Aproximación de Proximidades por Similitudes

L. Garmendia
Ingeniería del Software e Inteligencia Artificial
Facultad de Informática
Universidad Complutense de Madrid
C/ Profesor José García Santesmases, s/n
28040 Madrid
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
j.recasens@upc.edu

Resumen

En este trabajo se presenta un algoritmo para hallar una relación de similitud próxima a una relación de proximidad (i.e.: una relación de tolerancia) R dada. La relación de similitud obtenida es más próxima a R que su clausura transitiva o cualquier opening transitivo de R.

Palabras Clave: Relación de Proximidad, Relación de Tolerancia, Relación de Similitud, Clausura Transitiva, Opening Transitivo.

1 Introducción

Las relaciones de similitud son operadores de T-indistinguibilidad respecto a la t-norma Mínimo [9]. Son relaciones borrosas interesantes con diversas aplicaciones en Técnicas de Clasificación y Cluster analysis porque sus α-cortes son particiones del universo de discurso y al aumentar α las correspondientes particiones refinan a las anteriores. Dicho de otro modo, las relaciones de similitud generan árboles jerárquicos indexados. Recíprocamente, a partir de un árbol jerárquico indexado se puede generar una relación de similitud.

En muchas situaciones, los datos de un problema se dan en forma de una relación de proximidad en el universo de discurso. Se parte entonces de una relación reflexiva y simétrica. Si se necesita clasificar los objetos del universo mediante un árbol jerárquico, esta relación debe ser reemplazada por una relación de similitud lo más cercana a la original como sea posible.

El método más común de hallar una relación de similitud cercana a una relación de proximidad R dada es calcular su clausura transitiva, que es una relación mayor o igual que R, o buscar un opening transitivo de R, es decir, una relación de similitud maximal entre las que son menores o iguales que R.

Este trabajo presenta un algoritmo de simple ejecución para hallar una relación de similitud a partir de una relación de proximidad R dada que es más cercana a R que su clausura transitiva o cualquiera de sus openings transitivos. En contraste con la clausura transitiva y los openings transitivos, en general algunas entradas de la similitud obtenida son mayores y otras menores que las correspondientes entradas de la relación de proximidad.

En [3] se presentaron un par de métodos parra hallar buenas aproximaciones de una relación de proximidad por operadores de T-indistinguibilidad cuando T es una t-norma arquimediana continua. Estos métodos no funcionan para la t-norma Mínimo (i.e.: relaciones de similitud). Con los resultados de este trabajo podemos obtener buenas aproximaciones para las t-normas más usuales: Łukasiewicz, Producto y Mínimo.

2 Preliminares

Esta sección presenta algunos resultados sobre operadores de T-indistinguibilidad que serán necesarios a lo largo del trabajo.

Definición 2.1 *La residuación $\overrightarrow{T}$ de una t-norma T se define para todo $x, y \in [0, 1]$ mediante*

$$\overrightarrow{T}(x|y) = sup\{\alpha \in [0, 1] \mid T(x, \alpha) \leq y).$$

Definición 2.2 *El operador de T-indistinguibilidad natural E_T asociado a una t-norma T dada es la relación borrosa en* [0,1] *definida para todo $x, y \in [0, 1]$ mediante*

$$E_T(x, y) = T(\overrightarrow{T}(x|y), \overrightarrow{T}(y|x)).$$

E_T es de hecho un tipo especial de operador de T-indistinguibilidad (Definición 2.3) [1] y en el contexto de la Lógica Borrosa, donde T representa la conjunción, E_T se interpreta como la biimplicación asociada a T [4].

Definición 2.3 *Dada una t-norma T, un operador de T-indistinguibilidad E en un conjunto X es una relación borrosa $E : X \times X \to [0,1]$ que para todo $x, y, z \in X$ satisface*

1. $E(x,x) = 1$ *(Reflexividad)*
2. $E(x,y) = E(y,x)$ *(Simetría)*
3. $T(E(x,y), E(y,z)) \leq E(x,z)$ *(T-transitividad).*

Ejemplo 2.4

1. *Si T es la t-norma de Łukasiewicz, entonces $E_T(x,y) = 1 - |x-y|$ para todo $x, y \in [0,1]$.*
2. *Si T es la t-norma Producto, entonces $E_T(x,y) = Min(\frac{x}{y}, \frac{y}{x})$ para todo $x, y \in [0,1]$ donde $\frac{z}{0} = 1$.*
3. *Si T es la t-norma Mínimo, entonces $E_T(x,y) = \begin{cases} Min(x,y) & \text{si } x \neq y \\ 1 & \text{en otro caso.} \end{cases}$*

Teorema 2.5 *Teorema de Representación* [8]. *Sea R una relación borrosa en un conjunto X y T una t-norma continua. R es una relación de T-indistinguibilidad si, y sólo si, existe una familia $(h_i)_{i\in I}$ de subconjuntos borrosos de X tal que para todo $x, y \in X$*

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

$(h_i)_{i\in I}$ se llama una familia de generadores de R.

En particular, dada una relación de proximidad R en X (i.e. una relación borrosa reflexiva y simétrica), podemos construir el operador de T-indistinguibilidad $\underline{R}$ generado por el conjunto de las columnas de R (i.e. los subconjuntos borrosos $R(x,\cdot)$, $x \in X$).

Proposición 2.6 $\underline{R} \leq R$.

Definición 2.7 *Sea R una relación o matriz de proximidad en X y T una t-norma continua. La clausura T-transitiva $\overline{R}$ de R es el menor operador de T-indistinguibilidad en X que cumple $R \leq \overline{R}$.*

Definición 2.8 *Sean R y S dos relaciones borrosas de X y T una t-norma continua. El producto Sup-T de R y S es la relación borrosa $R \circ S$ on X definida para todo $x, y \in X$ mediante*

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

Gracias a que el producto Sup-T es asociativo para t-normas continuas, podemos definir la potencia n-ésima R^n_T de una relación borrosa R para cualquier $n \in N$:

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

Definición 2.9 *Sea R una relación borrosa en un conjunto X y T una t-norma continua. La clausura T-transitiva de R es la relación borrosa*

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

Proposición 2.10 *Sea R una relación de proximidad en un conjunto finito X de cardinal n. Entonces*

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

Definición 2.11 *Sea R una relación de proximidad en un universo X y T una t-norma. Un operador de T-indistinguibilidad E de X es un opening T-transitivo de R si, y sólo si, $E \leq R$ y E es maximal en el conjunto de operadores de T-indistinguibilidad menores o iguales que R.*

3 El algoritmo

En esta sección se dará un algoritmo que genera una relación de similitud E cercana a una relación de proximidad R dada. La idea es partir de su clausura transitiva o de una relación de similitud menor o igual que R y modificar sus valores para hallar una mejor aproximación de R.

Proposición 3.1 [5] *Sea E una relación de similitud en un universo finito X de cardinal n. El número de entradas diferentes de la matriz que representa a E es menor o igual que n.*

El algoritmo para aproximar relaciones de proximidad mediante relaciones de similitud se basa en el siguiente resultado.

Proposición 3.2 *Sea E una relación de similitud en un universo finito X de cardinal n y $a_1 < a_2 < \ldots < a_k = 1$ ($k \leq n$) las entradas de E. Si reemplazamos las entradas por $a'_1 \leq a'_2 \leq \ldots \leq a'_k = 1$ respectivamente, se obtiene una nueva relación de similitud en X.*

Demostración

La reflexividad y la simetría de la nueva matriz son triviales.

Transitividad: Si $Min(a_i, a_j) \leq a_k$, entonces $Min(a'_i, a'_j) \leq a'_k$.

Ejemplo 3.3

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

es una relación de similitud. Si reemplazamos 0.2, 0.3, 0.4, 1 por 0.1, 0.5, 0.8, 1 respectivamente, obtenemos la nueva relación de similitud

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

La idea que se sigue para hallar una relación de similitud cercana a una relación de proximidad R dada es sencilla. Podemos calcular la clausura *Min*-transitiva $\overline{R}$ de R y luego modificar las entradas de $\overline{R}$ para minimizar alguna distancia dada a R o para maximizar alguna medida de similitud a R. Sin duda, también podemos calcular primero un opening transitivo o la relación de similitud obtenida de R mediante el Teorema de Representación en vez de la clausura *Min*-transitiva.

Sin embargo, el método no es tan directo tal y como muestran los siguientes dos ejemplos.

Ejemplo 3.4 *Consideremos la relación de proximidad con matriz*

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

Su clausura Min-transitiva es

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

Si reemplazamos las entradas con valores 0.7 y 0.8 de $\overline{R}$ por a y b respectivamente, para minimizar la distancia euclídea d entre R y la nueva matriz debemos minimizar

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

$$\frac{\partial f}{\partial a} = 2(a-0.7) + 2(a-0.4) + 2(a-0.7) = 0$$
$$\frac{\partial f}{\partial b} = 2(b-0.8) + 2(b-0.3) = 0$$

y

$$a = \frac{0.7+0.7+0.4}{3} = 0.6$$
$$b = \frac{0.8+0.3}{2} = 0.55$$

obteniendo

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

Esta matriz no es Min-transitiva y por tanto no es una relación de similitud.

El mismo problema puede presentarse con aproximaciones por debajo tal y como muestra el siguiente ejemplo.

Ejemplo 3.5 *De la matriz de proximidad*

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

se obtiene la matriz $\underline{R}$ mediante el Teorema de Representación.

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

Si reemplazamos las entradas con valores 0.2 y 0.3 de $\underline{R}$ por a y b respectivamente, para minimizar la distancia euclídea d entre R y la nueva matriz debemos minimizar

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

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

y

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

La matriz

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

no es Min-transitiva y por tanto no es una relación de similitud.

Una posible solución en estos casos es reemplazar las entradas "'mal ordenadas"' por un único valor. Por ejemplo, en el Ejemplo 3.4 las entradas 0.6 y 0.55 se pueden reemplazar por $\frac{3\cdot 0.6+2\cdot 0.55}{5} = 0.58$ y en el Ejemplo 3.5 0.48 y 0.3 por $\frac{5\cdot 0.48+1\cdot 0.3}{6} = 0.45$. Nótese que gracias al siguiente lema estos nuevos valores son los que minimizan la distancia entre la matriz obtenida y R.

Lema 3.6 *Sea $P = (x_1, x_2, ..., x_n) \in R^n$. El punto Q más próximo a P respecto a la distancia euclídea de la forma $(a, a, ..., a)$ cumple $a = \frac{x_1+x_2+...+x_n}{n}$.*

El algoritmo que proponemos para hallar una relación de similitud cercana a una relación de proximidad R dada en un universo finito de cardinal n es el siguiente.

Algoritmo 3.7

1. *Calcular la clausura Min-transitiva $\overline{R}$ o una aproximación inferior de R.*

2. *Ordenar las entradas $a_1 < a_2 < ... < a_k = 1$ $(k \leq n)$ de $\overline{R}$.*

3. *Reemplazar cada a_i por la media a'_i de las entradas de R que ocupan el mismo lugar que los a_i.*

4. IF $a'_1 \leq a'_2 \leq ... \leq a'_k = 1$, *entonces la relación de similitud E se obtiene reemplazando las entradas $a_1, a_2, ..., a_k$ de $\overline{R}$ por $a'_1 \leq a'_2 \leq ... \leq a'_k$ respectivamente.*

5. ELSE, *para cada cadena maximal $C = \{a'_i, a'_{i+1}, ..., a'_{i+j}\}$ con $a'_i > a'_{i+j}$, reemplazar todos los elementos de C por su media aritmética ponderada a_C, pesando cada a'_l de C por el número de entradas de R que corresponden a a'_l. Reemplazando los elementos de C por a_C en E se obtiene la relación de similitud deseada.*

4 Conclusiones

En este trabajo hemos presentado un algoritmo para hallar aproximaciones de relaciones de proximidad por relaciones de similitud que en general son mejores que su clausura transitiva o sus openings transitivos. Por ejemplo, la distancia euclídea entre la proximidad del Ejemplo 3.4 y su clausura transitiva es 0.825, mientras que su distancia a la matriz obtenida usando el algoritmo propuesto es 0.795. La distancia euclídea entre la proximidad del Ejemplo 3.5 y la similitud obtenida mediante el Teorema de Representación es 1.183, y su distancia a la matriz obtenida usando el algoritmo propuesto es 0.768.

La sencillez del algoritmo permite su aplicación a problemas reales sin necesidad de aumentar la complejidad o tiempo de ejecución respecto a los métodos clásicos.

Los resultados de este trabajo junto con los de [3] permiten hallar buenas aproximaciones de una relación de proximidad mediante un operador de T-indistinguibilidad para las t-normas más usuales: *Mínimo* y t-normas arquimedianas continuas (incluidas la t-norma de *Łukasiewicz* y la t-norma *Producto*).

Agradecimientos

Este trabajo ha sido parcialmente financiado por los proyectos números TIN2006-14311 y TIN2006-06190.

Referencias

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

[2] J. Fodor, M. Roubens (1995). Structure of transitive valued binary relations. *Math. Social Sci.* 30, 71–94.

[3] L. Garmendia, J. Recasens (2007). Approximation of Proximities by Aggregating T-indistinguishability Operators. Mathware & Soft Computing 14, 171-181.

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

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

[6] E. P. Klement, R. Mesiar, E. Pap (2000). *Triangular norms.* Kluwer. Dordrecht.

[7] B. Schweizer, A. Sklar (1983) *Probabilistic Metric Spaces.* North-Holland. Amsterdam.

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

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