Approximating Proximities by Similarities

L. Garmendia
Ingeniería del Software e Inteligencia Artificial
Facultad de Informática
Universidad Complutense de Madrid
C/ Profesor José García Santesmases, s/n
28040 Madrid
Spain
lgarmend@fdi.ucm.es

J. Recasens
Secció Matemàtiques i Informàtica
ETS Arquitectura del Vallès
Universitat Politècnica de Catalunya
Pere Serra 1-15
08190 Sant Cugat del Vallès
Spain
j.recasens@upc.edu

Abstract

In this paper an algorithm to find a similarity close to a proximity or tolerance relation R is given. The obtained similarity is closer to R than its transitive closure or any transitive opening of R.

Keywords: Proximity, Tolerance relation, Similarity, Transitive closure, Transitive opening.

1 Introduction

T-indistinguishability operators with respect to the Minimum t-norm (Min-indistinguishability operators) are usually called similarity relations [6]. They are interesting fuzzy relations with many applications on Classification and Cluster analysis, because the α-cuts are partitions of the universe and when α increases the partitions are refinements of the previous ones. In other words, similarity relations generate indexed hierarchical trees and vice versa, from an indexed hierarchical tree a similarity relation can be built.

In many situations, the data are given as a proximity relation on the universe of discourse. This is a reflexive and symmetric fuzzy relation. If we need to classify our objects in a hierarchical tree, this relation must be replaced by a similarity relation as close to it as possible.

The most popular ways to find such similarity is calculating its transitive closure, which is greater or equal than R, or finding a transitive opening, which is a maximal similarity among the ones that are smaller or equal than R.

This paper presents a simple algorithm to find a similarity relation close to a given proximity relation which is closer than its transitive closure or any of its transitive openings. In contrast to the transitive closure or transitive openings, in general some entries of the obtained similarity are greater and some smaller than the corresponding entries of the proximity relation.

In [2], a couple of methods to calculate good approximations of a proximity relation by a T-indistinguishability operator when T is a continuous Archimedean t-norm were provided. Those methods do not work for the Minimum. With the results of the current work we can then obtain good approximations of proximity relations for the most popular t-norms, namely Łukasiewicz, Product and the Minimum.

2 Preliminaries

This Section contains some results on indistinguishability operators that will be needed later on in the paper.

Definition 2.1. [6], [1] *A similarity relation E on a set X is a fuzzy relation $E : X \times X \to [0,1]$ satisfying for all $x, y, z \in X$*

1. $E(x,x) = 1$ *(Reflexivity)*
2. $E(x,y) = E(y,x)$ *(Symmetry)*
3. $\mathrm{Min}(E(x,y), E(y,z)) \leq E(x,z)$ *(*Min-

L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 1332–1335
Torremolinos (Málaga), June 22–27, 2008

transitivity).

Definition 2.2. *The natural similarity relation E_{Min} is the fuzzy relation on* [0,1] *defined by* $E_{\mathrm{Min}}(x,y) = \begin{cases} Min(x,y) & \text{if } x \neq y \\ 1 & \text{otherwise.} \end{cases}$

E_{Min} is indeed a special kind of similarity relation and in a logical context where Min plays the role of the conjunction, E_{Min} is interpreted as the bi-implication associated to the Minimum t-norm [3].

Theorem 2.3. *Representation Theorem* [5]. *Let R be a fuzzy relation on a set X. R is a similarity relation if and only if there exists a family $(h_i)_{i \in I}$ of fuzzy subsets of X such that for all $x, y \in X$*

$$R(x,y) = inf_{i \in I} E_{\mathrm{Min}}(h_i(x), h_i(y)).$$

$(h_i)_{i \in I}$ is called a generating family of R.

In particular, given a proximity matrix or relation R on X (i.e. a reflexive and symmetric fuzzy relation), we can build the similarity relation $\underline{R}$ generated by the set of the columns of R (i.e. the fuzzy subsets $R(x,\cdot)$, $x \in X$).

Proposition 2.4. $\underline{R} \leq R$.

Definition 2.5. *Let R be a proximity matrix or relation on X. The* Min*-transitive closure $\overline{R}$ of R is the smallest similarity operator on X satisfying $R \leq \overline{R}$.*

Definition 2.6. *Let R and S be two fuzzy relations on X. The Sup-*Min *product of R and S is the fuzzy relation $R \circ S$ on X defined for all $x, y \in X$ by*

$$(R \circ S)(x,y) = sup_{z \in X} \mathrm{Min}(R(x,z), S(z,y)).$$

Since the Sup-Min product is associative, we can define for $n \in N$ the n^{th} power R^n of a fuzzy relation R:

$$R^n = \overbrace{R \circ \ldots \circ R}^{n \ times}.$$

Definition 2.7. *Let R be a fuzzy relation on a set X. The* Min*-transitive closure of R is the fuzzy relation*

$$\overline{R} = sup_{n \in N} R^n.$$

Proposition 2.8. *Let R be a proximity relation on a finite set X of cardinality n. Then*

$$\overline{R} = sup_{s \in \{1,\ldots,n-1\}} R^s.$$

Definition 2.9. *Let R be a proximity relation on a universe X. A similarity relation E on X is a* Min*-transitive opening of R if and only if $E \leq R$ and E is maximal among all similarity rerlations smaller or equal than R.*

3 The algorithm

In this section we give an algorithm that generates a similarity E close to a given proximity relation R. The idea is to start with the transitive closure or a similarity smaller than R and shift their entries to obtain a better approximation of R.

Proposition 3.1. *[4] Let E be a similarity relation on a finite universe X of cardinality n. The number of different entries of the matrix representing E is smaller or equal than n.*

The algorithm to approximate proximities by similarity relations is based on the following result.

Proposition 3.2. *Let E be a similarity relation on a finite universe X of cardinality n and $a_1 < a_2 < \ldots < a_l = 1$ $(l \leq n)$ the entries of E. If we replace the entries by $a'_1 \leq a'_2 \leq \ldots \leq a'_l = 1$ respectively, we obtain a new similarity relation on X.*

Proof. Reflexivity and symmetry are trivial.

Transitivity: if $Min(a_i, a_j) \leq a_k$, then $Min(a'_i, a'_j) \leq a'_k$. □

Example 3.3.

$$E = \begin{pmatrix} 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 1 & 0.3 & 0.3 \\ 0.2 & 0.3 & 1 & 0.4 \\ 0.2 & 0.3 & 0.4 & 1 \end{pmatrix}$$

is a similarity relation. If we replace 0.2, 0.3, 0.4, 1 by 0.1, 0.5, 0.8, 1 respectively we obtain the similarity relation

$$E' = \begin{pmatrix} 1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 1 & 0.5 & 0.5 \\ 0.1 & 0.5 & 1 & 0.8 \\ 0.1 & 0.5 & 0.8 & 1 \end{pmatrix}$$

The idea for finding a similarity relation close to a given proximity R is then very easy. We can calculate the Min-transitive closure $\overline{R}$ of R and then we can modify the entries of $\overline{R}$ in order to minimize some distance to R or to maximize some similarity measure to R. Of course, we can calculate a transitive opening or the similarity relation obtained from R by the Representation Theorem instead of the transitive closure.

Nevertheless, the procedure is not straightforward as is shown in the following two examples.

Example 3.4. *Let us consider the proximity with matrix*

$$R = \begin{pmatrix} 1 & 0.7 & 0.3 & 1 \\ 0.7 & 1 & 0.4 & 0.7 \\ 0.3 & 0.4 & 1 & 0.8 \\ 1 & 0.7 & 0.8 & 1 \end{pmatrix}$$

Its Min-transitive closure is

$$\overline{R} = \begin{pmatrix} 1 & 0.7 & 0.8 & 1 \\ 0.7 & 1 & 0.7 & 0.7 \\ 0.8 & 0.7 & 1 & 0.8 \\ 1 & 0.7 & 0.8 & 1 \end{pmatrix}$$

If we replace the entries with values 0.7 and 0.8 of $\overline{R}$ by a and b respectively in order to minimize the Euclidean distance d between R and the new matrix, we must minimize

$$\begin{aligned} f(a,b) &= (a-0.7)^2 + (a-0.4)^2 + (a-0.7)^2 \\ &\quad +(b-0.8)^2 + (b-0.3)^2. \end{aligned}$$

$$\frac{\partial f}{\partial a} = 2(a-0.7) + 2(a-0.4) + 2(a-0.7) = 0$$

$$\frac{\partial f}{\partial b} = 2(b-0.8) + 2(b-0.3) = 0$$

and

$$a = \frac{0.7+0.7+0.4}{3} = 0.6$$

$$b = \frac{0.8+0.3}{2} = 0.55$$

obtaining

$$\begin{pmatrix} 1 & 0.6 & 0.55 & 1 \\ 0.6 & 1 & 0.6 & 0.6 \\ 0.55 & 0.6 & 1 & 0.55 \\ 1 & 0.6 & 0.55 & 1 \end{pmatrix}$$

which is not Min-transitive and therefore not a similarity relation.

The same problem may occur with approximations from bellow as shown in the next example.

Example 3.5. *From the proximity matrix*

$$R = \begin{pmatrix} 1 & 0.2 & 0.3 & 0.4 \\ 0.2 & 1 & 0.7 & 0.8 \\ 0.3 & 0.7 & 1 & 0.5 \\ 0.4 & 0.8 & 0.5 & 1 \end{pmatrix}$$

we obtain the matrix $\underline{R}$ with the Representation Theorem.

$$\underline{R} = \begin{pmatrix} 1 & 0.2 & 0.2 & 0.2 \\ 0.2 & 1 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1 & 0.3 \\ 0.2 & 0.2 & 0.3 & 1 \end{pmatrix}$$

If we replace the entries with values 0.2 and 0.3 of $\underline{R}$ by a and b respectively in order to minimize the Euclidean distance d between R and the new matrix, we must minimize

$$\begin{aligned} f(a,b) &= (a-0.2)^2 + (a-0.3)^2 + (a-0.4)^2 \\ &+ (a-0.7)^2 + (a-0.8)^2 + (b-0.3)^2. \end{aligned}$$

$$\begin{aligned} \frac{\partial f}{\partial a} &= 2(a-0.2) + 2(a-0.3) + 2(a-0.4) \\ &\quad +2(a-0.7) + 2(a-0.8) = 0 \\ \frac{\partial f}{\partial b} &= 2(b-0.3) = 0 \end{aligned}$$

and

$$a = \frac{0.2+0.3+0.4+0.7+0.8}{5} = 0.48$$

$$b = 0.3$$

The matrix

$$\begin{pmatrix} 1 & 0.48 & 0.48 & 0.48 \\ 0.48 & 1 & 0.48 & 0.48 \\ 0.48 & 0.48 & 1 & 0.3 \\ 0.48 & 0.48 & 0.3 & 1 \end{pmatrix}$$

is not Min-transitive and therefore not a similarity relation.

A possible solution in these cases is replacing the "wrongly ordered" entries by a unified value. For instance, in Example 3.4 the entries 0.6 and 0.55 can be replaced by $\frac{3\cdot 0.6+2\cdot 0.55}{5} = 0.58$ and in Example 3.5 0.48

and 0.3 by $\frac{5 \cdot 0.48+1 \cdot 0.3}{6} = 0.45$. Note that, thanks to the next lemma, among all possible values 0.58 and 0.45 are the ones who minimize the distance between the obtained matrix and R.

Lemma 3.6. *Let $P = (x_1, x_2, ..., x_n) \in R^n$ The closest point Q to P with respect to the Euclidean distance of the form $(a, a, ..., a)$ satisfies $a = \frac{x_1+x_2+...+x_n}{n}$.*

The algorithm to find a close similarity to a given proximity relation R on a universe of cardinality n is then as follows

Algorithm 3.7.

1. *Calculate the Min-transitive closure $\overline{R}$ or a lower approximation of R.*
2. *Order the entries $a_1 < a_2 < ... < a_k = 1$ $(k \leq n)$ of $\overline{R}$.*
3. *Replace every a_i by the arithmetic mean a'_i of the entries of R that are in the same place than a_i.*
4. *If $a'_1 \leq a'_2 \leq ... \leq a'_k = 1$, then the desired similarity relation E is obtained by replacing the entries $a_1, a_2, ..., a_k$ of $\overline{R}$ by $a'_1 \leq a'_2 \leq ... \leq a'_k$ respectively.*
5. *Else, For every maximal chain $C = \{a'_i, a'_{i+1}, ..., a'_{i+j}\}$ with $a'_i > a'_{i+j}$, replace all the elements of C by the weighted mean a_C of them, weighting every a'_l of C by the number of entries of R that correspond to a'_l. Replacing the elements of C by a_C in E the desired similarity relation is obtained.*

4 Concluding Remarks

An easy algorithm to calculate a good approximation of a given proximity relation P by a similarity has been introduced. The resulting similarity fits better to P that its Min-transitive closure or any of its Min-transitive openings. In particular, the entries of the Min-transitive closure are greater or equal than the corresponding entries of P, while the entries of a Min-transitive opening are smaller or equal than the corresponding entries of P. In general, some entries of the obtained approximation using the preceding algorithm are greater and some of them smaller than the corresponding entries of P.

The simplicity of the algorithm makes it suitable for real applications where Min-transitivity is needed.

In [2], some methods to calculate approximations of a proximity relation by a T-indistinguishability operator when T is a continuous Archimedean t-norm were given. With the results of the current work we can then obtain good approximations of proximity relations for the most popular t-norms, namely Łukasiewicz, Product and the Minimum.

Acknowledgements

Research partially supported by DGICYT projects number TIN2006-14311 and TIN2006-06190.

References

[1] D. Boixader, J. Jacas, J. Recasens (2000). Fuzzy Equivalence Relations: Advanced Material. In Dubois, Prade Eds. *Fundamentals of Fuzzy Sets*, Kluwer, 261-290.

[2] L. Garmendia, J. Recasens (2007). Finding close T-indistinguishability Operators to a given Proximity. Proceedings EUSFLAT2007. Ostrava. (2007). Vol 1, 127-133.

[3] P. Hájek (1998) *Metamathematics of Fuzzy Logic*. Kluwer. Dordrecht.

[4] J. Jacas (1990). Similarity relations - the calculation of minimal generating families. *Fuzzy Sets and Systems* 35, 151–162.

[5] L.Valverde (1985). On the structure of F-indistinguishability operators, *Fuzzy Sets and Systems* 17, 313-328.

[6] L.A.Zadeh (1971). Similarity relations and fuzzy orderings, *Information Science* 3, 177-200.