Fuzzy order-equivalence for similarity measures

Maria Rifqi, Marie-Jeanne Lesot and Marcin Detyniecki

Abstract— **Similarity measures constitute a central component of machine learning and retrieval systems, and the choice of an appropriate measure is of major importance. In this paper, we consider this issue from the point of view of the order induced by the measures when comparing a set of objects to a given reference, i.e. the ranking from the most similar object to the least similar one. We introduce the notion of fuzzy order-equivalence, based on degrees that quantify the extent to which the induced orders differ. We define these degrees using the generalized Kendall's rank correlation, taking into account the number of order permutations as well as their positions.**

We then present an automatic and hierarchical classification of usual similarity measures that makes it possible to indicate, for a given number of tolerated variations, the measures that will yield rankings without significant changes; it thus provides a guideline for set data similarity measure selection.

I. Introduction

Most machine learning tasks rely on the use of a similarity measure, or a distance function, to compare objects one with another, for instance with the aim of determining whether they can be assigned to the same class. Now, in many applications, the similarity values themselves have no importance, only the induced rankings matter, as pointed out by [1], [2]. This is for instance the case for document retrieval systems: the user is interested in the list of documents most similar to her request, ignoring the similarity score of each document. Likewise, monotone equivariant cluster analysis [3] constitutes a clustering approach where only the similarity rankings matter, and not their numerical values.

In such cases, the choice between one similarity measure and another is of little interest if the two measures lead to the same ordered lists. To formalize this observation, several authors introduced the notion of order equivalence [4], [3], [5], [6], defining two measures as equivalent if they induce the same ranking. Classes of equivalent measures for set data, i.e. groups of measures that always lead to the same object ranking, were exhibited.

In this paper, we study the non equivalent similarity measures, and propose to differentiate more finely between them: even if measures do not lead to the same rankings, they can for instance differ to a small extent, e.g. if they only lead to few rank permutations, or to a large extent, e.g. if they lead to opposite rankings. Moreover, two measures can be considered as more or less alike depending on whether these rank permutations occur for the highest similarity values (corresponding to the first retrieved objects) or only for the smallest ones.

Maria Rifqi, Marie-Jeanne Lesot and Marcin Detyniecki are with the University Pierre et Marie Curie-Paris6, CNRS UMR 7606, LIP6, 104, av. Kennedy, Paris, F-75016, France, email: {Maria.Rifqi, Marie-Jeanne.Lesot, Marcin.Detyniecki}@lip6.fr.

To study this measure proximity, we introduce the notion of fuzzy order-equivalence, and propose to quantify it with degrees of equivalence based on the comparison of the top-k ordered similarity values. To compare rankings, we use the generalized Kendall measure [7], [8]

The paper is organized as follows: Section II recalls the notion of set data similarity measures and the order equivalence. In Section III we extend this notion to fuzzy order-equivalence and we define degrees to quantify its extent. Lastly in Section IV we present a classification of set data similarity measures, based on a hierarchical clustering: the latter yields nested decompositions of the set of similarity measures according to their relative fuzzy equivalence, providing a guideline for the similarity measure selection problem.

II. State of the art

A. Set data similarity measures

Similarity measures are functions that associate to data point couples numerical values in the interval $[0,1]$, indicating the extent to which the two data are similar one to another. In the following, we denote $\mathcal{X}$ a finite set of data, and we focus on set data, also called binary data, i.e. data described by means of m binary attributes that denote the presence or absence of characteristics. Indeed, existing works on equivalent measures (see Section II-B) belong to this framework.

Given two data points x and y, we denote $X = \{i | x_i = 1\}$ and $Y = \{i | y_i = 1\}$ the set of attributes present in data point x and y respectively, and $|.|$ the cardinality of a set.

Set data similarity measures can be expressed as functions of the following 4 quantities associated with the attributes of a couple of data points $(x, y) \in \{0,1\}^p \times \{0,1\}^p$:

- the number of attributes common to both points, $|X \cap Y|$, denoted a.
- the number of attributes present in x but not in y, $|X - Y|$, denoted b.
- the number of attributes present in y but not in x, $|Y - X|$, denoted c.
- the number of attributes in neither x nor y, $|\bar{X} \cap \bar{Y}|$, denoted d.

Table I gives the definition of 10 classic similarity measures in this context, that we will study in the following. All measures were normalized to $[0,1]$, thus the given definitions may not be exactly the classic ones.

These similarity measures can be divided into two groups. The first 4 measures only depend on the characteristics present in x or y and do not take into account the number of characteristics possessed by none of the objects (i.e. d); they

TABLE I

CLASSIC SET DATA SIMILARITY MEASURES.

	Similarity Measure	*Notation*	*Definition*
1	Jaccard	S_{Jac}	$\frac{a}{a+b+c}$
2	Dice	S_{Dic}	$\frac{2a}{2a+b+c}$
3	Ochiai	S_{Och}	$\frac{a}{\sqrt{a+b}\sqrt{a+c}}$
4	Kulczynski 2	S_{Kul}	$\frac{1}{2}\left(\frac{a}{a+b}+\frac{a}{a+c}\right)$
5	Rogers and Tanimoto	S_{RT}	$\frac{a+d}{a+2(b+c)+d}$
6	Simple Matching	S_{SM}	$\frac{a+d}{a+b+c+d}$
7	Sokal and Sneath 1	S_{SS1}	$\frac{a+d}{a+\frac{1}{2}(b+c)+d}$
8	Russel and Rao	S_{RR}	$\frac{a}{a+b+c+d}$
9	Yule Q	S_{YuQ}	$\frac{ad}{ad+bc}$
10	Yule Y	S_{YuY}	$\frac{\sqrt{ad}}{\sqrt{ad}+\sqrt{bc}}$

are called *type I similarity measures*. The 6 last measures take d into account and are called *type II similarity measures*.

As can be seen from the table, the first 2 measures follow the same scheme and are of the form $\kappa a/(\kappa a+b+c)$. They belong to the general contrast model proposed by Tversky [9] in which a similarity measure is of the form $S(x,y) = a/(a+\alpha b+\beta c)$ and correspond to the special case where $\alpha = \beta = 1/\kappa = 2^{-n}$ with $n = 0$ and $n = 1$ respectively. Other classic measures in this framework are the Sorensen (corresponding to $n = 2$), Symmetric Anderberg ($n = 3$) and Sokal and Sneath 2 ($n = -1$) measures.

B. Similarity measure order equivalence

1) Order equivalence definition: Several authors [3], [5], [6] have considered the problem of a theoretical comparison between similarity measures and have proposed the notion of equivalence. Two measures are said equivalent if they induce the same order when comparing objects with a given reference. More formally:

Definition 1: Two similarity measures S_1 and S_2 are *equivalent* if and only if, $\forall x,y,z,t \in \mathcal{X}^4$,

$$S_1(x,y) < S_1(z,t) \iff S_2(x,y) < S_2(z,t)$$

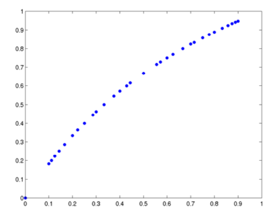
It has been shown [6], [10] that this definition is equivalent to the following one: S_1 and S_2 are equivalent if and only if there exists a strictly increasing function $f : Im(S_1) \rightarrow Im(S_2)$ such that $S_2 = f \circ S_1$, where $Im(S) = \{s \in [0,1]/\exists(x,y) \in \mathcal{X}^2, s = S(x,y)\}$.

Other definitions of equivalence have been proposed. For instance, Lerman [4] considers it from a different point of view: he defines the notion of equivalence with respect to a given data set, whereas Definition 1 requires it holds generally for any data set. Lerman studies the conditions on the given data set under which equivalence holds. He shows that if all data have the same number of present attributes, i.e. if there exists p such that $\forall x \in \mathcal{X}, |X| = p$, then all similarity measures are equivalent on $\mathcal{X}$.

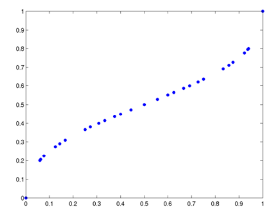
2) Classes of equivalent measures: Using Definition 1, several classes of equivalent similarity measures can be shown, with the property that measures of the same class always provide the same ranking when comparing a set of data to a given reference. The similarities defined in Table I can be grouped in the following classes:

- {Jaccard, Dice, symmetrical Tversky's measures}
- {Rogers and Tanimoto, Simple Matching, Sokal and Sneath 1},
- {Yule Q, Yule Y},
- each of the remaining measures forms a class by itself.

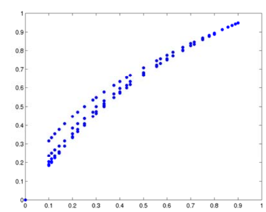
For Tversky's parametrized similarity measures, it was more generally shown [10] that two Tversky's measures with parameters (α,β) and (α',β') are equivalent if and only if $\alpha \cdot \beta' = \alpha' \cdot \beta$. This in particular implies that all symmetrical Tversky measures, for which $\alpha = \beta$, are equivalent.



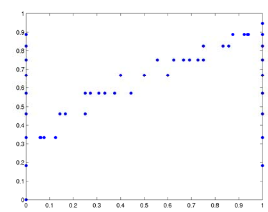
(a) S_{Jac} vs S_{Dic}



(b) S_{YuQ} vs S_{YuY}



(c) S_{Jac} vs S_{Och}



(d) S_{YuQ} vs S_{SS1}

Fig. 1. Graphical pairwise comparisons of some similarity measures, computed for data uniformly sampled in $\{0,1\}^{10}$.

3) Graphical representation: As stated in Definition 1, equivalent measures are linked by a function. This function can be graphically exhibited by plotting a measure S_1 against another one S_2. Figure 1 shows such plots for 4 similarity measures couples: data were uniformly sampled in $\{0,1\}^{10}$ and the similarity values were computed for all pairs of distinct data. Each point on the graph corresponds to such a couple, and has for coordinates the values taken by S_1 and S_2 for this data pair. It can be noticed that the plots only contain few points: the similarity measures only take few different values and many data couples lead to the same point on the plot.

From such graphs, in the case of equivalent measures (Fig. 1a and 1b), we observe the function linking the two measures. An analytical study leads to $S_{Dic} = f(S_{Jac})$ with $f(x) = 2x/(1+x)$ and $S_{YuQ} = g(S_{YuY})$ with $g(x) = x^2/(2x^2-2x+1)$.

For non-equivalent measures (Fig. 1c and 1d), the plot is not a curve, but a scattering of points. For instance Fig. 1d shows that a particular value of the Yule Q measure corresponds to several values of the Sokal and Sneath 1 measure and vice versa; thus YuQ cannot be written as a function of SS1.

III. Fuzzy order-equivalence degree definition

Two measures are said to be equivalent if they induce the same order. We propose to study in more detail the non-equivalent similarity measures and to examine more precisely the extent to which the rankings induced by two non-equivalent measures differ. To that aim, we introduce the notion of fuzzy order-equivalence and its quantification through a degree of equivalence. We first detail the principle of such degrees, and then recall the Kendall's rank correlation on which they are based; lastly we define them formally.

A. Principle

As Figure 1 shows, it is possible to discriminate between non-equivalent measures: from Figures 1c and 1d, it can be said that Jaccard and Ochiai have more correlated behaviors than Yule Q and Sokal and Sneath 1 as the spreading of points is smaller for the first pair than for the second one. Thus one would intuitively say that Jaccard and Ochiai are "more equivalent" one to another than the YuQ and SS1 measures. Indeed the spreading of the point scattering is related to the possibility of ranking permutation: if the value given by S_1 corresponds to a set of different values for S_2, it is likely that some object couples will be ranked differently by the two measures. In other words, a wider scattering increases the inversion risks.

Thus it is relevant to study and quantify the extent to which the non-equivalent similarity measures differ. The comparison of the induced ranks must take into account two components, namely the *quantity* and the *location* of the occurring rank inversions. On one hand, if the rankings are totally different, i.e. opposite, then the two similarity measures are completely different; if there are few rank inversions, then the two similarity measures are very close. On the other hand, given a fixed number of inversions, their locations also matter: if the inversions occur for the low similarity values, i.e. the least similar data points, the measures will be better fit as quasi-equivalent than in the case where the rankings differ for high similarity values, which correspond to the first retrieved objects in the case of a retrieval system.

Therefore we propose to define degrees of equivalence through a quantified comparison of the rankings, taking into account both the number of occurring permutations and their positions. To that aim, we exploit a generalization of the Kendall's rank correlation, as detailed in the following.

It is to be noticed that there exist other works about non-equivalent measures exploiting the inversion numbers: Lerman [4] generalizes his results about data sets on which measures are equivalent (see Section II-B.2), and shows that for data having a low variance of the number of present attribute for each data point $|X|$, the number of inversions is low. Batagelj et al. [6] also exploit the number of inversions, but they do not consider the positions of ranking inversions nor the case of ranking ties.

B. Comparing rankings

There are two major measures to compare rankings, Spearman's rank correlation and Kendall's rank correlation, which were proved to be equivalent [11]. In the following, we consider a generalization of the Kendall's tau.

Let n denote the number of objects to be ranked, and σ_1 and σ_2 two orders defined on this universe: they are vectors of dimension n whose i-th component σ_i indicates the rank of the i-th object. In the context of similarity measures, it is to be underlined that an object corresponds to the similarity value obtained by a data point pair: n corresponds to the number of distinct data pairs, i.e. $n = |\mathcal{X}|(|\mathcal{X}|-1)/2$.

1) Original Kendall's tau: The original Kendall's tau was defined as the frequency of pair-wise rank inversions: it computes the number of object pairs (i, j) that are not in the same order in σ_1 and σ_2, i.e. such that $\sigma_1(i) < \sigma_1(j)$ and $\sigma_2(i) > \sigma_2(j)$ or vice-versa. Such pairs are called discordant pairs; pairs for which the rankings agree are called concordant pairs.

As the total number of distinct pairs is $n(n-1)/2$, denoting D the number of discordant pairs, the Kendall's tau is formally defined as $K(\sigma_1, \sigma_2) = 2D/n(n-1)$. More generally,

$$K(\sigma_1, \sigma_2) = \frac{2}{n(n-1)} \sum_{i \neq j} P_{\sigma_1,\sigma_2}(i, j) \tag{1}$$

where $P_{\sigma_1,\sigma_2}(i, j)$ denotes the penalty associated to the pair (i, j): it equals 0 if (i, j) is a concordant pair, and 1 if it is discordant pair.

Thus K corresponds to the proportion of discordant pairs between the two rankings. It takes value 0 when the two orders are one and the same, and its maximum value, 1, when σ_1 is the reverse of σ_2.

2) Generalized Kendall's tau: The above definition applies to the case where $\sigma_1$and σ_2 are permutations, i.e. all objects are ranked and there is no object pair with identical ranks.

Fagin et al. [7], [8] consider the comparison of top-k lists, where only the top-k preferred objects for both rankings are taken into account. This implies that some objects may be present in one of the ranking, but not in the other one. Therefore they generalize the Kendall's tau to handle absent pairs. Moreover they propose to handle ties, i.e. several objects having the same rank.

To that aim, they identify categories other than concordant and discordant pairs and define associated penalties. More precisely, the generalized Kendall's tau distinguishes 4 different types of pairs (i, j):

- concordant and discordant pairs, as previously stated, respectively associated to penalty values 0 and 1;
- pairs that are tied in one ranking but not in the other one, i.e. $\sigma_1(i) = \sigma_1(j)$, but $\sigma_2(i) \neq \sigma_2(j)$, or vice-versa. In this case the penalty is $p \in [0, 1]$.
- pairs that are present in one ranking but not in the other, i.e. i and j both belong to the k preferred objects for σ_1, but not for σ_2, or vice-versa. In this case, the penalty

is $p' \in [0, 1]$ if both i and j are missing in the second ranking; if only one of the two is missing, the pair is treated as a normal one, and the penalty takes value 0, 1, or p depending on whether it is a concordant, discordant or tied pair respectively.

The generalized Kendall's tau is then defined as the total penalty divided by the total number of pairs. Thus its formal definition is identical to that in Equation (1), but the penalty function $P_{\sigma_1,\sigma_2}(i, j)$ can take values 0, 1, p or p'. To underline its parameters, we denote the generalized Kendall's tau as $K_{p,p'}(\sigma_1^k, \sigma_2^k)$, where σ_i^k represents the top-k values for the σ_i ranking.

C. Fuzzy order-equivalence degrees

Using the previous generalized Kendall's tau we can define, for two given similarity measures, equivalence degrees that measure the extent to which the orders they respectively induce are in agreement.

1) General form: As previously mentioned, equivalence degrees must take into account the number and positions of occurring inversions. The amount of inversions is directly computed in the Kendall's tau.

In order to take into account the positions, we compare the top-k lists for different k values. For example, if two measures only differ for low similarity values, the Kendall's tau computed for top-k lists with small k is low, and increases when k increases. This variation makes it possible to finely model the dependence on the inversion positions.

Thus we introduce a top-k degree of equivalence as

$$d_k(S_1, S_2) = 1 - K_{p,p'}(\sigma_1^k, \sigma_2^k)$$

where σ_i^k denotes the top-k values for the σ_i ranking induced from the S_i measure. Thus the equivalence degree equals 1 for equivalent measures and decreases when the disagreements between the two induced rankings increase.

2) Parameter choice: To completely define the degrees of equivalence, the penalty values p and p' have to be set.

We fix the tie penalty $p = 0.5$: indeed, if an object pair is tied in one ranking but not in the other one, the pair should be penalized, since it is not handled in the same way by both measures. Still, the disagreement between the two rankings is smaller than in the case of a discordant pair, thus the penalty should be lower than 1. Now when considering the tie, one has one chance out of two to come up with the same order as the one indicated by the untied ranking, therefore we define $p = 0.5$.

Regarding the absence penalty we set $p' = 1$: an object pair appearing in one top-k list but not in the other is penalized as a discordant pair. Indeed the absence of such an object pair indicates a major difference between the two rankings that can be penalized as much as a discordant pair.

3) Definitions: Thus the equivalence degree is finally defined as follows:

Definition 2: Given a finite data set $\mathcal{X}$, and two similarity measures S_1 and S_2, the *top-k degree of equivalence on $\mathcal{X}$* is defined as

$$d_k(S_1, S_2) = 1 - K_{0.5,1}(\sigma_1^k, \sigma_2^k)$$

where σ_i is the ranking on the pairs of data points from $\mathcal{X}$ compared using S_i, and σ_i^k denotes the top-k values for the σ_i ranking.

It is to be noticed that this definition is dependent on the set of data for which the similarity values are computed and ranked. In the case of equivalent measures, the data do not matter, as the notion of equivalence requires the same ordering for any data couples (see Definition 1); for non-equivalent measures, some inversions may not occur only for specific data sets. For a given data set, we thus define the notion fuzzy order-equivalent measures as

Definition 3: The *fuzzy order-equivalence of similarity measures at level k* is the fuzzy relation defined by the top-k degrees of equivalence as described in Definition 2.

IV. Fuzzy order-equivalence measure results

A. Equivalence degree values

In this section we perform an empirical study of the fuzzy order-equivalence, in the case of uniformly distributed data. The intuitive justification of this choice is to measure the probability of disagreement by uniformly covering the whole space.

Therefore, we sample data points uniformly in $\{0, 1\}^p$ and compute the similarity values obtained with the measures indicated in Table I. We then rank the obtained similarity values and compare the rankings using the proposed equivalence degrees (Definition. 2). Table II shows the obtained results for the full rank comparison and for the top-2 comparison.

1) Full rank comparison: It is first to be noted that the degrees for full rank comparison are globally high and remain above 0.69: this means that all measures are not so different one from another, the proportion of rank permutations does not exceed 31%.

The similarity measures couples for which the degree is 1 are those for which no discordant pairs occur, i.e. equivalent measures according to the equivalence definition given in Section II-B. One can read from the matrix the same equivalence classes as those recalled in Section II-B.2, which shows the consistency of the proposed equivalence degrees and the approach based on exhibiting the function linking the measures.

When the degree is strictly lower than 1, it is possible to quantify the extent to which non-equivalent measures differ: for instance one has $d_{\text{full}}(Jac, Och) = 0.972$ that is much higher that $d_{\text{full}}(YuQ, SS1) = 0.871$. This indicates that Jaccard is closer to Ochiai than Yule Q is to Sokal and Sneath 1. It corresponds to the intuitive comparison of Figures 1c and 1d where the point scattering spread is smaller for the first pair than for the second one.

Consequently this approach allows one to objectively assess the extent to which two similarity measures can be considered as equivalent, giving a more refined analysis than the concept of order equivalence. It provides hints regarding the choice of a similarity measure: for any given measure, it indicates those that will lead to very similar, and thus redundant, results and that should consequently not be

		Jac	Dic	Och	Kul	RR	SM	RT	SS1	YuY	YuQ
Full rank	Jac	1	1	0.972	0.947	0.891	0.819	0.819	0.819	0.790	0.790
	Dic		1	0.972	0.947	0.891	0.819	0.819	0.819	0.790	0.790
	Och			1	0.974	0.873	0.809	0.809	0.809	0.810	0.810
	Kul				1	0.858	0.805	0.805	0.805	0.823	0.823
	RR					1	0.709	0.709	0.709	0.691	0.691
	SM						1	1	1	0.871	0.871
	RT							1	1	0.871	0.871
	SS1								1	0.871	0.871
	YuY									1	1
	YuQ										1
Top-2	Jac	1	1	1	1	0.406	0.054	0.054	0.054	0.046	0.046
	Dic		1	1	1	0.406	0.054	0.054	0.054	0.046	0.046
	Och			1	1	0.406	0.054	0.054	0.054	0.046	0.046
	Kul				1	0.406	0.054	0.054	0.054	0.046	0.046
	RR					1	0.072	0.072	0.072	0.055	0.055
	SM						1	1	1	0.481	0.481
	RT							1	1	0.481	0.481
	SS1								1	0.481	0.481
	YuY									1	1
	YuQ										1

TABLE II

DEGREES OF EQUIVALENCE FOR THE FULL RANKING COMPARISON (TOP PART) AND TOP-2 COMPARISON (LOWER PART).

applied. It also gives, for any measure, the one that leads to the most different ranking, and thus could be considered. For instance it can be read from the table that the least similar measures to Russel and Rao are Yule Q and Yule Y.

2) Top-2 comparison: The lower part of Table II shows the equivalence degrees obtained when comparing the rankings of data pairs considering only those associated to the two highest similarity values. It can be seen that the obtained degrees highly differ as compared to the previous full ranking comparison: the lowest degree is now 0.046, which is actually mostly due to missing pairs and not to discordant ones. On the contrary, one can also notice that some measures now get equivalence degree 1, implying they are equivalent: when only considering the two highest similarity values, Ochiai and Kulczynski 2 belong to the same equivalence class as Jaccard and Dice. This corresponds to the graphical observation (see Fig. 1).

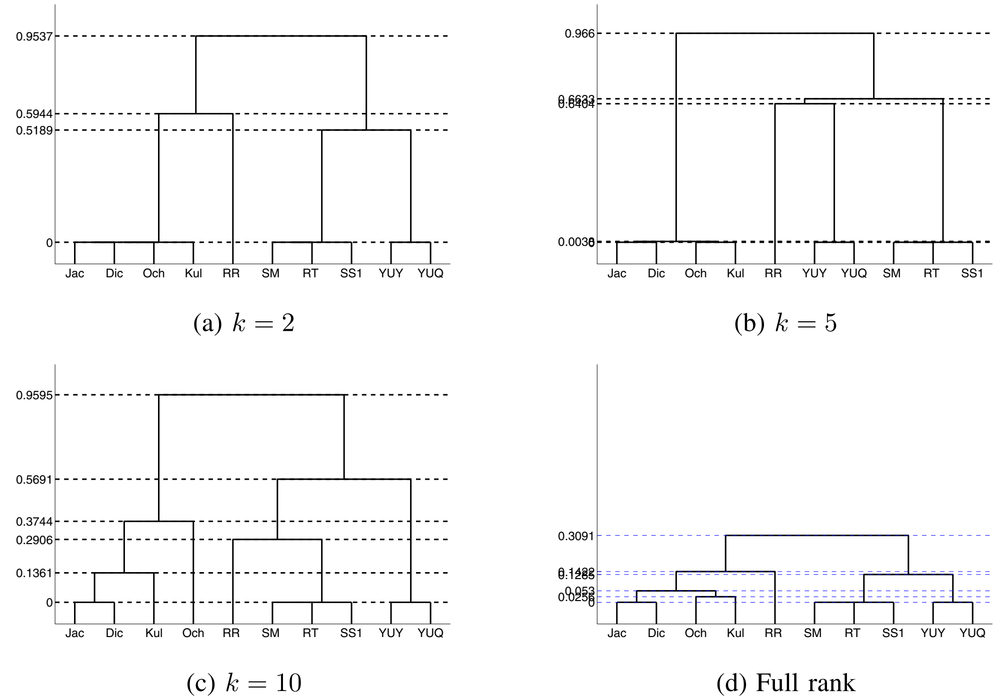
Thus the obtained results are quantified comparisons of similarity measures, of the following form: although Ochiai and Jaccard measures are considered as non-equivalent, they are equivalent if one focuses only on the 2 highest possible similarity values and they are equivalent to the degree 0.972 for the full rank comparison.

B. Classification of fuzzy order-equivalent measures

1) Methodology: In order to better exploit these results and to identify fuzzy order-equivalence classes, we apply a hierarchical clustering algorithm with complete linkage, using as distance measure the complement of the equivalence degrees, $1 - d_k$. Hierarchical clustering provides nested decompositions of the data, allowing to consider measures at different levels of equivalence. Furthermore, the complete linkage variant is based on the maximum operator for the cluster building, and thus provides upper bounds of the equivalence degrees within each cluster. Thus it provides information about the degrees with which measures within a given cluster can be considered as equivalent one to another.

2) Discussion: Figure 2 shows the obtained dendrograms for several k values. One can note that they differ from one k value to the other, but some constant features can be observed: in all cases, the studied measures can be separated into two major groups, that approximately correspond to the two measure types mentioned in Section II: the first one contains {Dice, Jaccard, Ochiai, Kulczynski2}, the second one {Simple Matching, Rogers and Tanimoto, SokalSneath1, Yule Y, Yule Q}, the Russel and Rao measure oscillates between these two groups. As expected, the obtained groups are superclasses of the order equivalence classes recalled in Section II-B.2.

Moreover, it can be observed that the first group is more homogeneous than the second one: measures in this group are equivalent to a degree higher than 0.95 (as their distance is lower than 0.05), except for $k = 10$. For the latter case, they are only equivalent to a degree 0.63. This is due to the fact that these measures agree for the highest similarity values and thus are equivalent for small k. For middle-range similarity values, more disagreements occur, leading to lower equivalence degrees. A better agreement is again achieved for the full rank comparison because the measures also agree for the least similar data couples.



(a) $k = 2$

(b) $k = 5$

(c) $k = 10$

(d) Full rank

Fig. 2. Dendrograms based on the equivalence degrees for different top-k comparisons.

Furthermore, through the nested partition it produces, the hierarchical clustering makes it possible to refine these measure groups. If one considers, for instance, the full ranking comparison, and if one sets a tolerance of 5% of rank inversions, 5 fuzzy equivalence groups can be identified. The previous first group $\{$Dice, Jaccard, Ochiai, Kulczynski2$\}$ is decomposed into 3 groups, $\{$Jac, Dice$\}$, $\{$Och, Kul$\}$ and $\{$RR$\}$ being a group of its own; the second cluster is split into the 2 equivalence classes recalled in Section II-B.2, $\{$SM, SS1, RT$\}$ and $\{$YuQ, YuY$\}$.

More generally, the hierarchical clustering with complete linkage enables a user to set the level of ranking permutations she accepts, and for this value, it indicates groups of measures that yield rankings with at least that level of agreement.

V. Conclusion and perspectives

In this paper we compared similarity measures extending the notion of order equivalence: when two measures do not induce exactly the same ordering of the data points, they can be compared finely, based on the extent to which they agree, and on the similarity values for which they agree. To formalize and quantify this notion we proposed to use the generalized Kendall's rank correlation at several top-k levels, and to apply a clustering algorithm to automatically extract quasi-equivalence classes and their associated degrees.

This study was performed at a theoretical level, considering as universe data sampled uniformly in $\{0, 1\}^p$. Since real data may follow a different distribution, a perspective of our work is to study fuzzy equivalence on real data, to examine in more details their specificity, if there is one.

References

[1] S. Santini and R. Jain, "Similarity measures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, Sept. 1999.

[2] J. S. Payne, L. Hepplewhite, and T. J. Stonham, "Perceptually based metrics for the evaluation of textural image retrieval methods," in *Int. Conf. on Multimedia Computing and Systems*, vol. 2, 1999, pp. 793–797.

[3] M. F. Janowitz, "Monotone equivariant cluster analysis," *Journal of Mathematical Psychology*, vol. 37, pp. 148–165, 1979.

[4] I. C. Lerman, "Indice de similarité et préordonnance associée," in *Séminaire sur les ordres totaux finis*, Aix-en-Provence, 1967, pp. 233–243.

[5] F. B. Baulieu, "A classification of presence/absence based dissimilarity coefficients," *Journal of Classification*, vol. 6, pp. 233–246, 1989.

[6] V. Batagelj and M. Bren, "Comparing resemblance measures," *Journal of Classification*, vol. 12, pp. 73–90, 1995.

[7] R. Fagin, R. Kumar, and D. Sivakumar, "Comparing top k lists," *SIAM Journal on Discrete Mathematics*, vol. 17, no. 1, pp. 134–160, 2003.

[8] R. Fagin, R. Kumar, M. Mahdian, D. Sivakumar, and E. Vee, "Comparing and aggregating rankings with ties," in *Symposium on Principles of Database Systems*, 2004, pp. 47–58.

[9] A. Tversky, "Features of similarity," *Psychological Review*, vol. 84, pp. 327–352, 1977.

[10] J.-F. Omhover, M. Rifqi, and M. Detyniecki, "Ranking invariance based on similarity measures in document retrieval," in *Adaptive Multimedia Retrieval AMR'05*. Springer LNCS, 2006, pp. 55–64.

[11] P. Diaconis and R. L. Graham, "Spearman's footrule as a measure of disarray," *Journal of the Royal Statistical Society, Series B*, vol. 39, no. 2, pp. 262–268, 1977.