Discrimination power of measures of resemblance

M. Rifqi, M. Detyniecki, and B. Bouchon-Meunier

LIP6 – UPMC – CNRS – Pôle IA
8, rue du Capitaine Scott
75015 Paris, France

Abstract. In this paper, we focus on exclusive measures of resemblance as introduced by [1]. We propose a method to compare them according to their discrimination power in order to help the user choosing one measure among others appropriately. We illustrate the behaviour of the main measures of resemblance.

1 Introduction

The evaluation of a proximity of two objects is a very common worry. Many methods lie on a measure of similarity (or dissimilarity) in different domains: clustering, case-based reasoning, image processing, ... This crucial step of evaluation of proximities implies an inspired choice of measure. There are very few comparative studies of existing measures of similarity. Generally, the comparisons are based on experiments or on mathematical properties like in [3], [10], [4], [2], [9] which is useful but not sufficient to understand the behaviour of a measure. Our approach is also based on basic mathematical properties (symmetry, exclusiveness, reflexivity,...): the study of these properties gives various types of measures of comparison [1] such as satisfiability, resemblance, inclusion, dissimilarity. We further propose to compare measures inside a same family on the basis of their power of discrimination. This power means that a measure can be qualified as severe or not, according to the values it takes when the objects to be compared are more or less close to each other. This global study is realized thanks to a geometrical representation of measures. In [5], we have studied measures of satisfiability, in this paper we focus on measures of resemblance.

2 General framework

Our study concerns measures of similarity that can be expressed in terms compatible with Tversky's contrast model [7]. It means that these measures can be decomposed in terms of distinctive features: $M(A \cap B)$ and distinctive features: $M(A - B)$ and $M(B - A)$. More particularly, we refer to the framework giving different families of similarities [1]. Measures of satisfiability were particularly studied in [5]. We focus on the second kind of mesures of similarity: the measures of resemblance. The problem is always the same: how can measures be compared inside a same family? Let us first define what a measure of resemblance is:

2.1 Measures of resemblance

Formally, for any set Ω of elements, let $F(\Omega)$ denote the set of fuzzy subsets of Ω. We suppose given a fuzzy set measure $M : \Omega \to R^+$ such that: $M(\emptyset) = 0$ and M is monotonous with respect to $\subseteq$. For instance:

- $M_1(A) = \int_\Omega f_A(x)dx$,
- $M_2(A) = \sup_{x\in\Omega} f_A(x)$,
- $M_3(A) = \sum_{count} f_A(x)$ if Ω is countable.

Definition 1 *An M-*measure of comparison *on Ω is a mapping $S : F(\Omega) \times F(\Omega) \to [0,1]$ such that $S(A,B) = F_S(M(A\cap B), M(B-A), M(A-B))$, for a given mapping $F_S : I\!R^+ \times I\!R^+ \times I\!R^+ \to [0,1]$ and a fuzzy set measure M on $F(\Omega)$.*

We denote:

- $X = M(A\cap B)$
- $Y = M(B-A)$
- $Z = M(A-B)$

Definition 2 *An M-*measure of resemblance *on Ω is a measure $S(A,B) = F_S(X,Y,Z)$ such that:*

- *F_S is increasing with X and decreasing with Y and Z*
- *$F_S(X,0,0) = 1$ for all X*
- *$F_S(X,Y,Z) = F_S(X,Z,Y)$.*

M-measures of resemblance which satisfy an additional property of t-transitivity, for a triangular norm t, are extensions of indistinguishability relations [6], [8] to fuzzy sets. In the case where t is the minimum, we obtain extensions of measures of similarity in the sense of Zadeh.

M-measures of resemblance satisfying the property of exclusiveness:

$$F_S(0,Y,Z) = 0 \quad \text{for all} \quad (Y,Z) \neq (0,0)$$

are called *exclusive M-measures of resemblance*. We focus on them in the sequel.

2.2 New representation

In order to compare all measures of a same family, we need to represent them in the same space. The new representation uses the spherical coordinates (r,μ,λ) of a point $P = (X,Y,Z)$ in cartesian representation (see figure 1). Cartesian coordinates are transformed into spherical coordinates by the following formulas:

$$X = r\sin\lambda\cos\mu \qquad (1)$$

$$Y = r\sin\lambda\sin\mu \qquad (2)$$

$$Z = r\cos\lambda \qquad (3)$$

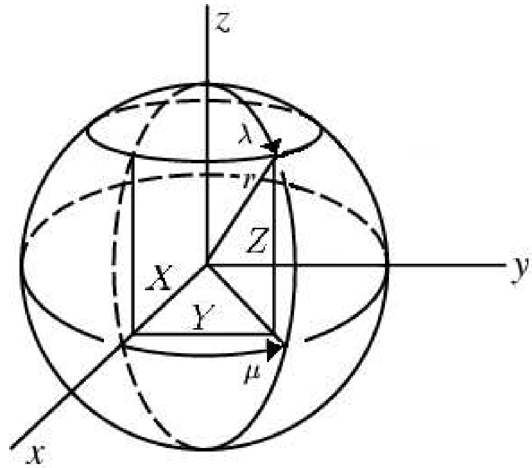

Fig. 1. New representation in spherical coordinates

where λ and μ belong to $[0, \frac{\pi}{2}]$ in our case.

Let us denote $\sigma(\mu, \lambda)$ the measure of resemblance expressed in spherical coordinates. The conditions of definition 2 become:

- σ is decreasing with respect to μ
- $\sigma(\frac{\pi}{2}, \lambda) = 0, \forall \lambda \neq \frac{\pi}{2}$
- $\sigma(0, \frac{\pi}{2}) = 1$

3 Comparison of classical measures of resemblance

A user who needs a measure of resemblance in his particular process needs to choose one measure of resemblance among several. This choice can be realized only if the different measures that are avalaible can be compared on the basis of some criteria. Thanks to this new space of representation, the behaviour of different measures of resemblance can be compared with respect to λ and μ. A crucial point lies in the fact that the new space is restricted into a sphere and then some operations are possible like computing the volume of a function of resemblance.

3.1 Examples of existing measures of resemblance

Let us consider the following usual measures with their expression in cartesian space and in spherical space:

Jaccard similarity	$S_1(X,Y,Z) = \frac{X}{X+Y+Z}$ $\sigma_1(\lambda,\mu) = \frac{\cos\mu}{\cos\mu + \sin\mu + \cot\lambda}$
Dice similarity	$S_2(X,Y,Z) = \frac{2X}{2X+Y+Z}$ $\sigma_2(\lambda,\mu) = \frac{2\cos\mu}{2\cos\mu + \sin\mu + \cot\lambda}$
Ochiai similarity	$S_3(X,Y,Z) = \frac{X}{\sqrt{(X+Y)(X+Z)}}$ $\sigma_3(\lambda,\mu) = \frac{\cos\mu}{\sqrt{(\cos\mu + \sin\mu)(\cos\mu + \cot\lambda)}}$
Fermi-Dirac Resemblance	$S_4(X,Y,Z) = \frac{F_{FD}(\phi) - F_{FD}(\frac{\pi}{2})}{F_{FD}(0) - F_{FD}(\frac{\pi}{2})}$

The last measure is not a usual one but a measure inspired by the Fermi-Dirac satisfiability [5]. This measure is based on the Fermi-Dirac function F_{FD} described by the following analytical form:

$$F_{FD}(\phi) = \frac{1}{1 + \exp^{\frac{(\phi-\phi_0)}{\Gamma}}}$$

where $\phi = \arctan(\frac{Y+Z}{X})$, Γ is a positive real and $\phi_0 \in [0, \frac{\pi}{2}]$.

3.2 Discrimination power of a measure of resemblance

A relevant way to observe the discrimination power of measures of resemblance is to study their severity. We consider that a measure σ_1 is more severe than a measure σ_2 for μ, λ if $\sigma_1(\mu,\lambda) < \sigma_2(\mu,\lambda)$. This intuitive definition is local. Thus, in order to compare two measures, we need to take into account all values. Thanks to our new representation, we can integrate the measure over the whole input space. The severity can be illustrated by the computation of the volume. In fact, a (non) severe measure gives (high) low values and therefore its volume is (large) small.

The total volume of each measure of resemblance is the following, by increasing order:

Resemblance	*Volume*
Fermi-Dirac, $\Gamma = 0.1$, $\phi_0 = \frac{\pi}{4}$	0.44
Jaccard	0.67
Dice	0.96
Ochiai	1.02

We observe that the Fermi-Dirac resemblance is globally the most severe whereas the Ochiai's resemblance is the least severe. The volume gives a global view of the severity. However, more refined studies can be done. In fact, if we focus on the situations where the objects to be compared are similar ($\mu \in [0, \frac{\pi}{4}], \lambda \in [\frac{\pi}{4}, \frac{\pi}{2}]$), the order of the local severity changes: Jaccard (volume=0.34), Fermi-Dirac (volume=0.36), Dice (volume=0.43), Ochiai (volume=0.43). This means that for rather similar objects, Jaccard's resemblance is the most severe eventhough globally it is not the case.

3.3 Graphical comparisons

In order to have a precise analyze of the behaviour of a measure of resemblance, we consider a reference measure which can be qualified as neutral with respect to the discrimination power. The reference function is the only measure of resemblance (i.e. satisfies the conditions of definition 2) that decreases linearly with respect to μ when λ is fixed, and with λ when μ is fixed. This function (see figure 2) is characterized by the following equation:

$$\sigma(\lambda, \mu) = \frac{(\pi - 2\mu) \cdot 2\lambda}{\pi^2}$$

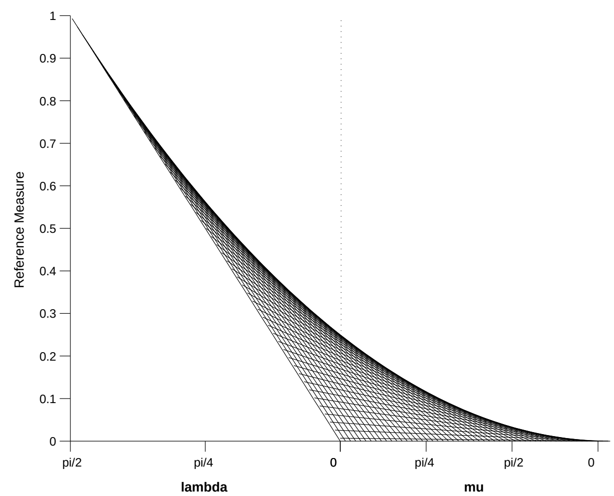

Fig. 2. The reference measure

We can now compare the other measures to this reference and focus on two aspects:

1. the relative position with respect to the reference.
2. the slope of the measure.

Severity In order to compare the severity, we look at the relative position of a measure of resemblance with respect to the reference measure. In fact, as we defined before, if the resemblance is above the reference, then the resemblance is less severe (white in Figure 3). Otherwise, it is more severe (black in Figure 3).

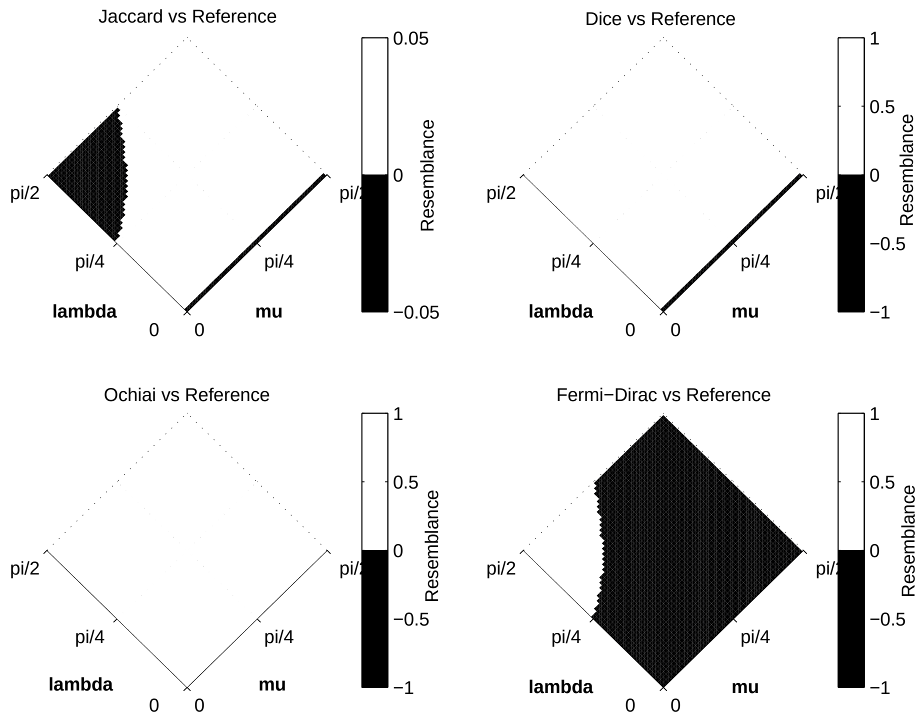

Fig. 3. Comparison of different measures of resemblance with respect to the reference measure

We notice that the Fermi-Dirac resemblance is the only measure that is more severe than the reference, almost everywhere. This explains that its volume is globally low. It is also the only one that is less severe than the reference for rather similar objects (μ close to 0 and λ to $\frac{\pi}{2}$). At the opposite, Jaccard resemblance, for the same kinds of objects, is the only one to be more severe than the reference. This symmetrical attitude explains the inversion of the order based on the volume over the square $\mu \in [0, \frac{\pi}{4}], \lambda \in [\frac{\pi}{4}, \frac{\pi}{2}]$.

Ochiai and Dice similarities have a non severe behaviour if anything, since they are always above the reference measure. And if we compare them (see right side of Figure 4), we see that they are very close to each other even though Dice similarity is always a bit more severe than Ochiai similarity. In return, on the left side of the same figure, we notice that Dice similarity is always less severe than the Jaccard similarity. We obtain like this a severity ranking: Jaccard, Dice, Ochiai.

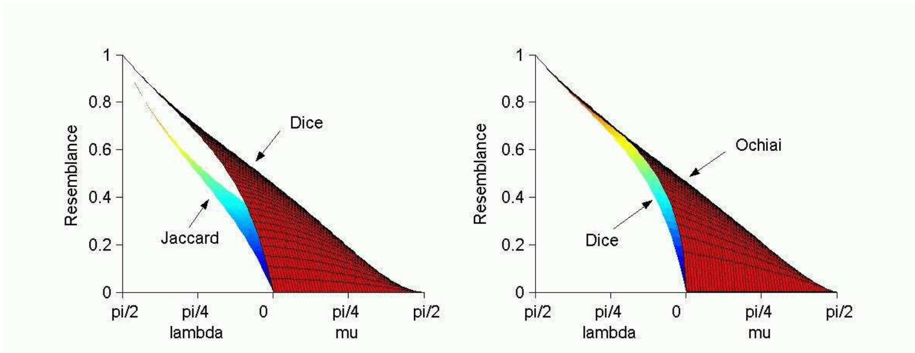

Fig. 4. Comparison of Ochiai, Dice and Jaccard measures of resemblance

Discrimination The relative position is not the only aspect to be considered for study of the behaviour of a measure. Indeed, the slope provides an idea of the variation of the discrimination power. Note that the reference measure has a linear variation (parallel to μ and to λ axes). Thus, we consider that if the decrease of a resemblance is bigger than the decrease of the reference measure, then the resemblance can be qualified as more discriminant. Otherwise, it is less discriminant.

The Fermi-Dirac resemblance is particularly interesting because the discrimination power of the measure can be controlled. For different small deviations between compared objects, the Fermi-Dirac resemblance makes little difference until a certain point, from where the resemblance decreases quickly. The discrimination at this point is very high. Finally, for large deviations, again the Fermi-Dirac resemblance makes little difference.

In this context, the discrimination power can be controlled by two parameters: ϕ_0 controls the point where the decrease will occur whereas Γ controls the decrease speed of the measure. In fact, when Γ is high, the resemblance varies in a uniform way (linearly), which means that every deviation between the compared objects is penalized in an equal manner. The right side of Figure 5 shows the behaviour of the Fermi-Dirac resemblance when Γ varies.

We observe two limit cases:

- when Γ is close to 0, the function tends to the binary function. This measure discriminates the comparisons in just two situations. The first one when the objects to be compared are rather similar, the resemblance equals 1 and the second one when the objects are rather different, the resemblance equals 0.
- when Γ tends to infinity, the function tends to the reference measure.

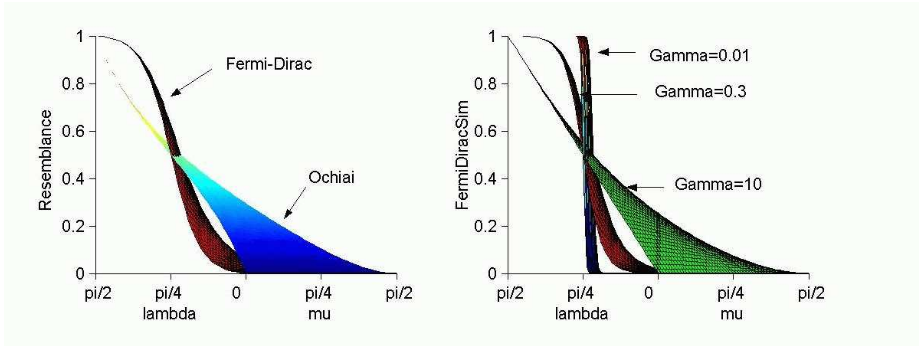

Fig. 5. The Fermi-Dirac measure of resemblance

4 Conclusion

We proposed a new representation (in spherical coordinates) that allows us to compare different measures of resemblance on all the input space, which is not possible otherwise. Thanks to this framework, we also identified a new measure of resemblance (Fermi-Dirac resemblance), which has not only a complementary behaviour with respect to the known measures but also provides an interesting parametrization, which enables us to control the discrimination power.

References

1. B. Bouchon-Meunier, M. Rifqi, and S. Bothorel. Towards general measures of comparison of objects. *Fuzzy Sets and Systems*, 84(2):143–153, 1996.
2. S.-M. Chen, M.-S. Yeh, and P.-Y. Hsiao. A comparison of similarity measures of fuzzy values. *Fuzzy Sets and Systems*, 72:79–89, 1995.
3. D. Dubois and H. Prade. A unifying view of comparison indices in a fuzzy set-theoretic framework. In R. R. Yager, editor, *Fuzzy set and possibility theory*, pages 3–13. Pergamon Press, 1982.
4. C. P. Pappis and N. Karacapilidis. A comparative assessment of measures of similarity of fuzzy values. *Fuzzy Sets and Systems*, 56:171–174, 1993.
5. M. Rifqi, V. Berger, and B. Bouchon-Meunier. Discrimination power of measures of comparison. *Fuzzy Sets and Systems*, 110(2):189–196, March 2000.
6. E. Trillas and L. Valverde. On implication and indistinguishability in the setting of fuzzy logic. In J. Kacprzyk and R. R. Yager, editors, *Management Decision Support Systems Using Fuzzy Sets and Possibility Theory*. Verlag TUV, Rheinland, 1984.
7. A. Tversky. Features of similarity. *Psychological Review*, 84:327–352, 1977.
8. L. Valverde. On the structure of t-indistinguishability operators. *Fuzzy Sets and Systems*, 17:313–328, 1985.
9. X. Wang, B. De Baets, and E. Kerre. A comparative study of similarity measures. *Fuzzy Sets and Systems*, 73:259–268, 1995.
10. R. Zwick, E. Carlstein, and D. V. Budescu. Measures of similarity among fuzzy concepts : a comparative analysis. *International Journal of Approximate Reasoning*, 1(2):221–242, 1987.