

Extensión de Proceso de Explotación de Información para la Detección de Asociaciones Espaciales Difusas

Giovanni Daián Rottoli^{1,2,3}, Hernán Merlino³

¹ Programa de Doctorado en Ciencias Informáticas, Facultad de Informática, Universidad Nacional de La Plata (UNLP). La Plata, Buenos Aires, Argentina.

² Programa de Beca Doctoral para el Fortalecimiento de Áreas I+D+i. Universidad Tecnológica Nacional (UTN). Argentina
rottolig@frcu.utn.edu.ar

³ Grupo de Investigación en Sistemas de Información (GISI), Universidad Nacional de Lanús (UNLa). Lanús, Buenos Aires, Argentina.
hmerlino@gmail.com

Resumen Las asociaciones espaciales brindan conocimiento que resulta de sumo interés para la inteligencia de negocios en dominios varios. Estas asociaciones están constituidas tanto de relaciones espaciales como no espaciales que se dan entre los objetos de una base de datos, las cuales invitan a ser modeladas haciendo uso de la teoría de conjuntos difusos debido a la falta general de límites claros para determinar cuándo un par de objetos pertenece o no a la relación en cuestión, producto de la misma naturaleza de los datos. Un ejemplo es la relación de cercanía entre dos objetos: ¿a qué distancia los mismos dejan de estar cerca entre sí? En consecuencia, en el presente trabajo se extiende un proceso de explotación de información para la detección de asociaciones espaciales basado en la minería de subgrafos frecuentes considerando además relaciones difusas. Se presenta una prueba de concepto utilizando datos generados para tales fines.

Keywords: Asociaciones Espaciales, Spatial Data Mining, Explotación de Información Espacial, Lógica Difusa, Subgrafos Frecuentes

1. Introducción

La explotación de información espacial se encarga de la búsqueda de conocimiento interesante y novedoso en bases de datos espaciales para dar soporte a la toma de decisiones estratégicas [1]. En este contexto, los patrones de asociación resultan de interés para los tomadores de decisiones [2].

Un dato espacial $D = \{\{s_j\}, \{v_k\}\}$ se compone por un conjunto de n atributos espaciales $\{s_j\}, j = 1..n$, que relacionan cada objeto espacial inequívocamente con una ubicación determinada del espacio, y un conjunto de m atributos no espaciales $\{v_j\}, k = 1..m$, que recogen datos relevantes para el dominio del problema en la ubicación detallada [3].

La dependencia espacial de la información admite extraer relaciones espaciales que se encuentran explícitas, las cuales a su vez pueden ser: 1. relaciones geométricas, si se basan en la distancia euclidiana (vecindad, por ejemplo); 2. relaciones topológicas, independientes de transformaciones como el escalado, la rotación y traslación (e.g., intersección, inclusión); 3. relaciones direccionales, relacionadas con la posición de un objeto con respecto a otro (norte, sur, arriba, abajo); 4. relaciones mixtas, si intervienen combinaciones de relaciones anteriormente mencionadas (perpendicularidad, paralelismo) [4–6].

Estas relaciones se asocian con conceptos cuya naturaleza no es posible de ser definida mediante umbrales estrictos, trasladándose esta incertidumbre a los patrones de asociación espaciales en los que las mismas intervienen. Frente a este escenario, la lógica difusa resulta una herramienta adecuada para modelarlas [8, 9].

En trabajos anteriores [7], se ha propuesto un proceso de explotación de información para el minado de asociaciones espaciales que pueden incluir relaciones y datos de distintos tipos y considera fenómenos como la heterogeneidad espacial.

El proceso de explotación de información mencionado deja abierta la posibilidad de utilizar relaciones difusas entre los datos, cuestión que se aborda en el presente trabajo. Para esto, en la sección 2 se ahonda en el concepto de relaciones espaciales (sección 2.1) y de asociaciones espaciales (sección 2.2). Luego, en la sección 3 se propone la adaptación del proceso en cuestión para contemplar tanto relaciones *CRISP* como difusas. Posteriormente, en la sección 4 se realizan pruebas de concepto del proceso propuesto utilizando datos generados para dicho propósito. Finalmente, en la sección 5 se presentan conclusiones y trabajos futuros.

2. Estado del Arte

2.1. Relaciones Espaciales

Sea D una base de datos espaciales, una relación $R \subseteq D^2$ es llamada *espacial* si y solo si está definida mediante un predicado binario $P(x, y) | x, y \in D$ que involucra a los atributos espaciales de los objetos evaluados [6].

La dependencia espacial de cada objeto de la base de datos es la que define implícitamente las distintas relaciones que se pueden extraer entre los mismos. A modo de ejemplo, la relación espacial $V \subseteq D^2$, con $x, y \in D$ y $\lambda \in R^+$, definida como:

$$xVy \leftrightarrow \sqrt{(lat_x - lat_y)^2 + (long_x - long_y)^2} < \lambda \quad (1)$$

es la relación de vecindad entre dos puntos en el espacio utilizando la distancia euclidiana entre sus atributos latitud y longitud.

El cómputo de estas relaciones implica una operación *join* entre dos subconjuntos de datos haciendo que su cálculo sea, por lo general, costoso, dependiendo de la cardinalidad de los mismos.

Las relaciones espaciales pueden clasificarse según su naturaleza en *Geométricas*, *Direccionales*, *Topológicas* e *Híbridas* [10, 11].

Tabla 1: Ejemplo de funciones de distancia.

Nombre	Fórmula
Distancia Euclidiana	$d_E(\vec{P}, \vec{Q}) = (\sum_{i=1}^n (p_i - q_i)^2)^{\frac{1}{2}}$
Distancia de Chebyshev	$d_C(\vec{P}, \vec{Q}) = \max_{i=1}^n (p_i - q_i)$
Distancia Manhattan	$d_M(\vec{P}, \vec{Q}) = \sum_{i=1}^n p_i - q_i $
Distancia Geográfica	$d_G(P, Q) = R((\Delta\phi)^2 + (\cos(\phi_m)\Delta\lambda)^2)^{\frac{1}{2}}$

Las relaciones geométricas están basadas en la geometría euclidiana y suelen ser el tipo de relaciones más utilizadas entre datos espaciales. Se requiere, por lo usual, el cálculo de distancias entre los objetos de la bases de datos (Tabla 1) [12].

Las relaciones direccionales involucran información sobre la orientación relativa de un objeto en un espacio 2D o 3D. Generalmente son definidas en función del concepto de “vecindario”, adoptando para tal propósito las ya mencionadas distancia Manhattan y la de Chebyshev, obteniéndose respectivamente cuatro vecinos (norte, sur, este y oeste) u ocho vecinos (norte, noreste, este, sudeste, sur, sudoeste, oeste, noroeste).

Las relaciones topológicas son independientes de los conceptos de distancia y dirección y no se ven afectadas por transformaciones lineales como traslación, escalado y rotación. Es utilizado frecuentemente el modelo de 9 intersecciones propuesto por [13] para modelar las distintas interacciones topológicas entre dos instancias de cualquier tipo de dato espacial vectorial: líneas, puntos o polígonos. Las mismas poseen diferentes configuraciones que varían entre los distintos tipos de objetos espaciales. Por ejemplo, entre dos polígonos existen ocho posibles relaciones: *disjuntos*, *contiene_a*, *dentro_de*, *iguales*, *conoce_a*, *ubre*, *cubierto_por*, *superpuestos*.

Las relaciones híbridas surgen a partir de la combinación de relaciones clasificadas en dos o más categorías de relaciones diferentes. Por ejemplo, las relaciones de paralelismo y perpendicularidad entre dos líneas son tanto geométricas, debido a que se basan en el ángulo de incidencia de ambos elementos, como topológicas, ya que la propiedad se mantiene invariante frente a transformaciones de traslación, rotación y escalado.

Esta variedad de funciones de tipos de relaciones espaciales y el costo computacional que requiere su cálculo, hacen de la preparación de los datos una tarea considerablemente más compleja que en el explotación de información tradicional.

2.2. Asociaciones Espaciales

Un patrón de asociación se define como un predicado compuesto con alta probabilidad de ocurrencia. Este patrón se dice espacial si al menos uno de los predicados simples que lo componen define una relación espacial [2]. Un ejemplo de una asociación espacial es:

$$Es(x, \text{'Escuela'}) \wedge Es(y, \text{'ZonaResidencial'}) \wedge Contiene(y, x)$$

La relevancia de un patrón es medida utilizando distintas métricas. El soporte (*support*) es una de las más utilizadas, calculándose sobre una base de datos transaccional como la probabilidad de ocurrencia conjunta de los predicados (o valores de sus atributos). En una base de datos espaciales esta medida es equivalente.

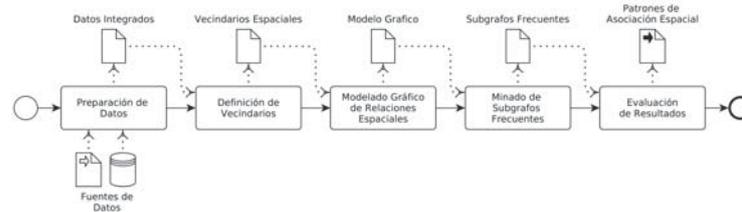


Figura 1: Proceso de Explotación de Información para el descubrimiento de Asociaciones Espaciales [7].

Asimismo, al igual que en la búsqueda de asociaciones tradicional, es posible generar reglas de asociación a partir del cálculo de la confianza de las posibles configuraciones de antecedente y consecuente sobre los patrones de asociación descubiertos [15], la cual se calcula utilizando la probabilidad condicional: $Es(x, \text{'Escuela'}) \wedge Contiene(y, x) \rightarrow Es(y, \text{'ZonaResidencial'})$

3. Extensión del Proceso de Explotación de Información

Previamente, en [7], se ha propuesto un proceso de explotación de información para el descubrimiento de asociaciones espaciales (Fig. 1) que tiene en consideración el fenómeno de la heterogeneidad espacial [16] y la utilización de datos de entrada de naturaleza y tipos varios.

El proceso en cuestión modela las relaciones espaciales mediante una estructura de multigrafo $G = (V, E, L, K, \phi, P_l, P_k)$. Los conjuntos que conforman dicha estructura son V , conjunto de vértices que modelan los objetos espaciales independientemente de su forma; E , conjunto de aristas que modelan las relaciones entre los objetos espaciales; L y K , conjuntos de etiquetas para los vértices y las aristas respectivamente que modelan las características de los mismos. Adicionalmente, la función $\phi : E \rightarrow \{\{v_1, v_2\} | v_1 \in V \wedge v_2 \in V \wedge v_1 \neq v_2\}$ es la función de incidencia de las relaciones espaciales sobre los objetos espaciales, $P_l \subseteq V \times L$ es la relación de etiquetado de características de los objetos espaciales, y $P_k \subseteq E \times K$ es la relación de etiquetado de características de las relaciones espaciales.

A partir de esta herramienta de modelado, se propone la búsqueda de subgrafos frecuentes en el grafo de relaciones original. Esta actividad se lleva a cabo mediante el uso de algoritmos especializados en esta tarea [17, 18]. Cada uno de estos subgrafos representan aquellas relaciones espaciales que se dan frecuentemente de forma conjunta.

El artículo previamente mencionado propone la posibilidad de utilización de relaciones difusas, mas no especifica el procedimiento para hacerlo. En consecuencia, a continuación se propone la representación de relaciones difusas mediante una estructura de multigrafo equivalente, el minado de los subgrafos frecuentes, y la obtención de un valor representativo de los grados de pertenencia de las instancias involucradas en los patrones obtenidos.

3.1. Modelado Gráfico de Relaciones Espaciales Difusas

Los umbrales rígidos para la definición de relaciones entre datos espacialmente referenciados no permiten modelar la realidad de manera apropiada. Póngase como

ejemplo la relación de vecindad: si el criterio tomado establece una distancia máxima de 50 metros entre dos objetos para que estos sean vecinos, ¿es justo excluir un par de datos por encontrarse separados por 51 metros?. Asimismo, ¿Cuál es el límite estricto de una zona de bosque? .

Generalmente hacemos uso de etiquetas lingüísticas como “cerca” o “lejos” para referimos a las relaciones espaciales. Estos conceptos no poseen propiamente límites bien establecidos, por lo que son factibles de ser modelada utilizando herramientas como la teoría de conjuntos difusos (*fuzzy sets*) [14].

Un conjunto difuso se define mediante una función de pertenencia $\mu_{\tilde{A}}(x) : U \rightarrow [0, 1]$, que representa el grado en el cual $x \in \tilde{A}$, siendo x un objeto del universal considerado. Consecuentemente, una relación difusa \tilde{R} es un subconjunto difuso del producto cartesiano clásico $A_1 \times A_2 \times \dots \times A_n$ [19].

A modo de ejemplo, dada la relación de vecindad previamente definida de manera rígida (Ec. 1), es posible extenderla de forma difusa mediante una función de pertenencia como la presentada en la ecuación 2. Consecuentemente, dos objetos espaciales que se encuentren a una distancia menor a λ_1 pertenecerán a la relación con $\mu = 1$; si se encuentran a una distancia comprendida entre λ_1 y λ_2 serán vecinos con $0 < \mu < 1$; y si se alejan a mayor no estarán relacionados.

La definición de las funciones de pertenencia dependerán del dominio del problema abordado en cada caso.

$$\forall x, y \in S, \mu_{\text{Vecinos}}((x, y)) = \begin{cases} 1 & \text{si } \delta \leq \lambda_1 \\ (-\frac{1}{25}\delta + 3)^{\frac{1}{2}} & \text{si } \lambda_1 < \delta < \lambda_2 \\ 0 & \text{si } \delta \geq \lambda_2 \end{cases} \quad (2)$$

siendo $\delta = \text{Distancia}(x, y)$

Para la construcción de las aristas del grafo, entonces, primero deben ser extraídos los grados de pertenencia de las tuplas de objetos espaciales a las relaciones difusas evaluadas. Luego, se modelará una arista por cada tupla perteneciente al soporte de cada relación difusa, siendo éste el conjunto rígido $S(\tilde{A})$ que contiene todos los elementos del universal U cuyo grado de pertenencia al conjunto difuso \tilde{A} sea no nulo. En símbolos: $S(\tilde{A}) = \{x/x \in U \wedge \mu_{\tilde{A}} > 0\}$ [20].

De esta forma, el conjunto de aristas se conforma a partir de las relaciones cuyas tuplas pertenezcan en algún grado a la relación, sin importar el valor de este. Posteriormente se evaluarán los grados de pertenencia de las instancias de los patrones de asociación detectados.

3.2. Minado de Subgrafos Frecuentes

A partir del grafo generado, tal como se especifica en el proceso presentado en [7], se buscan aquellos subgrafos que se repiten frecuentemente en el grafo original haciendo uso de algoritmos de búsqueda de subgrafos frecuentes en un único gran grafo [17, 18]. Los autores proponen el uso del algoritmo SUBDUE ya que hace uso de una medida de prevalencia de los subgrafos basada en la compresión de la información [21].

Posteriormente, por cada uno de los subgrafos frecuentes, se calculan los grados de pertenencia de las instancias de las relaciones que intervienen en el mismo, obteniendo la mínima pertenencia conjunta a las mismas. Para esto haremos uso de una T-norma sobre los grados de pertenencia de cada objeto espacial interviniente en cada una de las instancias del patrón, a los predicados en los que participa [15]. Por último calcularemos

mediante una regla heurística el grado de pertenencia conjunto de las instancias, por ejemplo, mediante el promedio de los mismos.

Si bien la T-norma Mín es la más habitual, esto no reflejaría correctamente la relación existente entre los objetos que conforman la instancia de la regla. Por ejemplo, sea un predicado compuesto $Vecinos(x, y) \wedge Norte(x, y)$ con grados de pertenencia $\mu_1 = (0,5; 1)$ y $\mu_2 = (0,5; 0,7)$, si se utiliza la T-norma Mín, el grado de cumplimiento de la instancia, en ambos casos es 0.5, si bien la segunda instancia cumple en menor medida el segundo predicado.

Por este motivo se propone utilizar la familia de T-normas de Hamacher [22], que permiten tener mayor flexibilidad en los posibles modelos al controlar la tasa de caída de la función mediante su parámetro p (Ec. 3, Fig 2). Además, esta T-norma es equivalente a la T-norma producto cuando $p=1$.

$$T_p^H = \begin{cases} 0 & \text{si } p=x=y=0 \\ \frac{xy}{p+(1-p)(x+y-xy)} & \text{en caso contrario} \end{cases} \quad (3)$$

Pongamos como ejemplo un predicado compuesto $P_1 \wedge P_2 \wedge P_3$ que constituye un subgrafo frecuente, siendo $a_k, k = 1,4$, instancias del mismo con grados de pertenencia a cada una de las relaciones mostrados en la tabla 2. Aplicando la T-norma de Hamacher en cada fila obtenemos el grado de pertenencia descrito en la última columna. Luego, el promedio de dichos grados de pertenencia es 0,36 resultando una medida representativa del cumplimiento del patrón de todos los objetos espaciales intervinientes.

4. Prueba de Concepto

4.1. Materiales y Métodos

El proceso de explotación de información para el descubrimiento de asociaciones espaciales (Fig. 1) posee 5 actividades principales entre las que se encuentran la preparación de los datos y la determinación de regiones de interés mediante la definición de vecindarios. Debido a que la extensión realizada en el presente documento no las involucra, se realizarán las pruebas a partir de la actividad de modelado de relaciones mediante grafos. Supondremos entonces que los datos se encuentran preparados y que cada conjunto de datos corresponde a un vecindario espacial.

Se han realizado pruebas en 3 conjuntos de datos (Fig. 4). Estos consisten en 300 puntos espaciales, clasificados en 3 tipos, 3 regiones y 2 líneas [23]. Las regiones parten

μ	P_1	P_2	P_3	$\otimes_{p=0}^H$
a_1	0.9	0.3	0.5	0.23
a_1	0.8	0.7	0.6	0.43
a_1	0.1	0.4	0.2	0.06
a_1	0.9	1.0	0.8	0.73
			μ_{\otimes}	0.36

Tabla 2: Ejemplo de uso de T-Norma de Hamacher con $p=0$

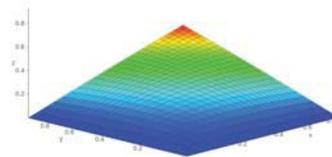


Figura 2: T-norma de Hamacher con $p = 0$.

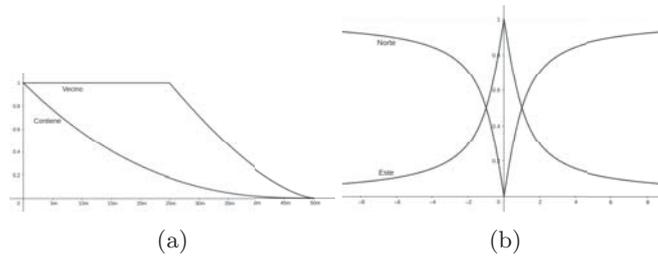


Figura 3: (a) Funciones de pertenencia para las relaciones Cerca y Contiene, en función de la distancia en metros. (b) Funciones de pertenencia para las relaciones Norte y Este, en función de la pendiente de la recta entre dos puntos.

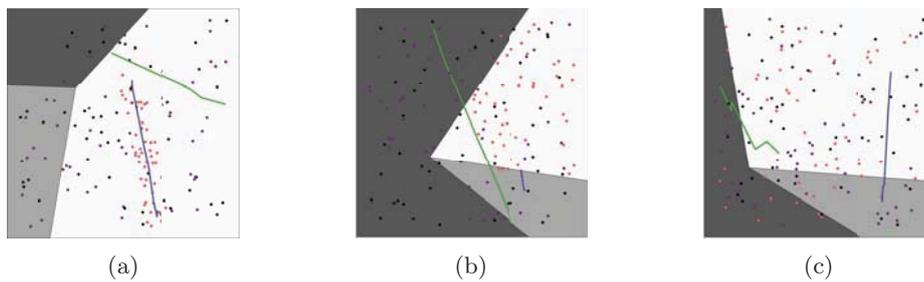


Figura 4: Conjuntos de datos de prueba.

el espacio en 3 partes de manera aleatoria. Las líneas y los puntos se disponen deliberadamente sobre el espacio para conformar ciertos patrones que se desean encontrar, incorporando también un error aleatorio. La prueba de concepto busca determinar si el proceso descubre adecuadamente los patrones de asociación que han sido forzados, y si el valor de pertenencia difuso describe adecuadamente la situación.

Las relaciones difusas binarias que se han utilizado son: “Cerca”, “Contiene”, “Norte”, “Este” y sus negaciones (Fig. 3), algunas de las cuales se detallan en las siguientes ecuaciones. Las relaciones euclidianas se evalúan entre los puntos más cercanos a la geometría involucrada, las direccionales se obtienen a partir del ángulo relativo a las posiciones de los objetos espaciales, y las topológicas se basan en el modelo de 9 relaciones mencionado previamente.

$$\mu_{Cerca}(x, y) = \begin{cases} 1 & \text{si } \delta \leq 25 \\ (-\frac{1}{25} Dist(x, y) + 2)^{\frac{3}{2}} & \text{si } 25 < Dist(x, y) < 50 \\ 0 & \text{si } Dist(x, y) \geq 50 \end{cases} \quad (4)$$

$$\mu_{Contiene}(x, y) = \begin{cases} 1 & \text{si } x \text{ contiene a } y \\ (-\frac{1}{50} Dist(x, y) + 1)^{\frac{5}{2}} & \text{si } Dist(x, y) < 50 \\ 0 & \text{si } Dist(x, y) \geq 50 \end{cases} \quad (5)$$

$$\mu_{Norte}(A, B) = \begin{cases} 1 & \text{si } A_x = B_x \wedge A_y > B_y \\ |\frac{2}{\pi} \tan^{-1} | \frac{A_y - B_y}{A_x - B_x} || & \text{si } A_x \neq B_x \wedge A_y > B_y \\ 0 & \text{en caso contrario} \end{cases} \quad (6)$$

Los cálculos han sido realizados alrededor de un tipo de punto de referencia (puntos rojos, *target*) y codificadas dentro de un archivo JSON adecuado para ser entrada del algoritmo SUBDUE. Una vez extraídos los subgrafos frecuentes, se ha calculado los grados de pertenencia de los predicados intervinientes.

Tanto la generación de los datos como la extracción de relaciones difusas fue realizada utilizando Python3.7 y las librerías Geopandas, Pandas, Shapely y Numpy, disponibles en el entorno Anaconda (<https://www.anaconda.com/>). QGIS3.8 (<https://www.qgis.org/es/site/>) fue utilizado para la validación de los datasets de pruebas. La extracción de los subgrafos frecuentes fue realizada con la herramienta SUBDUE (<http://ailab.wsu.edu/subdue/>).

4.2. Resultados

Por cada subgrafo se analizó cada una de las combinaciones de relaciones posibles para determinar cómo influyen los grados de pertenencia individuales en el grado de pertenencia conjunto del patrón. En el primer caso (Tabla 3) se puede observar que $\bar{\mu}_{\otimes} = 1$ para el primer patrón, al estar todos los puntos rojos contenidos en la región blanca (Fig. 4,a). Al contemplar también la relación de vecindad con la línea azul, se reduce el grado de pertenencia conjunto. Esto significa que existen ciertos puntos que no cumplen totalmente la relación de vecindad.

En el segundo caso se agrega un error aleatorio, por lo que existen ciertos puntos rojos que se encuentran incluidos en la región blanca pero no totalmente, lo cual reduce el valor de $\bar{\mu}_{\otimes}$ (Fig. 4,b). Al incluir la relación direccional Norte, se nota que los puntos se ubican en cierto grado al norte de la línea azul, cuestión que impacta en el grado de pertenencia. De igual forma sucede al considerar la misma relación pero con la línea verde, por lo que se concluye que los objetos espaciales cumplen de forma escasa con dicha propiedad. Si bien el subgrafo frecuente los incluye por haberse considerado el soporte de la relación difusa, es necesario entonces filtrar las aristas del subgrafo cuyo grado de pertenencia sea bajo.

Adicionalmente, con el tercer conjunto de datos se han encontrado tres subgrafos distintos que contemplan las relaciones consideradas. En el primer caso, se nota la posibilidad de que los puntos rojos no se encuentren al este de ninguna de las dos líneas, estando claramente al este de la línea verde. Esto sucede debido a la definición de la relación difusa (Tabla 3, fila 7,9).

Por último, en el tercer patrón detectado, se evidencia cómo la relación direccional no afecta significativamente al valor de la relación topológica, por lo que es un patrón de conocimiento que describe adecuadamente el conjunto de datos. Al agregar la relación direccional con la línea verde se reduce de nuevo, mas sigue estando dentro de valores tolerables.

Los patrones descubiertos describen adecuadamente los conjuntos de datos, y representan de forma apropiada los grados de pertenencia de las instancias de datos a cada uno de las regularidades descubiertas. Sin embargo, es considerable el tiempo y los recursos que demanda la ejecución del algoritmo de minado de subgrafos frecuentes debido a que esta es un problema cuya complejidad escala considerablemente al agregarse aristas y vértices. En futuros trabajos deberá evaluarse este aspecto del proceso.

5. Conclusión

Se ha presentado una extensión del proceso de explotación de información para el descubrimiento de asociaciones espaciales para incorporar relaciones difusas al mismo,

Patrón	Compresión	μ_{\otimes}
$rojo(p) \wedge blanca(r) \wedge contiene(r, p)$	0.219	1.0
$rojo(p) \wedge blanca(r) \wedge azul(l) \wedge contiene(r, p) \wedge cerca(p, l)$	0.219	0.838
$rojo(p) \wedge blanca(r) \wedge contiene(r, p)$	0.119	0.948
$rojo(p) \wedge blanca(r) \wedge contiene(r, p) \wedge azul(l) \wedge norte(p, l)$	0.119	0.815
$rojo(p) \wedge blanca(r) \wedge verde(l) \wedge contiene(r, p) \wedge norte(p, l)$	0.119	0.598
$rojo(p) \wedge blanca(r) \wedge azul(l_1) \wedge verde(l_2) \wedge contiene(r, p) \wedge norte(p, l_1) \wedge norte(p, l_2)$	0.119	0.618
$rojo(p) \wedge verde(l) \wedge \neg este(p, l)$	0.077	0.489
$rojo(p) \wedge azul(l) \wedge \neg este(p, l)$	0.077	0.951
$rojo(p) \wedge verde(l) \wedge este(p, l)$	0.068	0.578
$rojo(p) \wedge blanca(r) \wedge contiene(r, p)$	0.062	0.947
$rojo(p) \wedge blanca(r) \wedge contiene(r, p) \wedge azul(l) \wedge \neg este(p, l)$	0.062	0.954
$rojo(p) \wedge blanca(r) \wedge contiene(r, p) \wedge azul(l_1) \wedge \neg este(p, l_1) \wedge verde(l_2) \wedge este(p, l_2)$	0.062	0.832

Tabla 3: Resultados obtenidos.

lo que permite modelar la realidad asociada a los datos espacialmente referenciados de manera más adecuada.

Esta extensión se lleva a cabo mediante la aplicación de una T-norma sobre el grado de pertenencia de cada instancia del patrón, medida que se promedia para obtener un valor representativo del grado en el cual todos los datos involucrados cumplen la regularidad descubierta.

Se ha realizado una prueba de concepto con datos generados para tal fin, obteniendo resultados que representan adecuadamente las situaciones planteadas. Queda pendiente la prueba sobre datos reales y la evaluación de eficiencia de la aplicación de las herramientas asociadas al proceso.

Agradecimientos

La investigación presentada en este artículo fue parcialmente financiada por el Programa de Doctorado para el Fortalecimiento de Áreas I+D+i (2016-2020) de la Universidad Tecnológica Nacional y por el Proyecto de Investigación 80020160400001LA de la Universidad Nacional de Lanús. Los autores agradecen a Karen Nerak por sus comentarios sobre la redacción y al Prof. Dr. Ramón García-Martínez quien ha motivado estos estudios y ya no se encuentra entre nosotros.

Referencias

1. Mennis, J., & Guo, D. “Spatial data mining and geographic knowledge discovery—An introduction.” *Computers, Environment and Urban Systems* 33.6 (2009): 403-408.
2. Koperski, Krzysztof, and Jiawei Han. “Discovery of spatial association rules in geographic information databases.” *International Symposium on Spatial Databases*. Springer, Berlin, Heidelberg, (1995).
3. Leung, Yee. “Knowledge Discovery in Spatial Data.” Ed. Manfred M. Fischer et al. 1st ed. Vol. 45. Hong Kong: Springer, (2011).
4. Agrawal, Rakesh, and Ramakrishnan Srikant. “Fast algorithms for mining association rules.” *Proc. 20th int. conf. very large data bases(VLDB)*. Vol. 1215. (1994).
5. Appice, Annalisa et al. “Discovery of Spatial Association Rules in Geo-Referenced Census Data : A Relational Mining Approach.” *Intelligent Data Analysis 7* (2003): 541–566.

6. Rottoli, Giovanni Daián. "Formulación de Procesos para una Ingeniería de Explotación de Información Espacial." *XXI Ibero-American Conference on Software Engineering*. Doctoral Symposium. Bogotá, Colombia, (2018).
7. Rottoli, Giovanni Daián, Hernán Merlino, and Ramón García-Martínez. "Proceso para el descubrimiento de asociaciones espaciales mediante subgrafos frecuentes." *XXIV Congreso Argentino de Ciencias de la Computación* (La Plata, 2018).. 2018.
8. Shi, Wenzhong, Anshu Zhang, and Geoffrey I. Webb. "Mining significant crisp-fuzzy spatial association rules." *International Journal of Geographical Information Science* 32.6 (2018): 1247-1270.
9. Ladner, Roy, Frederick E Petry, and Maria A Cobb. "Fuzzy Set Approaches to Spatial Data Mining of Association Rules." *Transactions in GIS* 7.1 (2003): 123-138.
10. Appice, Annalisa, et al. "Discovery of spatial association rules in geo-referenced census data: A relational mining approach." *Intelligent Data Analysis* 7.6 (2003): 541-566.
11. Malerba, Donato. "A relational perspective on spatial data mining." *International Journal of Data Mining, Modelling and Management* 1.1 (2008): 103-118.
12. Maiti, Sandipan, and R. B. V. Subramanyam. "Mining behavioural patterns from spatial data." *Engineering Science and Technology, an International Journal* 22.2 (2019): 618-628.
13. Egenhofer, Max J., and Robert D. Franzosa. "Point-set topological spatial relations." *International Journal of Geographical Information System* 5.2 (1991): 161-174.
14. Gader, Paul D. "Fuzzy spatial relations based on fuzzy morphology." *Proceedings of 6th International Fuzzy Systems Conference*. Vol. 2. IEEE, 1997.
15. Laube, Patrick, Mark De Berg, and Marc Van Kreveld. "Spatial support and spatial confidence for spatial association rules." *Headway in Spatial Data Handling*. Springer, Berlin, Heidelberg, 2008. 575-593.
16. Miller, Harvey J. "Tobler's first law and spatial analysis." *Annals of the Association of American Geographers* 94.2 (2004): 284-289.
17. Bianco, Santiago. "Análisis Comparativo de Algoritmos de Minería de Subgrafos Frecuentes." *Revista Latinoamericana de Ingeniería de Software* 4.2 (2016): 111-142.
18. Jiang, Chuntao, Frans Coenen, and Michele Zito. "A survey of frequent subgraph mining algorithms." *The Knowledge Engineering Review* 28.1 (2013): 75-105.
19. Zadeh, Lotfi A. "Fuzzy sets." *Information and control* 8.3 (1965): 338-353.
20. Lazzari, L., et al. "Los conjuntos borrosos y su aplicación a la programación lineal." Facultad de Ciencias Económicas, Universidad de Buenos Aires, Buenos Aires (2001).
21. Ketkar, Nikhil S., Lawrence B. Holder, and Diane J. Cook. "Subdue: Compression-based frequent pattern discovery in graph data." *Proceedings of the 1st international workshop on open source data mining: frequent pattern mining implementations*. ACM, 2005.
22. Fodor, János C., and Tibor Keresztfalvi. "A characterization of the Hamacher family of t-norms." *Fuzzy Sets and Systems* 65.1 (1994): 51-58.
23. Rottoli, Giovanni Daián. "Datos para la prueba de la extensión difusa del proceso para SARM", *Mendeley Data*. (2019). v1. <http://dx.doi.org/10.17632/b254wfhfkg.1>