Similarity Relations Based on Distances as Fuzzy Concepts

Manuel Laguía
Dept. Leng. y Sistemas Informáticos
E.S. Ingeniería de Cádiz
Univ. de Cádiz, 11003 Cádiz, Spain
manuel.laguia@uca.es

Juan Luis Castro
Dept. CC. Computación e Inteligencia Artificial
E.T.S. Ingeniería Informática
Univ. de Granada, 18071 Granada, Spain
castro@decsai.ugr.es

Abstract

An usual and effective way to define a similarity relation is from a function of distance. Nevertheless, if we consider only usual distances, then some natural kinds of similarity relations cannot be obtained. In this paper, we analyze some of these kinds of similarity relations, and we find non–usual distances in order to obtain these ones.

Keywords: Similarity, Distance, Pseudo–metrics, Fuzzy Distance, Machine Learning, Classification Tasks, Case–Based Reasoning.

1 Introduction. The notion of similarity

It is often difficult to assign a degree of similarity $sim(x,y) : D \times D \to [0,1]$ between two objects x and y represented by two n-valued vectors $(x_1,x_2,\ldots,x_n)$ and $(y_1,y_2,\ldots,y_n)$, but it is easier to define partial measurements of similarity for each attribute $sim_i(x_i,y_i)$. How can we combine that information to obtain one single value that represents the whole similarity between x and y?

One usual and effective way to overcome these difficulties is by using distances, because we are familiar with them and, from a practical point of view, it is easy to transform one distance measurement in one similarity measurement.

Much work has been done about similarity ([2] [4] [5]), and about distances and methods based on distances (for example [3] [6]). It is interesting the soft or fuzzy sense of this kind of functions due to their gradual outcome (between 0 and 1 for similarity, and 0 and $+\infty$ for distance) and their close relation with fuzzy sets [1].

If we consider only usual distances, then some natural kinds of similarity relations cannot be obtained. For example, in classification problems we know the value of some characteristics and a class about some instances and our goal is finding the correct class for new instances. If we know the incomes and expenses of some companies and the class is the profits of the company, the examples belonging to the same class (and in that sense, similar) are along a line (see fig. 1). This is one example of a wide number of problems where the examples are, roughly speaking, grouped into bands, and usual distances fail.

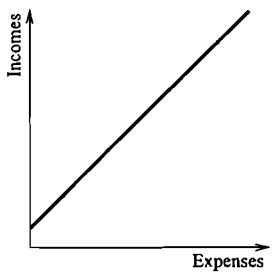

Figure 1: Companies with equal profits.

2 Functions of Distance and Similarity

In general, we will use a function of distance $d(x,y) : D \times D \to \mathrm{R}$ according to the following pattern:

$$d(x,y) = Comb(w_1,w_2,\cdots,w_n,d_1(x_1,y_1), d_2(x_2,y_2),\cdots,d_n(x_n,y_n))$$

where $d_i : D_i \times D_i \to \mathrm{R}$ is a partial measurement of distance over the i^{th} domain, $w_i \in [0,1]$ is a

weight that shows the relative significance of D_i, and $Comb : \mathrm{R}^{2n} \rightarrow [0,1]$ is a function that combines all the information and provides one measurement of distance in D. This definition includes all the usual geometric metrics, like Euclidean, Manhattan, Chebychev, ...

Once we have a distance, we can obtain a measurement of similarity according to [4]:

$$sim(x,y) = 1 - \frac{d(x,y)}{max}$$

if $d(x,y) \in$ [0,max], or

$$sim(x,y) = 1 - \frac{d(x,y)}{1+d(x,y)}$$

if $d(x,y)$ is unbounded.

3 Non-Usual Distances

The usual distances are useful in a lot of situations, but often other kind of distances are more appropriated, as we previously showed in the example in Section 1 (see fig. 1).

To deal with that kind of problems we propose a function of distance that groups the points according to bands along an hyperplane H in R^n (a line in R^2) $d_{\alpha,wide}(x,y) : \mathrm{R}^n \times \mathrm{R}^n \rightarrow \mathrm{R}_0^+$ as

$$d_{\alpha,wide}(x,y) = wide \left| \sum_{i=1}^{n} \cos\alpha_i (x_i - y_i) \right|$$

where $\alpha = (\alpha_1, \cdots, \alpha_n) \in [0,2\pi]^n$ is the set of angles between the axis and the unitary vector that is perpendicular to the desired hyperplane H (fig. 2), $wide \in \mathrm{R}_0^+$ controls the width of the band of points at a given distance (lower values imply approaching points to H), and $|.|$ is the absolute value in R. Moreover, α verifies that $\sum_{i=1}^{n} \cos^2\alpha_i = 1$, so there is only n degrees of freedom in the parameters. Notice that $(x_i - y_i)$ can be lower, equal or greater than 0. We use R^n for simplicity, and for symbolic domains $(x_i - y_i)$ represents the partial distance between x_i and y_i in that domain.

This distance observes $d(x,y) \geq 0$, $d(x,y) = d(y,x)$, and $d(x,z) \leq d(x,y) + d(y,z)$ $\forall x,y,z \in D$, but $d(x,y) = 0 \not\Rightarrow x = y$ $\forall x,y \in D$. So it is not a mathematical metric, but it is a mathematical pseudo–metric. Moreover, the binary relation $xRy \Leftrightarrow d(x,y) = 0$ observes the reflexive, symmetrical and transitive properties: R is a relation of equivalence that divides the original set in classes. Examples of this kind of distance are showed in fig. 3.

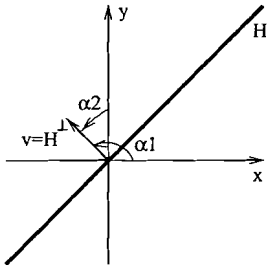

Figure 2: Definition of a band in R^2.

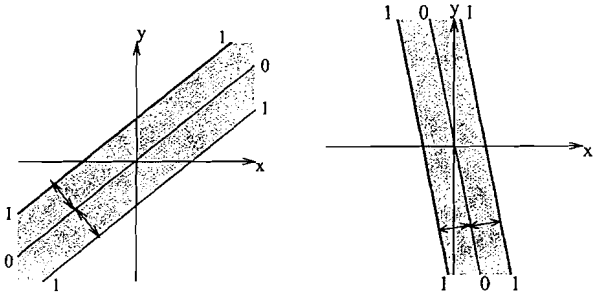

Figure 3: Examples of bands along hyperplanes in R^2.

Instead of restricting the weights α to $[0,2\pi]^n$ such as $\sum_{i=1}^{n} \cos^2\alpha_i = 1$, we also propose, given $w = (w_1, \cdots, w_n) \in \mathrm{R}^n$, the function of distance $d_w(x,y) : \mathrm{R}^n \times \mathrm{R}^n \rightarrow \mathrm{R}_0^+$ as

$$d_w(x,y) = \left| \sum_{i=1}^{n} w_i (x_i - y_i) \right|$$

Lemma 1 *For all $w \in \mathrm{R}^n$, there do exist some values $\alpha \in [0,2\pi]^n$ with $\sum_{i=1}^{n} \cos^2\alpha_i = 1$ and $wide \in \mathrm{R}_0^+$ (and vice versa), such that $d_w(x,y) = d_{\alpha,wide}(x,y)$ $\forall x,y \in \mathrm{R}^n$*

See appendix A for the demonstration. The meaning of this lemma is that given n real values, there exists a set of n angles whose cosines are proportional to those real values (and vice versa).

Given $w = (w_1, \cdots, w_n) \in \mathrm{R}^n$, the two following distances are also interesting

$$d_w^{\times}(x,y) = \sum_{i=1}^{n} w_i |x_i - y_i|$$

$$d_w^{|\times|}(x,y) = \left| \sum_{i=1}^{n} w_i |x_i - y_i| \right|$$

If $w_i > 0 \quad \forall i = 1,\ldots,n$ both distances are a weighted Chebychev distance, but if some $w_i \leq 0$ we obtain non-usual distances. For instance, if we consider $w_1 = -1$ and $w_2 = 1$ in R^2 we obtain the $d_w^{|\times|}(x,y)$ distance shown in fig. 4. On the left we can see the distance of some points from (0,0), and on the right the points with $d_w^{|\times|}(x,y) \in [0,1]$.

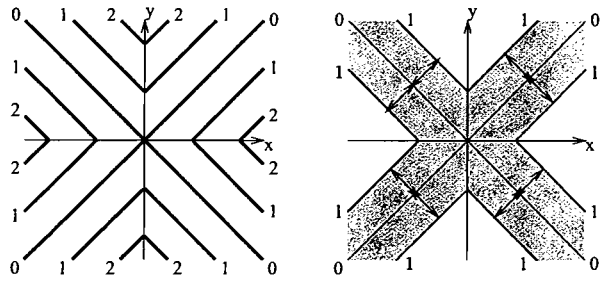

Figure 4: Example of distance $d_w^{|\times|}(x,y)$.

If we consider the same values $w_1 = -1$ and $w_2 = 1$ we obtain the $d_w^{\times}(x,y)$ distance shown in fig. 5.

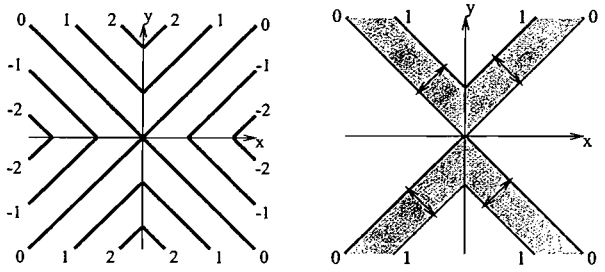

Figure 5: Example of distance $d_w^{\times}(x,y)$.

$d_w^{|\times|}(x,y)$ is restricted to zero or positive values, but $d_w^{\times}(x,y)$ may also return negative values. Here it is interesting to observe that $d_w^{\times}(x,y)$ can separate the space in two regions in a X–shaped style, according to the points with positive or negative values (fig. 6).

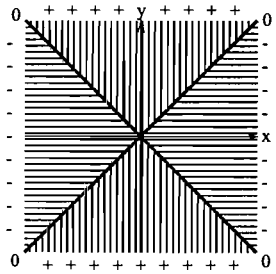

Figure 6: Positive and negative regions of $d_w^{\times}(x,y)$.

4 Conclusions and Future works

We have introduced some new functions of distance: $d_{\alpha,wide}(x,y)$, $d_w^{\times}(x,y)$ and $d_w^{|\times|}(x,y)$.

$d_{\alpha,wide}(x,y)$ assigns equal distance (and similarity) along hyperplanes in R^n. It is preferable in domains where similar instances are grouped into bands, for example when we have a set of marks or values and the assigned class is the weighted average (see fig. 7 for $\frac{1}{3}v_1 + \frac{2}{3}v_2$).

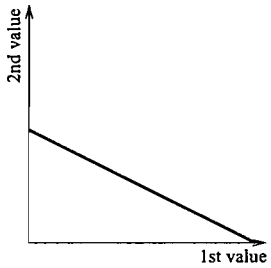

Figure 7: Points with equal weighted average.

$d_w^{\times}(x,y)$ and $d_w^{|\times|}(x,y)$ assign distance in an X-shaped style, for example, we can express the set of legal moves of a Bishop in chess, giving distance 0 to the squares where the bishop can go in the next movement.

We conclude that $d_{\alpha,wide}(x,y)$, $d_w^{\times}(x,y)$, $d_w^{|\times|}(x,y)$ are useful in domains where the information is grouped according to patterns that are different to the usual notion of proximity. In such domains, usual distances do not work properly, and other techniques like k–NN will choose the non–appropriate points. Much work must be done with this kind of distances, like estimation of the values for the parameters or definition of new non–usual distances.

Another interesting field is using functions of distance to establish similarity between fuzzy sets over D or D_i, or when the values of some attributes A_i are fuzzy sets.

A Demonstration of Lemma 1

Given $w \in R^n$.

If $w = (0,0,\ldots,0)$, we can choose $wide = 0$ ($\in R_0^+$), and any $\alpha \in [0,2\pi]^n$ with $\sum_{i=1}^n \cos^2\alpha_i = 1$ (for instance $\alpha_1 = 0$ and $\alpha_i = \frac{\pi}{2}$ $i = 2,\ldots,n$: $\alpha_1 = 0 \wedge \alpha_i = \frac{\pi}{2}$ $i = 2,\ldots,n \Rightarrow \cos\alpha_1 = 1 \wedge \cos\alpha_i = 0$ $i = 2,\ldots,n \Rightarrow \cos^2\alpha_1 = 1 \wedge \cos^2\alpha_i = 0$ $i = 2,\ldots,n$

$\Rightarrow \sum_{i=1}^{n} \cos^2 \alpha_i = 1$). Then

$d_w(x,y) = |\sum_{i=1}^{n} w_i(x_i - y_i)| = |\sum_{i=1}^{n} 0(x_i - y_i)| = |\sum_{i=1}^{n} 0| = 0 = 0|\sum_{i=1}^{n} \cos\alpha_i(x_i - y_i)| = d_{\alpha,wide}(x,y)$

If $w \neq (0,0,\cdots,0)$, we can choose $wide = \sqrt{\sum_{j=1}^{n} w_j^2}$ and α so that $cos(\alpha_i) = \frac{w_i}{\sqrt{\sum_{j=1}^{n} w_j^2}}$. Then

$\sqrt{\sum_{j=1}^{n} w_j^2} \geq 0 \Rightarrow wide \in \mathrm{R}_0^+$,

$|w_i| = \sqrt{w_i^2} \leq \sqrt{\sum_{j=1}^{n} w_j^2} \Rightarrow \frac{|w_i|}{\sqrt{\sum_{j=1}^{n} w_j^2}} \Leftrightarrow$
$-1 \leq \frac{w_i}{\sqrt{\sum_{j=1}^{n} w_j^2}} \leq 1 \Leftrightarrow \cos\alpha_i \in [-1,1] \Leftrightarrow$
$\alpha_i = \arccos \frac{w_i}{\sqrt{\sum_{j=1}^{n} w_j^2}} \in [0,2\pi]$,

$\sum_{i=1}^{n} \cos^2 \alpha_i = \sum_{i=1}^{n} \frac{w_i^2}{\sum_{j=1}^{n} w_j^2} = \frac{\sum_{i=1}^{n} w_i^2}{\sum_{j=1}^{n} w_j^2} = 1$

so *wide* and α verify their restrictions,

$d_{\alpha,wide}(x,y) = wide|\sum_{i=1}^{n} \cos\alpha_i(x_i - y_i)|$
$= \sqrt{\sum_{j=1}^{n} w_j^2}\left|\sum_{i=1}^{n} \frac{w_i}{\sqrt{\sum_{j=1}^{n} w_j^2}}(x_i - y_i)\right| =$
$\sqrt{\sum_{j=1}^{n} w_j^2}\left|\frac{1}{\sqrt{\sum_{j=1}^{n} w_j^2}}\sum_{i=1}^{n} w_i(x_i - y_i)\right| \overset{\sqrt{\sum_{j=1}^{n} w_j^2} \geq 0}{=}$
$\left|\frac{\sqrt{\sum_{j=1}^{n} w_j^2}}{\sqrt{\sum_{j=1}^{n} w_j^2}}\sum_{i=1}^{n} w_i(x_i - y_i)\right| = |\sum_{i=1}^{n} w_i(x_i - y_i)| = d_w(x,y)$

so, with those values of *wide* and α_i $d_w(x,y) = d_{\alpha,wide}(x,y) \quad \forall x,y \in \mathrm{R}^n$.

Given $wide \in \mathrm{R}_0^+$ and $\alpha \in [0,2\pi]^n$ such as $\sum_{i=1}^{n} \cos^2 \alpha_i = 1$ we choose $w_i = wide \cos\alpha_i$ ($\in$ R). Then

$d_w(x,y) = |\sum_{i=1}^{n} w_i(x_i - y_i)|$
$= |\sum_{i=1}^{n} wide \cos\alpha_i(x_i - y_i)| =$
$|wide \sum_{i=1}^{n} \cos\alpha_i(x_i - y_i)| \overset{wide \in \mathrm{R}_0^+}{=}$
$wide|\sum_{i=1}^{n} \cos\alpha_i(x_i - y_i)| = d_{\alpha,wide}(x,y)$

so, with those values of w_i $d_w(x,y) = d_{\alpha,wide}(x,y) \quad \forall x,y \in \mathrm{R}^n$.

Acknowledgements

The authors gratefully acknowledge the support of the Spanish *Dirección General de Enseñanza Superior e InvestigaciÓn Científica*, under grant PB98-1379-C02-01.

References

[1] Baets, B., Mesiar, R. (1997) *Pseudo-metrics and T-equivalences*. Journal Fuzzy Mathematics Vol. 5, No. 2, pp. 471–481.

[2] Plaza, E., López, R., Armengol, E. (1996) *On the Importance of Similitude: An Entropy-Based Assessment*. Proceedings Third European Workshop on Case-Based Reasoning (EWCBR-96). Springer-Verlag.

[3] Ricci, F., Avesani, P. (1995) *Learning a Local Similarity Metric for Case-Based Reasoning*. Proceedings of the First International Conference on Case-Based Reasoning, Sesimbra, Portugal (ICCBR-95). Springer-Verlag.

[4] Ritcher, M.M. (1992) *Classification and Learning of Similarity Measures*. Proceedings der 16. Jahrestagung der Gesellschaft für Klassifikation (GFKL-92). V. Springer Verlag

[5] Ritcher, M.M. (1995) *On the Notion of Similarity in Case-Based Reasoning*. Mathematical and Statistical Methods in Artificial Intelligence, Springer Verlag 1995, pp. 171–184. Eds. G. del Viertl.

[6] Wettschereck, D. (1994) *A Study of Distance–Based Machine Learning Algorithms*. PhD. thesis, Oregon State University.