions. If nothing else, good recovery rates on data with "known" cluster structure at least provide some psychological reassurance that the clustering algorithm can recover "good clusters" (sometimes!). As an additional caveat, "true" reference partitions fuel our expectation that the best algorithmic partitions will recover the same number of clusters. However, labels that are physically correct do not necessarily reflect the geometric situation in the numerical data, because the numerical features chosen may not provide the separation that a model needs to recover the physical classes. Thus, the Iris data have three subsets of vectors with different physical labels but only two clusters that are clearly (mathematically) separable in four spaces using most standard clustering algorithms.

Aside: There are other ways besides (i.e., internal and external) to classify validation methods. For example, $v(U)$'s are divided into (i.e., direct and indirect) validity indices in [5], depending on whether U is crisp or soft, respectively. This is a minor confusion; however, we must be aware of this "conflicting-jargon" problem.

There are far too many indices to be surveyed by us, or by any other authors. Hubert and Arabie [16] stated that a comprehensive review of $s(U, V)$ comparison methods alone was impossible in any journal paper since it would require "the length of a monograph"—in 1985! Beyond the discussions of many schemes involving $v(U)$ and/or $s(U, V)$ that can be found in texts [4]–[13], there are many number of surveys that limit themselves to one of the validation strategies; for references beyond those available via Google (which can supply you with an effectively infinite supply of links toward papers on this topic), see [17]–[21].

Here is the plan for the rest of this paper. Section II develops the background, which we need to describe many of the measures of the $s(U, V)$ type that have been used to compare pairs of crisp partitions. Section III discusses previous approaches to generalize several of the well-known measures to fuzzy partitions, and then introduces our approach to apply these measures to any pair of partitions in $M_{pcn} \times M_{pcn}$. Our method produces well-defined generalizations of many measures for all four partition types. Section IV contains some numerical examples that illustrate various facets of the new measures. Section V

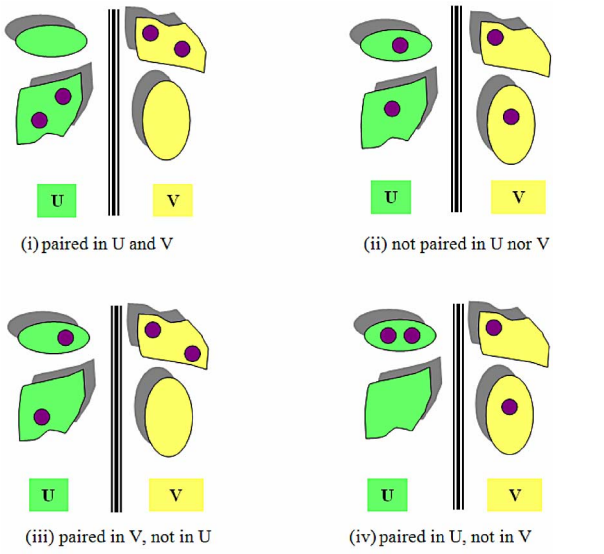

Fig. 1. Four types for a pair of objects from $\mathbf{O} \times \mathbf{O}$ relative to clusters in U and V. (i) Paired in U and V. (ii) Neither paired in U nor in V. (iii) Paired in V but not in U. (iv) Paired in U but not in V.

TABLE I
CONTINGENCY TABLE AND FORMULAS USED TO COMPARE CRISP PARTITIONS U AND V

		Partition V V_j = row j of V	
	Class	V_1 V_2 ... V_c	Sums
Partition U U_i = row i of U	U_1 U_2 ⋮ U_r	$N = \begin{bmatrix} n_{11} & n_{12} & \cdots & n_{1c} \\ n_{21} & n_{22} & \cdots & n_{2c} \\ \vdots & \vdots & & \vdots \\ n_{r1} & n_{r2} & \cdots & n_{rc} \end{bmatrix} = UV^T$	$n_{1\bullet}$ $n_{2\bullet}$ ⋮ $n_{r\bullet}$
	Sums	$n_{\bullet 1}$ $n_{\bullet 2}$ ⋯ $n_{\bullet c}$	$n_{\bullet\bullet} = n$

contains our conclusions and a discussion about conjectures and extensions of the new measures.

II. COMPARISON INDICES AND THE CONTINGENCY TABLE FOR (U, V)

Let $U \in M_{hrn}$ and $V \in M_{hcn}$ be two crisp partitions of n objects. Note that U and V need not possess the same number of clusters, i.e., $r \neq c$. The classical approach to compare U and V begins by considering the four possible combinations for pairs of objects from the set $\mathbf{O}$ in clusters of U and V, as shown in Fig. 1. We have numbered the four types, i.e., (i)–(iv), following Hubert and Arabie [16, p. 194]. The comparison of U with V with a similarity measure s begins with the $r \times c$ contingency table, as shown in Table I, that contains counts of the number of occurrences of each of the four types over the $n(n-1)/2$ distinct, unordered pairs in $\mathbf{O} \times \mathbf{O}$.

Entry n_{ij} from $N = UV^T$ is the number of objects common to classes U_i and V_j. For example, let us consider the following:

$$U = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \quad V = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

$$N = UV^T = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{matrix} 3 \\ 2 \end{matrix}$$
$$\begin{matrix} 2 & 1 & 2 \end{matrix} \quad 5.$$

Clusters U_1 and V_1 have two objects in common, i.e., $\{o_1, o_2\}$, U_1 and V_2 have 1 object in common, i.e., $\{o_3\}$, and U_1 and V_3 have 0 objects in common. Additionally, there are two paired in U and V, i.e., type (i), i.e., $\{\{o_1, o_2\}, \{o_4, o_5\}\}$, six in type (ii), i.e., $\{\{o_1, o_4\}, \{o_1, o_5\}, \{o_2, o_4\}, \{o_2, o_5\}, \{o_3, o_4\}, \{o_3, o_5\}\}$, zero paired in type (iii), and two paired in type (iv), i.e., $\{\{o_1, o_3\}, \{o_2, o_3\}\}$. Equations (3a)–(3d), which are functions of the entries in Table I, are the building blocks of many similarity measures for $s(U, V)$. These four equations simply count the number of occurrences among the $n(n-1)/2$ pairs of each of the four types of unordered pairs shown in Fig. 1. Thus, a in (3a) is the number of pairs that are paired in some cluster of U and some cluster of V; b is the number of pairs neither paired in U nor in V, etc. It is easy to verify that the calculation of the matrix N in Table I as $N = UV^T$ is correct; however, to our knowledge, it has never been exhibited quite this way. This simple observation will enable us to easily define soft generalizations of many of the comparison indices that are to be discussed next. Many well-known crisp comparison indices are functions of the values in Table I. Some of these indices have "canonical stories" attached to them that depend on interpretations of the next four functions. Here

$$a = \frac{1}{2} \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij}(n_{ij} - 1) \tag{3a}$$

is the number paired in U and V

$$d = \frac{1}{2} \left(n^2 + \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij}^2 - \left(\sum_{i=1}^{r} n_{i\bullet}^2 + \sum_{j=1}^{c} n_{\bullet j}^2 \right) \right) \tag{3b}$$

is the number paired in neither U nor V

$$b = \frac{1}{2} \left(\sum_{j=1}^{c} n_{\bullet j}^2 - \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij}^2 \right) \tag{3c}$$

is the number paired in V, but not in U, and

$$c = \frac{1}{2} \left(\sum_{i=1}^{r} n_{i\bullet}^2 - \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij}^2 \right) \tag{3d}$$

is the number paired in U, but not in V.

The sums $(a + d)$ and $(b + c)$ are usually interpreted, respectively, as (the total number of) *agreements* and *disagreements* between U and V. Table II contains a [nonexhaustive] list of coefficients that have been proposed for $s(U, V)$ based on functions of a, b, c, and d.

TABLE II
COMPARISON MEASURES AND SOME GENERALIZATIONS BETWEEN PARTITIONS U AND V

			Fuzzy Generalizations			
Name/Reference	**Crisp Measure**	**Eqn**	**[35]**	**[36]**	**[37]**	**[Us]**
Rand [15]	$s_r(U,V) = (a+d)/(a+b+c+d)$	(4a)	✔	✔	✔	✔
Johnson [22] Mirkin & Chernyi [23] Arabie & Boorman [24]	$(b+c)/(a+b+c+d))$	(4b)				✔
Hubert [25]	$((a+d)-(b+c))/(a+b+c+d)$	(4c)				✔
Wallace [26]	$a/(a+b)$ and $a/(a+c)$	(4d)				✔
Fowlkes and Mallow [27]	$a/\sqrt{(a+b)(a+c)}$	(4e)	✔			✔
Jacard [28]	$a/(a+b+c)$	(4f)	✔		✔	✔
Adjusted Rand Hubert and Arabie [16]	$\dfrac{\left(a-\dfrac{(a+c)(a+b)}{a+b+c+d}\right)}{\left(\dfrac{(a+c)+(a+b)}{2}-\dfrac{(a+c)(a+b)}{a+b+c+d}\right)}$	(4g)				✔
Adjusted Rand Campello [35] Assumed: r = c	$\dfrac{\left(a-\dfrac{(a+c)(a+b)}{d}\right)}{\left(\dfrac{(a+c)+(a+b)}{2}-\dfrac{(a+c)(a+b)}{d}\right)}$	(4h)	✔			✔
Adjusted Rand Brouwer [37]	$\dfrac{2(ad-bc)}{c^2+b^2+2ad+(a+d)(b+c)}$	(4i)			✔	✔
Minkowski [30]	$s_{\text{Mink}}(U,V) = \sqrt{(b+c)/(b+a)}$	(4j)	✔			✔
Hubert's Gamma [6]	$\dfrac{\left(\binom{n}{2}a-(a+b)(a+c)\right)}{\sqrt{(a+b)(a+c)\left[\binom{n}{2}-(a+b)\right]\left[\binom{n}{2}-(a+c)\right]}}$	(4k)	✔			✔
Yule [14]	$((ad)-(bc))/(ab+cd)$	(4l)				✔
Chi-Squared [16]	$\chi^2 = n\left(\sum_{j=1}^{c}\sum_{i=1}^{r}\dfrac{n_{ij}^2}{n_{i\bullet}n_{\bullet j}}-1\right)$	(4m)				✔
Goodman-Kruskal [38]	$\tau_b = \left[n\left(\sum_{j=1}^{c}\sum_{i=1}^{r}\dfrac{n_{ij}^2}{n_{i\bullet}}\right)-\sum_{j=1}^{c}n_{\bullet j}^2\right]\Big/\left[n^2-\sum_{j=1}^{c}n_{\bullet j}^2\right]$	(4n)				✔

There are other comparison indices based on the entries of the contingency matrix $N = UV^T$ in Table II that are not easily expressible in terms of the quantities a, b, c, and d in (3a)–(3d). The last two rows in Table II exhibit functions that fall into this category; for several other measures of this type, see [16]. Table II has the look of an extensive collection of comparison indices; however, it barely scratches the surface of this monolithic topic. You can find many more indices like these in the literature—[14] is a good starting point for such an expedition.

A. Resubstitution Error Rate

Let us consider the use of $s(U, V)$ to compare soft partitions U to a crisp reference partition V in the special case when $r = c$, i.e., when U and V have the same number of clusters. When U is crisp, we are sometimes able to interpret the value of $s(U, V)$ in terms of a canonical story. For example, we know that the Rand index is the ratio of pair agreements to the number of pairs. However, when U is soft, we are unable to provide a canonical situation that describes what the index (physically) represents, even though we may find utility in computing the extended value.

The traditional approach to this dilemma is to first harden any soft partition U. Then, the hardened version of U, say $H(U)$, can be compared with V, and the number of label matches counted. This amounts to a comparison method that is similar to the retrieval problem A1, which circumvents the difficulty posed by U being soft in a different way. In order to do this, soft labels in N_{pc} must be transformed into crisp labels. Usually, noncrisp labels ($\mathbf{p}$) are converted to crisp ones ($\mathbf{e}_i$) using the conversion function $H : N_{pc} \mapsto N_{hc}$, which is defined as

$$H(\mathbf{p}) = \mathbf{e}_i \Leftrightarrow \|\mathbf{p}-\mathbf{e}_i\|_2 \le \|\mathbf{p}-\mathbf{e}_j\|_2 \Leftrightarrow p_i \ge p_j, \quad j \ne i \tag{5}$$

where ties are broken arbitrarily.

Geometrically, H finds the crisp label vector $\mathbf{e}_i$ in N_{hc} closest to $\mathbf{p}$ in the Euclidean norm $\|\mathbf{p}-\mathbf{e}_i\|_2$ and uses it versus $\mathbf{p}$ to label the associated object unequivocally as belonging to cluster i.

Now, let U be any soft partition of $\mathbf{O}$. Let us apply H to each column U^k of U, which results in the crisp label vector $[\mathbf{H}(U)]^k$. The matrix$[H[U]]$ obtained by applying this transformation to each of the n columns of U is regarded as the best approximation to U by a crisp partition of $\mathbf{O}$ in the sense of maximum: *memberships* when U is fuzzy, *posterior probabilities* when U is probabilistic, or *typicalities* when U is possibilistic. We note that hardening U does not guarantee that $H(U)$ is a crisp c-partition of $\mathbf{O}$; however, if hardening happens to produce a zero row in $H(U)$, this does not alter the number of matching columns between V and $H(U)$. We compute the number of label matches in the columns of $H(U)$ to the reference partition V as

$$e(H(U),V)=\frac{\sum_{k=1}^{n}\left\|[\mathbf{H}(U)]^k-V^k\right\|_1}{2n}. \tag{6}$$

The 1-norm is used in (6), and $e(H(U),V)$ is called the *resubstitution error rate* for the soft clusters in U. This measure takes the value 1 when no columns match, and is 0 when all n columns match. To make (6) comparable with the Rand index $s_r(H(U),V)$, we compute

$$\begin{aligned}s_e(H(U),V)&=1-e(H(U),V)\\&=1-\frac{\sum_{k=1}^{n}\left\|[\mathbf{H}(U)]^k-V^k\right\|_1}{2n}.\end{aligned} \tag{7}$$

Let us consider the following example:

$$U=\begin{bmatrix}0.1 & 0.7 & 0.4 & 1 & 0.9\\ 0.9 & 0.3 & 0.6 & 0 & 0.1\end{bmatrix}\underbrace{\rightarrow\rightarrow\rightarrow}_{\text{hardening}}H(U)$$

$$=\begin{bmatrix}0 & 1 & 0 & 1 & 1\\ 1 & 0 & 1 & 0 & 0\end{bmatrix},\quad V=\begin{bmatrix}1 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 1 & 1\end{bmatrix}$$

$s_e(H(U),V)=1-\sum_{k=1}^{n}\|[\mathbf{H}(U)]^k-V^k\|/2n=0.4$, and here, $s_r(H(U),V)=0.4$ as well. While this suggests that the Rand index coincides with $s_e(H(U),V)$ in this special case, a counterexample is easy to find, which is given by

$$U=\begin{bmatrix}1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0\end{bmatrix}$$

$$V=\begin{bmatrix}1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1\end{bmatrix}$$

for which $s_e(H(U),V)=1-\sum_{k=1}^{n}\|[\mathbf{H}(U)]^k-V^k\|/2n=0.75$, while $s_r(U,V)=0.5909$.

The general situation in this special case (recall that $r=c$ here) is shown in Fig. 2. The lower part of this figure shows that we can use any comparison index in Table II to compute $s(H(U),V)$ once U is hardened by H, and we can also compute $s_e(H(U),V)$. On the other hand, once $s(U,V)$ is generalized to cover the case of soft partitions, we can calculate any comparison index in Table II directly on the pair (U,V), as shown in the upper part of Fig. 2. In summary, when, and only when, $r=c$—and *after* we have solved the correspondence problem for $H(U)$ below—we will have two approaches to generalize the comparison of a soft U to a crisp reference partition V.

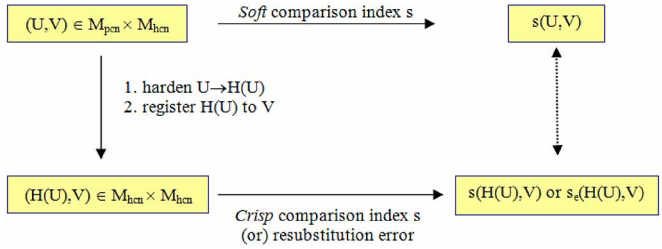

Fig. 2. Comparison indices versus resubstitution error when $r=c$ with reference partition V.

$V=\begin{bmatrix}1 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 1 & 1\end{bmatrix}$		$V=\begin{bmatrix}1 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 1 & 1\end{bmatrix}$
$U=\begin{bmatrix}0 & 0 & 1 & 1 & 1\\ 1 & 1 & 0 & 0 & 0\end{bmatrix}$	$\underbrace{\rightarrow\rightarrow\rightarrow}_{\text{relabeling}}$	$U'=\begin{bmatrix}1 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 1 & 1\end{bmatrix}$
$s_r(U,V)=1$ $s_e(U,V)=0$		$s_r(U',V)=1$ $s_e(U',V)=1$

Fig. 3. Correspondence problem in clustering.

B. Correspondence Problem

Before using $s_e(H(U),V)$, we must account for the correspondence (or registration) problem. The problem is easiest to understand when $s_e(H(U),V)$ is used to compare a reference partition V to candidate partitions $H(U)$ and $r=c$—i.e., when $s_e(H(U),V)$ assesses the retrieval success of $H(U)$ in recovering the known labels in V. However, the problem exists, even when $r\neq c$ so that U and V represent different numbers of clusters in $\mathbf{O}$. Fig. 3 depicts the problem for two partitions U and V of a set of five objects, say $\mathbf{O}=\{o_1,\ o_2\ ,o_3,o_4,\ o_5\}$.

There are a variety of approaches for relabeling $H(U)$ so that its clusters are "aligned" with the reference clusters in V. When the data are vectorial, there are a number of relabeling algorithms based on the available cluster prototypes [5]. In the case of relational data, one may use an accuracy metric that relies on permuting the cluster labels to register $H(U)$ to V [39]–[41]. Toward this end, let $\mathbf{I}_{\text{gt}}$ be the vector of ground truth labels in V for n objects. Let us assume without loss that $\mathbf{I}_{\text{gt}}$ is arranged lexicographically—i.e., the objects are arranged so that $\mathbf{I}_{\text{gt}}$ has the form

$$\mathbf{I}_{\text{gt}}=(\underbrace{11\ldots1}_{n_1},\underbrace{22\ldots2}_{n_2},\ldots,\underbrace{cc\ldots c}_{n_c})$$

where n_i is the number of objects bearing label i, $1\leq i\leq c$, and $\sum_{i=1}^{c}n_i=n$. Let $\mathbf{I}_A$ denote the vector of labels in $H(U)$ obtained by clustering the n objects with clustering algorithm A, $\mathbf{I}_A\in\{1,2,\ldots,c\}^n$. It is possible that A finds the ground-truth clusters in V exactly; however, $H(U)$ does not present the labels in ground-truth order—this is exactly the case for U and V on the left side of Fig. 3. There, we see that if we alter U to U' by switching the rows of U, then V and U' are in 100% agreement, and $s_e(U,V)=1$.

Let π be the permutation functions on n slots, with $\pi \in \Pi$, $\pi : \{1, 2, \ldots, n\} \mapsto \{1, 2, \ldots, n\}$, and let us define

$$AC(\mathbf{I}_{\rm gt}, \mathbf{I}_A) = \underbrace{\max}_{\pi \in \Pi} \left\{ \frac{\sum_{i=1}^{n} \delta([\mathbf{I}_{\rm gt}]_i, [\pi(\mathbf{I}_A)]_i)}{n} \right\} \quad (8)$$

where

$$\delta(x, y) = \left\{ \begin{array}{ll} 1, & x = y \\ 0, & x \neq y \end{array} \right\}.$$

The global solution of optimization problem (8) identifies the permutation of $\mathbf{I}_A$ that is the best match to $\mathbf{I}_{\rm gt}$. The solution to (8) grows with factorial complexity but can be solved by dynamic programming with, for example, the Kuhn–Munkres algorithm [42].

A last observation about this topic—but an important one—is that all of the indices in Table II depend on double sums, row sums, or column sums of the entries of the matrix $N = UV^T$. Consequently, crisp comparison indices, such as Rand's index, are essentially independent of the correspondence problem that plagues evaluation of retrieval success for soft clustering algorithms by the "harden and count" method represented in (6) or (7). This is an important advantage for the crisp-comparison-index method, which remains true, even when $r \neq c$, and the resubstitution error rate cannot even be computed!

III. Generalizing $s(U, V)$ When U and/or V are Soft Partitions

This section contains our proposal to generalize *all* of the indices in Table II, as well as any others that depend only on the entries of $N = UV^T$, to the cases where U and/or V are *any* partitions in $M_{pcn} \times M_{pcn}$. We begin with *Rand's index* (see (4a); see Table II). This index apparently first appeared in 1958 [29], where it was called a simple matching coefficient. Rand reintroduced this function in 1971 [15], and the literature has consistently referred to it as "Rand's index" since then. Hubert and Arabie [16] give a nonexhaustive set of 11 other citations that track the history of the Rand index. Because it has a simple, natural interpretation, the Rand index continues to be one of the most-popular comparison indices, which recently appeared in many papers concerned with clustering microarrays and gene-expression data [30]–[34].

The *adjusted Rand index* was first discussed by Morey and Agresti, who suggested that the Rand index be normalized by accounting for the possibility that U and V are random partitions of $\mathbf{O}$. Hubert and Arabie [16] use an adjustment for chance, which is based on the formula

$$s^*(U, V) = \frac{s(U, V) - E[s(U, V)]}{\max\{s(U, V)\} - E[s(U, V)]}. \quad (9)$$

The quantity $E[s(U, V)]$ in [16] is the expected value of the index under the hypergeometric assumption. Hubert and Arabie note that Morey and Agresti's adjustment is incorrect, and then derive the adjusted Rand-index formula (4g) in Table II. We think that (4g) is usually accepted as the correct adjustment of the Rand index; however, note that Hubert and Arabie listed at least three alternate forms of this index under different assumptions about the term $\max\{s(U, V)\}$ in (5).

The recent use of Rand's original index [30]–[34] has renewed interest to generalize it, along with some of the other comparison indices shown in Table II, to various noncrisp cases. Specifically, we mention the papers of Campello [35], Hullermeier and Rifqi [36], and Brouwer [37], all of which generalize the Rand index to the case of U and/or V being fuzzy partitions of the n objects. Two of these authors also present generalizations of the adjusted Rand index; however, as you can see in Table II, neither Campello's [see (4h)], nor Brouwer's [see (4i)] adjusted Rand index agrees with the adjustment derived by Hubert and Arabie [see (4g)]. In fact, there seem to be many other "adjusted Rand indices" in the literature; for two other adjusted Rand indices, see [32]. We will use (4g) as our basis for a soft generalization of the adjusted Rand index.

Campello [35] presents a method to fuzzify the original Rand index in (4a), as well as counterparts of the other five related indices given in the fourth column of Table II. Campello presents a special version of the adjusted Rand index, as shown in (4h) of Table II. Campello asserts that his generalization of these six indices is valid for any soft degenerate (one or more empty clusters) or nondegenerate partition U. Campello consider neither the case where V might be fuzzy nor the applications A2–A6 for $s(U, V)$ listed above. His scheme is based on presenting (3a)–(3d) in an equivalent form using (cardinalities of) intersections of the crisp subsets of $\mathbf{O} \times \mathbf{O}$ corresponding to each of the four totals, and then replacing the crisp sets with fuzzy sets to arrive at a partial generalization, which we denote below as $s_{\rm FRC}(U, V)$. For later comparison, we record Campello's generalization of (3a)

$$a = \frac{1}{2} \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij}(n_{ij} - 1) \underbrace{\mapsto}_{\rm Campello}$$

$$a = \sum_{j=1}^{i-1} \sum_{i=2}^{n} \left(\left(\bigvee_{k=1}^{r} (U_{ki} \wedge U_{kj}) \right) \wedge \left(\bigvee_{k=1}^{c} (V_{ki} \wedge V_{kj}) \right) \right). \quad (10)$$

Rand [15] pointed out that $s_r(U, V) = (a + d)/(a + b + c + d)$ in (4a) satisfied the following properties: 1) $s_R(U, V) = 0 \Leftrightarrow U = \mathbf{1}_n$ and $V = I_n$ (or the reverse); 2) $s_R(U, V) = 1 \Leftrightarrow U = V$; and 3) $d_R(U, V) = 1 - s_R(U, V)$ is a metric on the set $Y = \bigcup_{k=1}^{n} M_{hkn}$. Campello's generalization of Rand's index [i.e., fuzzy Rand of Campello (FRC)], $s_{\rm FRC}(U, V)$, is valued in [0, 1]; however, condition 2) on $U = V$ is only necessary in the fuzzy case and not necessary and sufficient for the index to take the value 1. Curiously, Campello states that "the fuzzy-Rand index cannot be seen as a general measure for comparing two fuzzy partitions," even though his index is well defined for this case. Instead, he advocates using it only as a measure to compare a fuzzy partition against a hard partition (possibly with a different number of categories). Consequently, this is the only numerical case illustrated in [35], and this is done only with a "toy" example—no real data are clustered, nor are any computed partitions evaluated, in this paper.

Hullermeier and Rifqi [36] take issue with Campello's definition of the fuzzy-Rand index. They assert, and subsequently, demonstrate that the function $d_{\mathrm{FRC}}(U,V) = 1 - s_{\mathrm{FRC}}(U,V)$ is not a true distance measure on $M_{frn} \times M_{fcn}$. They do not formulate their index in terms of (3a)–(3d). Instead, their generalization is guided by the fact that Rand's index counts the number of paired agreements, i.e., $(a+d)$, divided by the total number of possible pairs, i.e., $(a+b+c+d)$, and this leads them directly to the fuzzy Rand of Hullermeier and Rifqi (FRHR) index given by

$$s_{\mathrm{FRHR}}(U,V) = 1 - \left[\sum_{j=i+1}^{n} \sum_{i=1}^{n-1} \left| \|\mathbf{V}^i - \mathbf{V}^j\| - \|\mathbf{U}^i - \mathbf{U}^j\| \right| \Big/ \binom{n}{2} \right]. \tag{11}$$

In (11) and the sequel, $\mathbf{A}_i$ and $\mathbf{A}^j$ denote the vectors corresponding to the ith row and jth column of any matrix $\mathbf{A}$, and $\langle \mathbf{A}_i, \mathbf{A}^j \rangle$ is the dot product of these two vectors.This index is a direct generalization of the original Rand index in (4a) to the case where U and V are both fuzzy partitions of $\mathbf{O}$. Hullermeier and Rifqi show that the function $d_{\mathrm{RHR}}(U,V) = 1 - s_{\mathrm{RHR}}(U,V)$ is a pseudometric on $M_{frn} \times M_{fcn}$, and a full metric on a very sparse subset of $M_{frn} \times M_{fcn}$, which is called the "normal" partitions of $\mathbf{O}$—these are partitions that have at least one 1 in each row of the partition. This paper contains no examples of the use of this index to compare fuzzy partitions; therefore, it is difficult to say much more about it until we report on its values in our numerical experiments.

Brouwer [37] discusses a different generalization to the three indices given in Table II: the Rand, namely, the Rand, a third variant of the adjusted Rand, and Jaccard's index. Brouwer's generalization of (3a) is given by

$$\begin{aligned} a &= \frac{1}{2} \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij}(n_{ij}-1) \underbrace{\quad \mapsto \quad}_{\text{Brouwer}} \\ a &= \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{\langle \mathbf{U}_i^T, \mathbf{U}^j \rangle}{\|\mathbf{U}_i^T\| \|\mathbf{U}^j\|} \frac{\langle \mathbf{V}_i^T, \mathbf{V}^j \rangle}{\|\mathbf{V}_i^T\| \|\mathbf{V}^j\|} \right)}{2} - \frac{n}{2} \\ &= \frac{\left(\sum_{i=1}^{n} \sum_{j=1}^{n} \left(\cos(u_i^j) \cos(v_i^j) \right) \right) - n}{2}. \end{aligned} \tag{12}$$

In (12), $\cos(u_i^j)$ is the angle between the vectors $\mathbf{U}_i^T$ and $\mathbf{U}^j$, and $\cos(v_i^j)$ is the angle between $\mathbf{V}_i^T$ and $\mathbf{V}^j$. We see from (12) that the maximum occurs when both cosines are 1. Then, $a = n(n-1)/2$, which will occur if and only if $d = 0$ in Rand's formula, from which we conclude that $U = V$.

Our method begins by observing that the contingency Table I for U and V is $N = UV^T$. If the only case of interest concerned softening these indices for use with fuzzy or probabilistic partitions, this matrix would serve as a basis for our approach. However, we want to include the situation wherein U and/or V are possibilistic. Let us consider a problem that naturally exhibits $\sum_{i=1}^{r} u_{ik} > 1$ and/or $\sum_{j=1}^{c} v_{jk} > 1$, or coincidental clusters, via an algorithm such as the possibilistic c-means [43]. In the possibilistic case, a measure, such as Rand's index, can yield both positive and negative values outside of [0,1]. To see how this is possible, let us consider the value of d in (3b), the number paired in neither U nor V, and $\sum_{i=1}^{r} n_{i\bullet} > n$, or $\sum_{j=1}^{c} n_{\bullet j} > n$, respectively. One or both of the terms $\sum_{i=1}^{r} n_{i\bullet}^2$ and $\sum_{j=1}^{c} n_{\bullet j}^2$ can make d in (3b) relatively large and negative. Depending on the values of b in (3c) and c in (3d), a measure, such as the (soft) Rand index, can result in $s_r(U,V) < 0$. While this value is still of use in terms of optimization (max or min) or identification of a "big jump," indices valued in [0,1] are intuitively easier to interpret, and facilitate comparison to other indices. Instead of calculating $N = UV^T$, we scale it with $\phi = n/\sum_{i=1}^{r} n_{i\bullet}$. Alternatively, we could scale N with $\varphi = n/\sum_{j=1}^{c} n_{\bullet j}$ because $\sum_{i=1}^{r} n_{i\bullet} = \sum_{j=1}^{c} n_{\bullet j}$, and hence, $\varphi = \phi$. Consequently, we define

$$N^* = \phi UV^T = \left[n \Big/ \sum_{i=1}^{r} n_{i\bullet} \right] UV^T. \tag{13}$$

An advantage of using this scaling factor is that in the cases of crisp, fuzzy, or probabilistic partitions, $\phi = 1$, and thus, $N^* = N = UV^T$. This shows that using the entries of (13) with *ANY* of the indices in Table II will reduce to the original indices when U and V are both crisp partitions of the n objects, and otherwise, they will constitute valid soft generalizations of all of those indices. Moreover, in the case of possibilistic partitions, the normalization in (13) also produces index values in the range [0,1]. Let us consider the following example, which illustrates this when V is a crisp reference partition for eight objects, and U is an *extreme* possibilistic partition of the eight objects:

$$U = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$$V = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Using the entries of $N = UV^T$ and (3a)–(3d), we find $a = 33$, $b = 75$, $c = 0$, and $d = (-)85$. Substituting these into (4a), the possibilistic Rand index is $s_r(U,V) = (-)2.2609$. Using instead the entries of $N^* = \varphi UV^T$, we find $a = 4.29$, $b = 14.81$, $c = 0$, $d = 8.88$, and a possibilistic Rand index, $s_r(U,V) = 0.4709$.

For comparison, our generalization of (3a), for $N = UV^T$, is

$$\begin{aligned} a &= \frac{1}{2} \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij}(n_{ij}-1) \underbrace{\quad \mapsto \quad}_{\text{Anderson } et\ al.} \\ a &= \frac{1}{2} \sum_{i=1}^{r} \sum_{j=1}^{c} \langle \mathbf{U}_i, (\mathbf{V}^T)^j \rangle (\langle \mathbf{U}_i, (\mathbf{V}^T)^j \rangle - 1) \end{aligned} \tag{14}$$

and using N^* in (13) instead of N, we have

$$a = \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij}(n_{ij}-1) \underbrace{\mapsto}_{\text{Anderson } et\,al.}$$

$$a = \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{c}\left(\frac{n\left\langle \mathbf{U}_i, (\mathbf{V}^T)^j\right\rangle}{\sum_{k=1}^{r}\sum_{p=1}^{c}\left\langle \mathbf{U}_k, (\mathbf{V}^T)^p\right\rangle}\right) \times\left(\left(\frac{n\left\langle \mathbf{U}_i, (\mathbf{V}^T)^j\right\rangle}{\sum_{k=1}^{r}\sum_{p=1}^{c}\left\langle \mathbf{U}_k, (\mathbf{V}^T)^p\right\rangle}\right)-1\right). \quad (15)$$

Comparison of (15) with the fuzzy Rand indices of Campello (10), Hullermeier and Rifqi (11), and Brouwer (12) shows that the four indices are different from each other. The big advantage of our method of forming the contingency table as $N^* = \phi UV^T = [n/\sum_{i=1}^{r} n_{i\bullet}]UV^T$ is that this formulation directly generalizes all 15 indices in Table II (and any other indices that are functions only of the entries in the contingency table) to every combination of (U, V). There are 16 possible pair types accordingly as each of U and V are crisp, fuzzy, probabilistic, or possibilistic; therefore, we have, for example, 16 Rand indices, 16 Jaccard indices, etc. Each formula there is recovered when U and V are crisp, i.e., these are true generalizations to every case—by definition.

Our generalization is a similarity measure and not a metric. The job of $s(U, V)$ is to compare pairs of partitions. The values calculated in this study are of use in terms of optimization (max or min) or identification of a "big jump." This is consistent with the work of Campello [35] and Hullermeier and Rifqi [36], where the latter is a pseudometric valid for only a subclass of fuzzy partitions. Again, the job of $s(U, V)$ is to compare pairs of partitions. One way to do this is to measure the distance between them, and any norm on matrices will do this. However, this only works when U and V are of the same size, and moreover, this is not necessarily the optimal way to compare clusters represented by those matrices.

Finally, we currently compute $n_{ij} = \langle \mathbf{U}_i, (\mathbf{V}^T)^j\rangle = \sum_{k=1}^{n} u_{ik}v_{jk}$ (not including the scaling factor). Using theproduct and sum for the inner and outer operations in matrix multiplication is but one instance of the more general form based on S and T norms. In general, $n_{ij} = S(T(u_{ik}, v_{jk})$, which then reduces to the form that we use here for the choices $T =$ product and $S =$ sum. Another common choice would be $T =$ min and $S =$ max. We leave the possibility that such a change would result in better results than the ones reported here for a future study.

IV. Numerical Examples on Cluster Validity

A. Example 1

The Rand index given by (4a) is used with the entries of N^* in (13), where partition V is the reference matrix and partitions U are obtained using HCM, FCM, and PCM.The FCM partitions are obtained using the *fcm* function from the MATLAB Fuzzy Logic Toolbox with $c = 4$, $m = 2$, maximum number of iterations MAXIT $= 100$, objective function error EPS $=$ 1e–5, and random-partition initialization.The HCM partitions are obtained using the MATLAB Statistics Toolbox function *k-means*, with random initialization and MAXIT $= 100$. The PCM partitions are generated using our MATLAB implementation of the algorithm from [43] with $c = 4$, $m = 3$, MAXIT $= 100$, EPS $=$ 0.001, and a fuzzy-c-mean-based initialization.

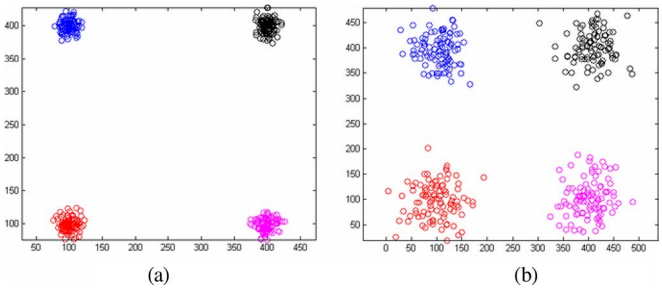

Fig. 4. Gaussians datasets used in Example 1. (a) Dataset 1. (b) Dataset 2.

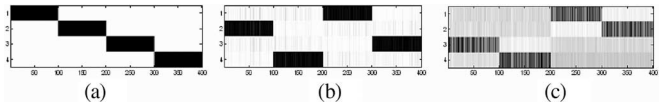

Fig. 5. Terminal partitions U acquired by (a) HCM, (b) FCM, and (c) PCM for $c = 4$ on dataset 2.

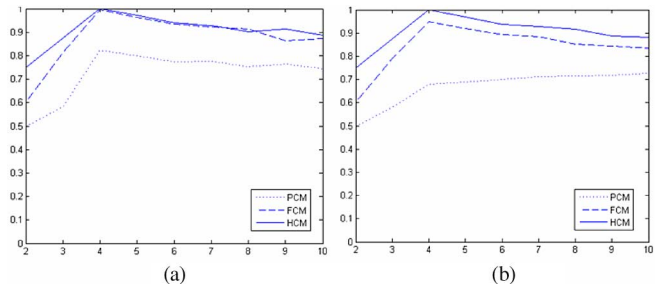

Fig. 6. Original Rand for HCM and our soft generalizations, fuzzy and possibilistic Rand for FCM and PCM, respectively, for (a) dataset 1 and (b) dataset 2.

We use two datasets: dataset 1 [see Fig. 4(a)] and dataset 2 [see Fig. 4(b)]. The two datasets are made up of four Gaussian clouds each with 100 samples per cloud. The means of the Gaussians are $[100\ \ 400]^T$, $[100\ \ 100]^T$, $[400\ \ 100]^T$, and $[400\ \ 400]^T$. All clouds in dataset 1 have a standard deviation of 10, while in dataset 2, they have a standard deviation of 30. The reference partition V for both datasets is the 4×400 matrix with four "diagonal" blocks of one hundred 1's in each row.

Fig. 5 displays the terminal partition matrices obtained for dataset 1 by the three clustering algorithms.

The partitions in Fig. 5(a) (i.e., HCM) and in Fig. 5(b) (i.e., FCM) exemplify the correspondence problem mentioned above. The points with index $[1\ \ 100]^T$ are in cluster 1 in U_{HCM}, while they are in cluster 2 in U_{FCM} and likewise for clusters 3 and 4. Therefore, before assessing the validity of a crisp version of FCM partition (5b) using (7) along the lower path in Fig. 2, we would need to swap clusters (1,2) and (3,4) using an algorithm such as (8) on $H(U)$.

Fig. 6 shows the Rand index obtained by varying the number of clusters in the HCM, FCM, and PCM runs from 2 to 10

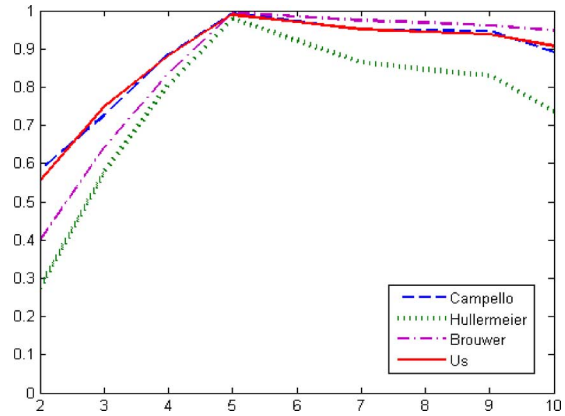

Fig. 7. Comparison of the four fuzzy-Rand indices on FCM partitions of dataset 3, which has $c = 5$ clusters.

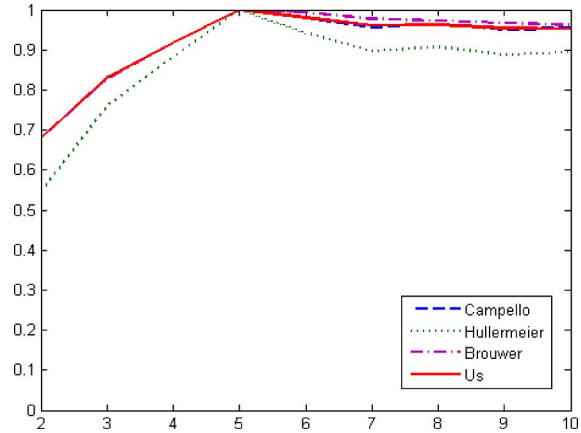

Fig. 8. Comparison of the four fuzzy-Rand indices on EM partitions of dataset 3, which has $c = 5$ clusters.

with each of the two datasets shown in Fig. 4. All three Rand indices (original Rand for HCM, and fuzzy and possibilistic Rand for FCM and PCM, respectively) achieve a maximum at $c = 4$ in Fig. 6(a), thereby indicating that among the nine partitions for each case, the best match for V is U_4. This implies that our generalizations of the Rand index behave as expected when the data form compact, well-separated clusters like the ones shown in Fig. 4(a). The graphs in Fig. 6(b) show that our Rand and fuzzy-Rand indices for HCM and FCM behave in a similarly desirable fashion, but the possibilistic Rand index fails to reach a maximum at $c = 4$. We believe that this is due to the characteristics of dataset 2, which favor HCM and FCM solutions. PCM is designed to account for inliers and outliers that do not seem to belong strongly to any cluster; however, looking back at Fig. 4(b), there are no points of these types in the data.

B. Example 2

In this example, we compare our fuzzy-Rand index with the indices of Campello [35], Hullermeier and Rifqi [36], and Brouwer [37]. We use FCM, same parameters as in Example 1, to generate fuzzy c-partitions of dataset 3, which consists of five Gaussians clouds in two dimensions with 100 samples each, means $[1\ \ 1]^T$, $[1\ \ 40]^T$, $[20\ \ 20]^T$, $[40\ \ 1]^T$, and $[40\ \ 40]^T$, and a standard deviation of 1.5. These clusters are fairly compact and well separated. The reference partition V for this example is the 5×500 matrix with five "diagonal" blocks of one hundred 1's in each row. Fig. 7 shows graphs of the four fuzzy-Rand indices for terminal FCM partitions of dataset 3 as c varies from 2 to 10; therefore, there are nine candidate partitions in CP. This graph shows two things: First, the four indices are indeed different; and second, they have very similar values on this well-behaved dataset. All four indices have clear maxima at $c = 5$, which points to the most-preferable partition in CP. Fig. 7 might tempt you to conjecture that the index of Hullermeier and Rifqi is bounded above by the other three indices, but we have not attempted a proof of this. The other three indices all cross each other in Fig. 7; therefore, this is the only conjecture suggested by these graphs.

C. Example 3

This example uses the same dataset 3 and reference partition V as Example 2, but candidate partitions are generated by the *expectation–maximization* (EM) algorithm for Gaussian-mixture decomposition [4, 8]. The EM partitions are generated using our MATLAB implementation, where MAXIT = 100, EPS = 1e–5, means are initialized using HCM (MATLAB *k-means* function), initial random but normalized mixing coefficients, and identity covariance matrices. Thus, the elements of CP are probabilistic c-partitions of dataset 3. We compare the same four fuzzy-Rand indices, as in Example 2, but as probabilistic comparison indices. Fig. 8 graphs values of these four indices. Comparison of Figs. 7 and 8 shows that the c-partitions found by EM on dataset 3 are quite similar to those produced by FCM, a result that could be anticipated in view of the structure of dataset 3. The trends and shapes of the four graphs are quite similar, again supporting our conjecture about the relationship of the Hullermeier and Rifqi index to the other three. The main point of this example is to show that fuzzy generalizations of comparison indices may be quite effective when used to compare candidate partitions from probabilistic clustering.

D. Example 4

In this last example, we show that all 15 formulas for comparison indices in Table II can be used with our soft-generalization scheme. Nine candidate partitions are again generated with FCM, using the same parameters, as in Example 1. Dataset 4 consists of three Gaussians clouds with 100 samples each, means $[1\ \ 1]^T$, $[1\ \ 40]^T$, and $[20\ \ 10]^T$, and standard deviation of 1.5. The reference partition V for this example is the 3×300 matrix with three "diagonal" blocks of one hundred 1's in each row. Table III lists the values of each index on the nine terminal partitions for $c = 2$–10. Scanning this table, with the shaded cells indicating the choice of each index, shows that each of the 14 indices points to $c = 3$, which is the preferred solution for this well-separated set of three clusters. Equations (4b) and (4j) correctly have minima at $c = 3$, while all other equations correctly have maxima at $c = 3$. This example verifies that we

TABLE III
RESULTS FOR THE 15 FORMULAS PROVIDED IN TABLE II, DATASET 4, THE FCM, AND N^*

	c=2	c=3	c=4	c=5	c=6	c=7	c=8	c=9	c=10
Eqn 4a	0.7410	0.9883	0.9341	0.8787	0.9032	0.8447	0.7921	0.8422	0.8203
Eqn 4b	0.2590	0.0117	0.0659	0.1213	0.0968	0.1553	0.2079	0.1578	0.1797
Eqn 4c	0.4819	0.9765	0.8681	0.7575	0.8064	0.6895	0.5842	0.6844	0.6405
Eqn 4d a/(a+b)	0.5686	0.9823	0.9837	0.9816	0.9847	0.9867	0.9848	0.9867	0.9879
Eqn 4d a/(a+c)	0.9023	0.9823	0.8144	0.6496	0.4896	0.5290	0.4053	0.3517	0.3345
Eqn 4e	0.7163	0.9823	0.8950	0.7962	0.8414	0.7288	0.6101	0.7235	0.6762
Eqn 4f	0.5356	0.9651	0.8035	0.6381	0.7109	0.5344	0.3758	0.5268	0.4602
Eqn 4g	0.4907	0.9735	0.8444	0.7004	0.7655	0.6040	0.4442	0.5967	0.5313
Eqn 4h	-2.7182	0.9646	0.8014	0.6347	0.7087	0.5327	0.3738	0.5251	0.4590
Eqn 4i	0.4907	0.9735	0.8444	0.7004	0.7655	0.6040	0.4442	0.5967	0.5313
Eqn 4j	0.7021	0.1883	0.4905	0.7461	0.6327	0.9271	1.2791	0.9415	1.0764
Eqn 4k	0.5310	0.9735	0.8521	0.7295	0.7841	0.6541	0.5300	0.6483	0.5980
Eqn 4l	1.5212	37.1486	4.2533	1.7979	2.5160	1.1575	0.6036	1.1220	0.8567
Eqn 4m	238.6712	584.1974	582.0073	581.3560	572.3248	582.3302	586.2185	573.7462	578.7262
Eqn 4n	0.7956	0.9737	0.7459	0.5502	0.6318	0.4403	0.2922	0.4327	0.3685

TABLE IV
COMPUTATIONAL COMPLEXITY FOR THE FIVE FUZZY-RAND INDICES

Method	**Computational Complexity**		**n=1000, r=c=5**
Our approach for (3a)	$\boldsymbol{O}$(2rcn+3rc)	$\boldsymbol{O}$(n)	50,075
Brouwer [37] for (3a) (assuming cosine as a single operation)	$\boldsymbol{O}(2n^2+n+2)$	$\boldsymbol{O}(n^2)$	2,001,002
Brouwer [37] for (3a) (using dot product and magnitude form)	$\boldsymbol{O}(4n^2+ rn^2+ cn^2-n+3rn+3cn+2)$	$\boldsymbol{O}(n^2)$	14,029,002
Campello [35] for (3a)	$\boldsymbol{O}(rn^2+cn^2+(n-n^2)/2-rn-cn)$	$\boldsymbol{O}(n^2)$	9,490,500
Hullermeier and Rifqi [36] (for the fuzzy Rand)	$\boldsymbol{O}((3rn^2+3cn^2-3rn-3cn)/2+5)$	$\boldsymbol{O}(n^2)$	14,985,005

have valid fuzzy generalizations of all 15 formulas in Table II. Moreover, as noted above, there are (many) other comparison indices that depend only on the entries of the contingency matrix $N = UV^T$; hence, our method is indeed quite general.

Lastly, we make a remark about computational complexity. Assuming similar cost for different operations, the cost of evaluation of (3a) for all but the index of Hullermeier and Rifqi (who form the fuzzy-Rand index directly) are reported in Table IV.

The example in the last column of Table IV shows that in terms of computational costs, the Anderson *et al*. method is (at least 3 and at best 5) orders of magnitude less than the other generalizations of Rand's index. Computation of the Rand index in (4a) involves calculation of all four equations, i.e., (3a)–(3d). Combining the factors, as in (4a), uses only addition and subtraction and will cost all methods equally. Hence, we can extend the results of Table IV from just (3a) to (4a) without loss.

V. DISCUSSION AND CONCLUSION

We have developed a method to generalize comparison indices $s(U, V)$ to all possible cases for U and V: crisp, fuzzy, probabilistic, and possibilistic. Our method is applicable to *every* index whose value is a function of a certain $r \times c$ contingency table whose row sums and column sums count numbers of each of the four pair types (i.e., together in both, together in neither, and together in either but not both). We compared our generalization of the classical Rand index with three other fuzzy generalizations of it both computationally, as well as in terms of computational complexity. We showed that our extension of the Rand index is $O(n)$, while the other three are all $O(n^2)$. Our numerical examples included the use of our soft comparison indices to validate both probabilistic and possibilistic partitions of the data.

The use of comparison indices for validation of clustering algorithms has the significant advantage of being independent of the correspondence problem to compare clustering solutions with known reference partitions. Rand's index has enjoyed a real resurgence of interest and usage in the recent bioinformatics community [30]– [34]; therefore, the present work, which provides a method to generalize this index for use with soft clustering algorithms, is both timely and important. In this paper, we have exemplified the use of our indices in only one (viz., A1) of the six problems (i.e., A1–A6) listed in Section I. Future studies will concentrate on the use of soft comparison indices for one or more of the remaining problems. Specifically, we believe that we can use soft comparison indices as internal criteria to compare candidate partitions and will focus our next efforts on this problem.

To conclude, we offer this observation. Validity indices have surprising and, very often, unpredictable dependency on

elements of the solution that seem, at first glance, to be rather unrelated to their job—which indicates whether or not to believe the outputs. However, clustering is used extensively in real applications, such as data mining; therefore, it is important to continue studying ways in which to validate candidate solutions.

Even if the objects being clustered are well separated into c recognizable subsets, there are many reasons why we may not discover this structure through clustering. For example, the numerical representation of the objects may not possess adequate information to discriminate between clusters of objects. Further, even if the data possess the desired substructure, the algorithm used may not extract it from the data. (For example, an algorithm, which looks for hyperspherical clusters, will not extract shell-type clusters.) Finally, the objects may have structure, the data may represent it, and the algorithm may be capable of finding it, but the appropriate parameters of the algorithm that yield a successful interpretation of X are never used. Even if all of these obstacles are met, validity indices may fail to indicate that the great clusters are indeed great!

REFERENCES

[1] E. Anderson, "The **IRIS**es of the Gaspe peninsula," *Bull. Amer. IRIS Soc.*, vol. 59, pp. 2–5, 1935.

[2] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.*, vol. 7, no. 2, pp. 179–188, 1936.

[3] D. Titterington, A. Smith, and U. Makov, *Statistical Analysis of Finite Mixture Distributions*. New York: Wiley, 1985.

[4] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum, 1981.

[5] J. C. Bezdek, J. M. Keller, R. Krishnapuram, and N. R. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*. Norwell, MA: Kluwer, 1999.

[6] A. Jain and R. Dubes, *Algorithms for Clustering Data*. Englewood Cliffs, NJ: Prentice-Hall, 1988.

[7] B. S. Everitt, *Graphical Techniques for Multivariate Data*. New York: North-Holland, 1978.

[8] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 3rd ed. New York: Academic, 2006.

[9] R. Duda and P. Hart, *Pattern Classification and Scene Analysis*. New York: Wiley Interscience, 1973.

[10] R. Xu and D. C. Wunsch, *Clustering*. Piscataway, NJ: IEEE Press, 2009.

[11] J. Valente de Oliveira and W. Pedrycz, *Advances in Fuzzy Clustering and Its Applications*. West Sussex, U.K.: Wiley, 2007.

[12] P. Arabie, L. J. Hubert, and R. Desoete, *Clustering and Classification*. River Edge, NJ: World Scientific, 1996.

[13] F. Hoppner, F. Klawonn, R. Kruse, and T. Runkler, *Fuzzy Cluster Analysis*. Chichester, U.K.: Wiley, 1999.

[14] P. H. A. Sneath and R. R. Sokal, *Numerical Taxonomy—The Principles and Practice of Numerical Classification*. San Francisco, CA: Freeman, 1973.

[15] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *J. Amer. Stat. Assoc.*, vol. 66, no. 336, pp. 846–850, 1971.

[16] L. J. Hubert and P. Arabie, "Comparing partition," *J. Classification*, vol. 2, pp. 193–218, 1985.

[17] J. C. Bezdek, W. Q. Li, Y. A. Attikiouzel, and M. P. Windham, "A geometric approach to cluster validity," *Soft Comput.*, vol. 1, pp. 166–179, 1997.

[18] S. Bandyopadhyay and U. Maulik, "Non-parametric genetic clustering: Comparison of validity indices," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 31, no. 1, pp. 120–125, Feb. 2001.

[19] M. Halkidi, Y. Batistakis, and M. Vazigiannis "Cluster validity methods: Part I," *ACM Sigmod. Record*, vol. 31, no. 2, pp. 41–45, 2002.

[20] M. Halkidi, Y. Batistakis, and M. Vazigiannis, "Cluster validity methods: Part II," *ACM Sigmod. Record*, vol. 31, no. 3, pp. 19–27, 2002.

[21] W. Wang and Y. Zhang, "On fuzzy cluster validity indices," *Fuzzy Sets Syst.*, vol. 158, pp. 2095–2117, 2007.

[22] S. C. Johnson, "*Metric clustering*," unpublished.

[23] B. G. Mirkin and L. B. Chernyi, "Measurement of the distance between distinct partitions of a finite dataset," *Autom. Remote Control*, vol. 31, pp. 786–792, 1970.

[24] P. Arabie and S. A. Boorman, "Multidimensional scaling of measures of distance between partitions," *J. Math. Psychol.*, vol. 10, pp. 148–203, 1973.

[25] L. Hubert, "Nominal scale response agreement as a generalized correlation," *Brit. J. Math. Stat. Psychol.*, vol. 30, pp. 98–103, 1977.

[26] D. L. Wallace, "Comment," *J. Amer. Stat. Assoc.*, vol. 78, pp. 569–579, 1983.

[27] E. B. Fowlkes and C. L. Mallows, "A method for comparing two hierarchical clusterings," *J. Amer. Stat. Assoc.*, vol. 78, pp. 553–569, 1983.

[28] P. Jacard, "Novelles recherches sur la distribution florale," *Bull. Soc. Vaud. Sci. Nat.*, vol. 44, pp. 223–270, 1908.

[29] R. R. Sokal and C. D. Michener, "A statistical method for evaluating systematic relationships," *Univ. Kansas Sci. Bull.*, vol. 38, pp. 1409–1438, 1958.

[30] D. Jiang, C. Tang, and A. Zhang, "Cluster analysis for gene-expression data: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 11, pp. 1370–1386, Nov. 2004.

[31] F. Duan and H. Zhang, "Correcting the loss of cell-cycle synchrony in clustering analysis of microarray data using weights," *Bioinformat.*, vol. 20, no. 11, pp. 1766–1771, 2004.

[32] A. Thalamuthu, I. Mukhopadhyay, X. Zheng, and G. Tseng, "Evaluation and comparison of gene clustering methods in microarray analysis," *Bioinformat.*, vol. 22, no. 19, pp. 2405–2412, 2006.

[33] D. S. V. Wong, F. K. Wong, and G. R. Wood, "A multi-stage approach to clustering and imputation of gene expression profiles," *Bioinformat.*, vol. 23, no. 8, pp. 998–1005, 2007.

[34] Z. Yu, H. S. Wong, and H. Wang, "Graph-based consensus clustering for class discovery from gene expression data," *Bioinformat.*, vol. 23, no. 21, pp. 2888–2896, 2007.

[35] R. J. G. B. Campello, "A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment," *Pattern Recognit. Lett.*, vol. 28, pp. 833–841, 2007.

[36] E. Hullermeier and M. Rifqi, "A fuzzy variant of the Rand index for comparing clustering structures," in *Proc. IFSA*, Lisbon, Portugal, 2009, pp. 1–6.

[37] R. K. Brouwer, "Extending the Rand, adjusted Rand, and Jaccard indices to fuzzy partitions," *J. Intell. Inf. Syst.*, vol. 32, pp. 213–235, 2009.

[38] L. A. Goodman and W. H. Kruskal, "Measures of association for cross-classifications," *J. Amer. Stat. Assoc.*, vol. 49, pp. 732–764, 1954.

[39] H. Z. Ning, W. Xu, Y. Chi, Y. Gong, and T. S. Huang, "Incremental spectral clustering with application to monitoring of evolving blog communities," in *Proc. SIAM Int. Conf. Data Mining*, 2007, pp. 261–272.

[40] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proc. ACM SIGIR Conf. Inf. Retrieval*, 2003, pp. 267–273.

[41] D. Cai, X. He, and J. Han, "Document clustering using locality preserving indexing," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 2, pp. 1624–1637, Dec. 2005.

[42] L. Lovasz and M. Plummer, *Matching Theory*. Amsterdam, The Netherlands: North-Holland, 1986.

[43] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 2, pp. 98–110, May 1993.

Derek T. Anderson (S'06) received the M.S. degree in computer science in 2005 from the University of Missouri, Columbia, where he is currently working toward the Ph.D. degree with the Electrical and Computer Engineering Department.

He is currently a National Library of Medicine Predoctoral Fellow and a Resident Instructor of information technology with the Computer Science Department, University of Missouri. His research interests include pattern recognition, computer vision, linguistic summarization of human activity, information fusion, and clustering.

Mr. Anderson received the Best Student Paper Award at the IEEE International Conference on Fuzzy Systems at the World Congress on Computational Intelligence 2008.

James C. Bezdek (M'80–SM'90–F'92–LF'09) received the Ph.D. degree in applied mathematics from Cornell University, Ithaca, NY, in 1973.

He retired in 2007. He is currently the Founding Editor of *International Journal of Approximate Reasoning*. His current research interests include woodworking, optimization, motorcycles, pattern recognition, cigars, clustering in very large data, fishing, coclustering, blues music, visual clustering, and hold'em poker.

Dr. Bezdek is the Past President of the North American Fuzzy Information Processing Society, the International Fuzzy Systems Association (IFSA), and the IEEE Computational Intelligence Society (CIS). He is currently the Founding Editor of the IEEE TRANSACTIONS ON FUZZY SYSTEMS. He is a Life Fellow of the IFSA. He is the recipient of the IEEE 3rd Millennium, the IEEE CIS Fuzzy Systems Pioneer, and the IEEE CIS technical field award Rosenblatt medals.

Mihail Popescu (M'08) received the B.S. degree in electrical engineering from the Polytechnic Institute, Bucharest, Romania, in 1987 and the first M.S. degree in medical physics, the second M.S. degree in electrical engineering, and the Ph.D. degree in computer science, all from the University of Missouri, Columbia, in 1995, 1997, and 2003, respectively.

He is currently an Assistant Professor with the Informatics Institute, University of Missouri. During 1990–1993, he was an Assistant Professor of electrical engineering with the Polytechnic Institute.

Dr. Popescu is currently a member of the Engineering in Medicine and Biology Society.

James M. Keller (M'79–SM'92–F'00) received the Ph.D. degree in mathematics from the University of Missouri, Columbia, in 1978.

He holds the Curators Professorship with the Department of Electrical and Computer Engineering and the Department Computer Science, University of Missouri, Columbia, where he is also the R.L. Tatum Professor with the College of Engineering. His industrial and government funding sources include the Electronics and Space Corporation, the Union Electric, the Geo-Centers, the National Science Foundation, the Administration on Aging, The National Institutes of Health, The National Aeronautics and Space Administration/Johnson Space Center, the Air Force Office of Scientific Research, the Army Research Office, the Office of Naval Research, the National Geospatial Intelligence Agency, the Leonard Wood Institute, and the Army Night Vision and Electronic Sensors Directorate. He has authored or coauthored more than 300 technical publications. He is currently an Associate Editor of the *International Journal of Approximate Reasoning* and a member of the Editorial Board of the *Pattern Analysis and Applications*, *Fuzzy Sets and Systems,* the *International Journal of Fuzzy Systems*, and the *Journal of Intelligent and Fuzzy Systems.* His research interests include computational intelligence: fuzzy-set theory and fuzzy logic, neural networks, and evolutionary computation with a focus on problems in computer vision, pattern recognition, and information fusion, including bioinformatics, spatial reasoning in robotics, geospatial intelligence, sensor and information analysis in technology for eldercare, and landmine detection.

Prof. Keller is a Fellow of the International Fuzzy Systems Association, a Distinguished Lecturer of the IEEE Computational Intelligence Society, a National Lecturer for the Association for Computing Machinery from 1993 to 2007, and a Past President of the North American Fuzzy Information Processing Society (NAFIPS). He received the 2007 Fuzzy Systems Pioneer Award from the IEEE Computational Intelligence Society. He had a full six-year term as Editor-in-Chief of the IEEE TRANSACTIONS ON FUZZY SYSTEMS. He was the Vice President for Publications of the IEEE Computational Intelligence Society from 2005 to 2008 and is currently an elected Adcom member. He was the Conference Chair of the 1991 NAFIPS Workshop, a Program Cochair of the 1996 NAFIPS meeting, a Program Cochair of the 1997 IEEE International Conference on Neural Networks, and the Program Chair of the 1998 IEEE International Conference on Fuzzy Systems. He was the General Chair of the 2003 IEEE International Conference on Fuzzy Systems.