

Minería de datos espaciales

Spatial data mining – An overview

Denisse Cangrejo Aljure¹ M. Sc. & Juan G. Agudelo² Ing.

1. Universidad Nacional de Colombia sede Bogotá, ANGeoSc.

2. Universidad Nacional de Colombia sede Bogotá, ANGeoSC.

ldcangrejoa@unal.edu.co, jgagudelog@unal.edu.co

Recibido para revisión 18 de junio de 2011, aceptado 18 de octubre de 2011, versión final 24 de noviembre de 2011

Resumen — El objetivo de este artículo es ofrecer una definición formal de Minería De Datos Espaciales además de proveer una introducción describiendo las principales características, patrones comunes y métodos que permiten identificar las fortalezas, oportunidades y aplicaciones para extraer información nueva y relevante que se encuentra contenida en las bases de datos espaciales.

Palabras Clave — Minería de Datos Espaciales, SDM, Base de Datos Espaciales, SDB, Patrones, Métodos, Aplicaciones.

Abstract — This paper is intended to provide a formal definition and the introduction of Spatial Data Mining, SDM, describe the patterns, methods and the main characteristics to identify current strengths, weakness, opportunities and the use of the SDM to extract new and useful information embedded in to the spatial database.

Keywords — SDM, Spatial Data Mining, SDB, Spatial Database, Mining Pattern, Mining Methods, SDM Applications.

I. INTRODUCCION

En la actualidad los sistemas de información geográfica involucran el almacenamiento y procesamiento de grandes volúmenes de datos espaciales con atributos que proveen información directa asociada a una localización en el espacio, además de metadatos que registran la información de los datos; sin embargo, existen numerosos patrones de comportamiento de los datos geográficos que pueden proveer información nueva, útil, relevante y de gran importancia para las organizaciones [31] [35]. Es en torno a la identificación de dichos modelos donde la minería de datos espaciales juega un papel preponderante.

En el presente documento se abordarán los diferentes aspectos de la Minería de Datos Espaciales, SDM, a partir de la definición general, de sus características generales, funcionamiento, métodos, aplicaciones y tendencias en el mundo actual.

II. DEFINICIÓN DE MINERÍA DE DATOS ESPACIALES

El desarrollo de las bases de datos espaciales y el uso de métodos como la minería de datos espaciales, tienen origen

en desarrollos conceptuales y tecnológicos en diversas áreas vinculadas a la información geográfica en su dimensión espacial; en particular, avances en procesamiento de la información espacial y herramientas como los sistemas de información geográfica o las bases de datos espaciales.

La información espacial ha estado asociada en forma directa con la cartografía, en el logro de objetivos específicos concernientes con operaciones de análisis y gestión de datos espaciales, en las cuales se representa dicha información con modelos que usan de mapas y símbolos [30][31]. Por otra parte, la evolución de la informática y la electrónica ha dado paso al surgimiento de Sistemas de Información Geográfica, SIGs, que constituyen una poderosa herramienta que considera las componentes espacial, temática y temporal de los datos y provee un sistema para la gestión de dichos datos georeferenciados, con herramientas de visualización, consulta, edición y análisis espacial. Los SIGs manejan información de los datos temáticos y disponen de una capa para manejar la geometría asociada a los datos y la ubicación espacial de dicha geometría que es almacenada en una Base de Datos Espacial, SDB [2][14][17].

En la actualidad se dispone de bases de datos espaciales extensas, en constante crecimiento, que brindan grandes cantidades de información útil. Se hace evidente por tanto, la necesidad de interpretar y descubrir información relevante sobre ellos [35] y de realizar búsquedas orientadas a identificar objetos y relaciones emergentes con valor significativo, tarea que puede involucrar niveles de complejidad importantes [1][3].

En respuesta a esta necesidad surge la minería de datos espaciales, SDM como el proceso de análisis automático, mediante la implementación de algoritmos, que brinda la posibilidad de buscar correlaciones no evidentes y potencialmente útiles entre objetos geográficos [33][34]. La aplicación de la SDM conduce al descubrimiento de conocimiento implícito que puede ser evidenciado sobre la información de los datos geográficos, en la deducción de patrones o categorizaciones, en la información convencional de las bases de datos y en el caso de los objetos geográficos de manera asociativa, en forma de estructuras, agrupaciones y diversos tipos de relaciones espaciales [1][4]. La minería de datos espaciales es considerada una rama de la minería de datos tradicional [28] y hace énfasis en la extracción

de conocimiento relevante inherente a la naturaleza espacial de los datos. [7][8][9].

III. TAREAS EN LA MINERÍA DE DATOS ESPACIALES

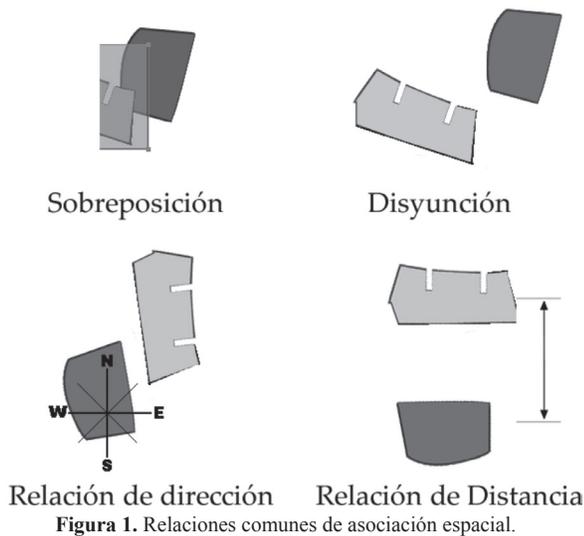
La minería de datos espaciales abarca 4 grandes tareas y para la realización de cada una de ellas existen múltiples y diferentes métodos provenientes del área computacional, estadística y visual, además de algunas combinaciones de estos métodos [35]. Las tareas comunes que se realizan en la minería de datos son:

A. Clasificación espacial y predicción

La clasificación hace referencia directa a la agrupación de datos puntuales en clases o categorías a partir de valores de los atributos contenidos en los datos; este tipo de clasificación también es llamada “Clasificación supervisada” [2] ya que se necesitan conjuntos de datos de formación entrenados para configurar el modelo de clasificación [49], conjuntos de datos de validación para optimizar la configuración y un conjunto de datos para pruebas que permiten evaluar el desempeño del modelo entrenado. Los métodos de clasificación incluyen árboles de decisión, redes neuronales, estimación de máxima verosimilitud, análisis de discriminante lineal, máquinas de soporte vectorial, k-nn (K nearest neighbors) y razonamiento basado en casos entre otros [38][39][40].

B. Reglas de asociación espacial

Representan las relaciones existentes entre los objetos y los predicados, contenidas en los predicados espaciales. Esta tarea originalmente se destinó a identificar regularidades entre objetos involucrados en transacciones de bases de datos [50].



Estas reglas también pueden representar relaciones topológicas entre objetos espaciales como disyunción, intersección, adyacencia, sobreposición, vecindad e igualdad entre otras [4]; también pueden representar la orientación o

el orden espacial o contener información acerca de distancias como un objeto que está cerca de, o lejos de, etc [5][9].

C. Agrupación espacial, regionalización y análisis de patrones de punto

Estas tareas han sido ampliamente utilizadas en análisis de datos que organizan un conjunto de datos en grupos o clusters, de tal forma que los elementos del mismo grupo son similares entre sí y diferentes de los otros. En cuanto a los métodos de clustering [33], se pueden clasificar en dos grandes tipos como son: Agrupación por Separación o Agrupación Jerárquica. La primera divide un conjunto de datos en un número de grupos que no se superponen, un dato se asigna al grupo más cercano basado en la proximidad o en una medida de disimilitud.



Figura 2. Incidencia de eventos y áreas de afectación.

De otro lado, la agrupación jerárquica [34] organiza los datos en jerarquías con una secuencia de particiones o agrupaciones. El término regionalización se refiere a una forma particular de agrupación que busca objetos espaciales en grupos contiguos, mientras optimiza una función objetivo.

En cuanto al análisis de patrones de punto [48], también conocido como Análisis de Punto Caliente, Hot Spot, el cual se enfoca en la detección de la concentración de eventos inusuales en el espacio, como grupos geográficos de enfermedades, crímenes, o accidentes de tránsito. El objetivo en este caso es saber si hay un exceso de ocurrencias del objeto observado en un área determinada.

D. Geovisualización

Se ocupa del desarrollo de la teoría y los métodos que facilitan la construcción de conocimiento a través de la exploración visual, del análisis de datos geoespaciales y de la implementación de herramientas visuales para la posterior recuperación, síntesis, comunicación y uso de conocimiento [46]. La geovisualización a diferencia de la cartografía hace énfasis en el desarrollo de

mapas altamente interactivos y herramientas asociadas para la exploración de datos, la generación de hipótesis y la generación de conocimiento [30][45].

IV. PATRONES COMUNES EN LA MINERÍA DE DATOS ESPACIALES

El descubrimiento de patrones en la minería de datos espaciales es mucho más complejo ya que se ocupa no sólo de los patrones correspondientes a datos categóricos y numéricos de la minería de datos tradicional [6], sino que además tiene en cuenta la identificación, localización y distinción de los objetos [32] y sus múltiples relaciones topológicas como son intersecciones, superposiciones, adyacencias y disyunciones entre otros, además de las relaciones de orientación espacial y distancia entre los objetos del ambiente y el objeto observado [4][13][32].

A. Generalización de datos no espaciales dominantes

Para los datos no espaciales un primer paso es la generalización, que consiste en una colección que involucra todos los atributos relevantes de la búsqueda, posterior a ello se realiza la creación de regiones que son combinaciones de vecindad que comparten un alto nivel de semejanza en las descripciones de los atributos [36]; en este punto se colectan los apuntadores a los objetos espaciales y se agrupan como datos no espaciales generalizados [14] [17][12].

B. Generalización de datos espaciales dominantes

Para el caso de los datos espaciales se colectan todos los datos brindando una jerarquía de datos espaciales; en este punto la generalización puede ser presentada como la partición en regiones y la posterior fusión de regiones espaciales de acuerdo a los atributos espaciales de los datos, donde la fusión está determinada por los parámetros definidos en la estructura jerárquica discriminante [36], este proceso es realizado hasta que se alcance el límite mínimo de tamaño definido [14][17], posteriormente se realiza un análisis de los objetos emergentes donde la calidad en los patrones encontrados está directamente relacionada con la conveniencia y precisión dadas en las jerarquías de los datos [12].

C. Patrones de asociación espacial

Las reglas de asociación espacial describen la implicación de características espaciales del objeto de análisis con respecto a las características de otros objetos, siendo éstos, únicos o de agrupación espacial. En las bases de datos espaciales las reglas de asociación espacial son de la forma $X \rightarrow Y$ [47], donde X , antecedente y Y , consecuente, son conjuntos de predicados (espaciales y no espaciales), los cuales generalmente incluyen relaciones topológicas, de orientación espacial y de distancia [13][17][37]. Co-localización es un tipo particular de asociación espacial que hace referencia a dos objetos que se encuentran en la misma localización espacial y significativamente cerca; esta

relación es un caso particular debido a su relación de vecindad que no obedece a una transacción natural entre antecedente y consecuente [17][47].

D. Autocorrelación especial

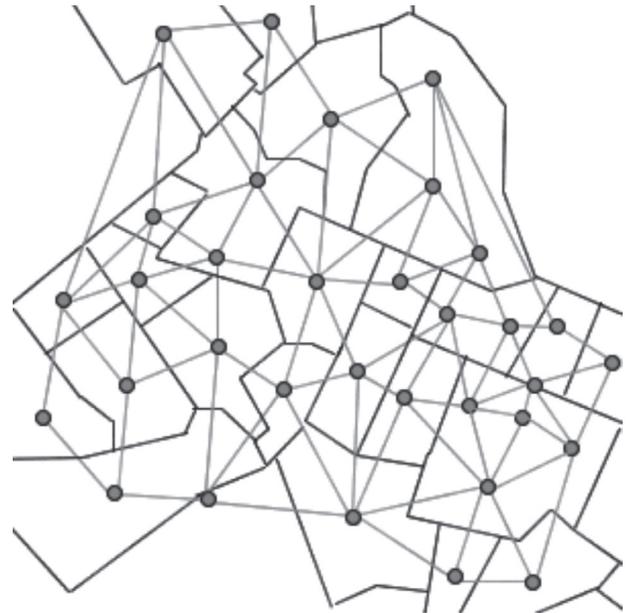


Figura 3. Relaciones de autocorrelación espacial.

Este patrón hace referencia a la medida de la relación de una variable con respecto a sí misma, se utiliza en análisis de exploración de datos espaciales y específicamente en técnicas de agregación espacial [37][47]; la autocorrelación espacial se presenta cuando ocurren Clusters o agrupaciones con valores de variables semejantes y donde los objetos espaciales se encuentran en una diferente localización en el espacio [17] [14].

V. MÉTODOS DE MINERÍA DE DATOS

Las técnicas de minería de datos se pueden categorizar dependiendo sus objetivos específicos y los métodos que utilizan, como se lista a continuación [14]:

A. Modelamiento predictivo

La predicción de eventos que ocurren en una ubicación geográfica particular es muy importante en varios tipos de aplicaciones, por ejemplo, en análisis de criminalidad, en redes celulares y en desastres naturales como los terremotos y sequías, entre otros [16][27]. El objetivo es predecir algunos campos en una base de datos a partir de otros.

Este modelo es un problema de regresión si el campo a predecir es una variable numérica continua, pero si el campo es categórico, entonces es un problema de clasificación. En cuanto a técnicas de regresión y clasificación, se tienen varias opciones [14][15], como los algoritmos de Markov Random Fields,

MRF [39] y Spatial Auto Regression, SAR [41] algoritmos que permiten hacer predicción y clasificación de datos espaciales, siendo el objetivo, determinar el valor de una variable, partiendo de los valores de otros campos de la base de datos [14][15][27].

Esta clase de método es útil en aplicaciones de administración del medio ambiente, seguridad pública, transporte, y evaluación de la capacidad de crédito, además de diversos tipos de modelos de predicción como pueden ser clima, comportamiento y trayectorias entre otros [2][42][43].

B. Resumen de datos espaciales

El objetivo es extraer patrones compactos que describen los subconjuntos de datos. Esta clase de métodos a diferencia de los dos anteriores, no predice, clasifica, ni agrupa; el fin es identificar las relaciones existentes entre los campos. Un método conocido es el método de Asociación de Reglas que se definió en la sección III [14][30][47].

Para describir los datos en forma global con este método, se pueden usar métodos estadísticos, como la varianza o análisis factorial de las estructuras espaciales, o bien, métodos de generalización de los datos espaciales [26]. En el primero, se usa el método de auto correlación global que aplica estadística elemental como el promedio y la varianza y gráficas como los histogramas, diagramas de barras entre otros; además se han desarrollado nuevos métodos como la varianza local, la covarianza local y la auto correlación espacial para medir la dependencia de vecindades a nivel global [26]. De otra parte el método de análisis de densidad no requiere conocimiento alguno de los datos; se pretende estimar la densidad, computando la intensidad de cada punto o área en el espacio para extraer los patrones de punto asociados. En cuanto al segundo, la generalización consiste en elevar el nivel de abstracción de los atributos no espaciales y reducir el detalle de la descripción geométrica mediante la fusión de objetos adyacentes [26].

C. Clustering

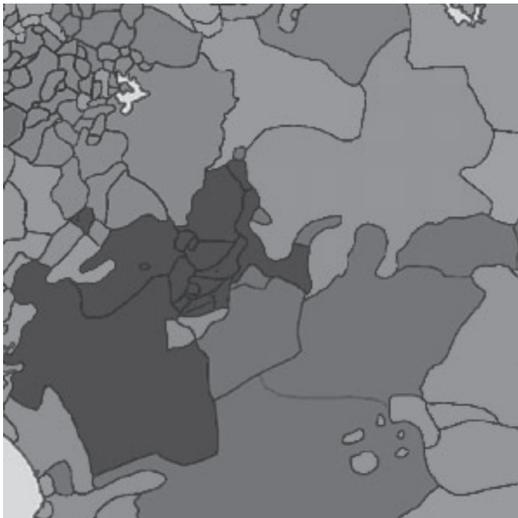


Figura 4. Clustering o segmentación espacial.

EL método clustering, también conocido como segmentación, no especifica los campos que va a predecir y separa los datos en subconjuntos de elementos similares. En este caso no se conoce el número deseado de grupos a diferencia del método de clasificación. Los algoritmos de clustering [29][33] usualmente utilizan dos fases de búsqueda: un ciclo externo sobre el mayor número posible de grupos y un ciclo interno para adaptarse al mejor grupo posible [14]. Estos métodos realizan tareas de agrupación de objetos de una base de datos en grupos llamados clusters, conformados por miembros tan similares como sea posible, es decir que en los otros clusters se encuentran objetos diferentes [2][24][25][27].

VI. APLICACIONES CON SOPORTE DE MINERA DE DATOS ESPACIALES

El campo de aplicación de minería de datos espaciales es amplio y diverso centrándose en objetivos puntuales para aplicaciones específicas que aplican diversas técnicas de SDM donde los objetos espaciales resultantes cuentan adicionalmente con atributos no espaciales propios y se convierten en los datos de entrada de algoritmos para inducir conocimiento y realizar análisis multidimensionales, para ello es posible el uso de árboles de decisión sobre los procesos de clasificación y redes Bayesianas, entre otros, para identificar las relaciones existentes entre los atributos de los datos [18], a continuación se muestran algunas aplicaciones que tienen funcionalidad para realizar procesos de minería de datos espaciales o que tienen como base dichos procesos para obtener resultados específicos.

A. SD-Miner

Uno de los sistemas más conocidos es el SD-Miner el cual soporta los métodos específicos de SDM como lo son clustering, clasificación espacial, caracterización espacial y reglas de asociación espacial. La aplicación está compuesta de tres partes: la Interfaz Gráfica de Usuario, GUI, el módulo SD-Miner y el módulo de administración de bases de datos. La interfaz gráfica de usuario se encarga de mostrar los resultados en tablas, gráficas y mapas, donde cada método de minería incluye diversas variables de entrada y diferentes formas de presentación de los resultados. El SD-Miner es el módulo que se encarga de procesar los datos con los métodos de minería y transferir los resultados a una base de datos. Finalmente, el módulo de administración soporta la comunicación entre el módulo SD-Miner y las bases de datos y maneja cuatro categorías de datos: jerarquías, datos espaciales, datos no espaciales y datos temporales. Los beneficios de la aplicación se enfocan en la facilidad de uso de datos espaciales y no espaciales, con la capacidad de detectar su naturaleza. Además los algoritmos están desarrollados como una librería, lo cual permite que sean implementados por otros sistemas permitiendo el desarrollo de nuevas funciones [19].

B. GeoDMA

Geographical Data Mining Analyst, GeoDMA, es un complemento del software TerraView, y soporta el uso de

datos espaciales para la comparación de imágenes y regiones resultantes de procesos de segmentación en análisis de imágenes, optimizando la segmentación de imágenes y la extracción de varias características incluyendo atributos de selección, clasificación, validación y visualización [44]. DeoGMA utiliza arboles de decisión y algoritmos para mapas auto-organizados; dentro de los retos futuros de este software se encuentra la implementación de algoritmos de minería de datos multi-temporal, ampliar los mecanismos de clasificación y la precisión en la segmentación [44].

C. Ontology-based data mining approach implemented for sport marketing

Es un sistema que combina técnicas de minería de datos espaciales con ontologías para la comercialización de artículos deportivos y toma como caso de estudio ha ADIDAS en Taiwan, donde se buscan estrategias de mercado para las cuales resulta útil saber quiénes son los clientes y además se conocer la segmentación del mercado para saber cuáles son los perfiles de éstos. Para ello proponen los autores un algoritmo de asociación de reglas y de clustering, técnicas de minería de datos espaciales basada en ontologías para obtener patrones, reglas y mapas, encaminadas a formular soluciones para problemas de compañías que están en el negocio [21], llegando incluso a identificar las preferencias deportivas de los clientes [22].

VII. ACTUALIDAD Y TENDENCIAS

Esta sección está dedicada a la revisión de algunos trabajos que hacen contribución a la minería de datos espaciales y a la identificación de algunas líneas de investigación. Dado el apogeo de los sensores y los sistemas de posicionamiento global, GPS, la minería de datos espaciales ha encontrado una nueva área de aplicación donde los objetivos principales son la caracterización de perfiles poblacionales y la identificación de las posibles trayectorias futuras de los usuarios de dispositivos móviles, es decir, se identifica el movimiento de los objetos almacenando una secuencia de coordenadas espaciales asociadas a un instante en el tiempo y posteriormente al aplicarse técnicas de minería de datos espaciales es posible establecer patrones de comportamiento y estimar las posibles posiciones futuras de los objetos de estudio en forma de categorías, donde la clasificación de las trayectorias puede ser desde la perspectiva global, es decir, desde las propiedades que personalizan la trayectoria de los objetos o desde la perspectiva local a partir de las propiedades que determinan las partes de la trayectoria del objeto incluyendo características como la longitud de una ruta, velocidad y aceleración [20].

Así encontramos aplicaciones fundamentales por ejemplo en ecología donde el análisis de rutas de movimientos de especies de animales permite entender el comportamiento natural de las especies, determinar las zonas de protección animal, la reacción

de las especies de estudio ante fenómenos extraños o desastres naturales con base en el estudio de cambios de movimientos abruptos en el comportamiento de las trayectorias [51]; Por su parte Seth Spielman y Jean-Claude Thill, en su trabajo Social Area Analysis, Data Mining and GIS, construyeron mapas de problemas sociales asociados al espacio geográfico, con lo que los sistemas de información toman importancia en este caso porque facilitan la visualización y la administración de datos georeferenciados [23]. Ellos presentan una integración de técnicas de minería de datos con las características propias de un SIG, para caracterizar la población urbana de Nueva York usando grandes colecciones de datos [23].

Al combinar el análisis de trayectorias, dispositivos móviles, sistemas basados en localización, LBS, mercadeo y comercio electrónico, encontramos aplicaciones de minería de datos que permiten realizar estudios y proyecciones de mercado con base en el análisis de trayectorias, que permiten hacer interpolación con diferentes eventos que afectan el mercado para así identificar el comportamiento de los clientes dependiendo de los eventos ocurridos y realizar proyecciones basadas en las características propias de los usuarios y determinar dónde y cuándo promocionar y realizar un evento o una campaña, para garantizar que la cobertura y el mensaje llegue al mayor número posible de usuarios específicos y cuyo interés es acorde con el objeto del mensaje, además de identificar estrategias y posibles alianzas de negocio [51][20].

VIII. CONCLUSIÓN

La Minería de Datos Espaciales se constituye como una herramienta que provee la base teórica y metodológica para la identificación de patrones sobre los datos, los cuales se convierten en la herramienta fundamental en el análisis de los datos y la realización de proyecciones para el soporte a decisiones, así al considerar las características principales de los datos espaciales se realizan tareas comunes de SDM que permiten la predicción a partir de clasificación y categorización. Los retos modernos de la minería de datos están vinculados al análisis y presentación de patrones complejos que deben ofrecer un procesamiento más eficiente, donde el tiempo de respuesta y la visualización de los datos se convierten en el principal objetivo de optimización y desarrollo de algoritmos.

REFERENCIAS

- [1] Keim, D.; Panse, C.; Sips, M.; North, S. 2004. Pixel based visual data mining of geo-spatial data. *Computers & Graphics*. Vol. 28, pp. 327-344.
- [2] D. Guo and J. Mennisb, 2009. Spatial data mining and geographic knowledge discovery—An introduction, *Computers, Environment and Urban Systems*, Vol. 6, pp. 403-408.

- [3] M. Klemettinen, P. Lanzi and R. Meo, 2004. Database support for data mining applications: discovering knowledge with inductive queries, Springer, pp. 96-116.
- [4] L. Amaral and M. Santos, 2004. Mining geo-referenced data with qualitative spatial reasoning strategies, *Computers & Graphics*, Vol. 28, pp. 371-379.
- [5] H. Mori, 2006. State-of-the-Art Overview on Data Mining in Power Systems, *Power Systems Conference and Exposition, PSCE '06*, 2006 IEEE PES.
- [6] L. Xiang and H. YueShun, 2009. A Study of Spatial Data Mining Technique Based on Web, *Management and Service Science, MASS International Conference on 2009*.
- [7] C. Chen, C. Chang and J. Chen, 2005. Spatial Knowledge discovery using spatial data mining method, *Geoscience and Remote Sensing Symposium, IGARSS Proceedings 2005 IEEE International*.
- [8] M. Ester, H. Kriegel and J. Sander, 2009. Algorithm for Characterization and Trend Detection in Spatial Databases, 4th Int. Conference on Knowledge Discovery and Data Mining (KDD '98), Institute for Computer Science, University of Munich, Germany.
- [9] J. Han and K. Koperski, 1995. Discovery of Spatial Association Rules in Geographic Information Databases, *Advances in Spatial Databases, 4th Symposium, SSD95 (Aug.6-9, Portland, Maine)*, pp 47-66.
- [10] M. Ester, H. Kriegel, J. Sander and X. Xu, 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases, *Institute for Computer Science, Second International Conference on Data Mining*, pp 226-231.
- [11] M. Ester, H. Kriegel, and J. Sander, 1999. Knowledge Discovery in Spatial Databases, *Institute for Computer Science, 23rd German Conf. on Artificial Intelligence*.
- [12] Y. Manolopoulos, A. Papadopoulos and G. Sasilakopoulos, 2005. *Spatial Databases: Technologies, Techniques and Trends*. En: J. Wang, W. Hsu and M. Lee, *Mining in Spatio – Temporal Data Bases, USA, IDEA Group Publishing*, pp 272 – 293.
- [13] E. Clementini and P. Di Felice, 1996. A model for representing topological relationships between complex geometric features in spatial databases, *Information Sciences*, Vol. 90, pp. 121-136.
- [14] U. Fayyad and P. Stolorz, 1997. Data mining and KDD: Promise and challenges, *Future Generation Computer Systems*, Vol. 13, pp. 99-117.
- [15] C. Hu, X. Qin and J. Zhang, 2007. Spatial Prediction Models for Mining Spatial Data, *College of Information Science and Technology Nanjing University of Aeronautics & Astronautics, International Conference on Integration Technology on 2007*.
- [16] K. Koperski, 1999. A progressive refinement approach to spatial data mining, *Tesis Dr Philosophy, California, Simon Fraser University, School of Computing Science*, 175 P.
- [17] B. Hall and A. Yeung, 2007. *Spatial Database Systems: Design Implementation and Project Management*, Toronto, Canada, Springer, 549 P.
- [18] Z. Jia and Y. Liu, 2008. Integreted GIS and Interactive Spatial Data Mining for Land Evaluation, *Intelligent Systems and Applications, 2009. ISA 2009. International Workshop on 2008*, pp 1-4.
- [19] D. Bae, J. Baek, H. Oh, J. Song and S. Wook, 2009. SD-Miner: A Spatial Data Mining System, *College of Information & Communitations, Hanyang University*, pp 803-807.
- [20] S. Dodge, E. Forootan and R. Weibel, 2009. Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects, *Computers, Environment and Urban Systems*, Vol. 33, pp 419-434.
- [21] J. Chen, T. Hsu and S. Liao, 2009. Ontology-based data mining approach implemented for sport marketing, *Expert Systems with Applications*, Vol. 36, pp. 11045-11056.
- [22] O. Corcho, M. Lopez and A. Perez, 2003. Methodologies, tools and languages for building ontologies, Where is their meeting point?, *Data and Knowledge Engineering*, Vol. 46, pp. 41-64.
- [23] T. Jean-Claude and E. Seth, 2008. Social area analysis, data mining, and GIS, *Computers, Environment and Urban Systems*, Vol. 32, pp. 110-122.
- [24] M. Ester, H. Kriegel and J. Sander, 2001. Algorithms and Applications for Spatial Data Mining. *Geographic Data Mining and Knowledge Discovery, Research Monographs in GIS*, Taylor and Francis, University of Munich, Germany.
- [25] A. Gordon, 1996. *Clustering and classification*, World Scientific Publisher. pp. 65-122.
- [26] K. Zeitouni, 2000. *A Survey of Spatial Data Mining Methods Databases and Statistics Point of Views*, University of Versailles.
- [27] Z. Pusheng, S. Shashi, R. Ranga and H. Yan, 2003. Trends in Spatial Data Mining, *Department of Computer Science and Engineering, University of Minnesota*.
- [28] J. Han and M. Kamber, 2000. *Data mining: Concepts and Techniques*, New York: Morgan-Kaufman.
- [29] D. Birant and A. Kuta, 2007. ST-DBSCAN: An algorithm for clustering spatial-temporal data, *Data & Knowledge Engineering*, Vol. 60, pp. 208-221.
- [30] C. Kim, *Spatial Data Mining, 2009. Geovisualization, International Encyclopedia of human Geography*, pp 332-336.
- [31] M. Bertolotto, P. Compieta, S. Di Martino, F. Ferrucci and T. Kechadi, 2007. Exploratory spatio-temporal data mining and visualization, *Journal of Visual Languages & Computing*, Vol. 18, pp. 255-279.
- [32] G. Andrienko, N. Andrienko and P. Gatalisky, 2003. Exploratory spatio-temporal visualization: an analytical review, *Journal of Visual Languages & Computing*, Vol. 14, pp. 503-541.
- [33] C. Eick and T. Ryu, 2005. A database clustering methodology and tool, *Information Sciences*, Vol. 17, pp. 29-59.
- [34] Y. Chen and C. Shen, 2005. Mining generalized knowledge from ordered data through attribute-oriented induction techniques, *European Journal of Operational Research*, Vol. 166, pp. 221-245.
- [35] D. Mattfeld and S. Meisel, 2010. Synergies of Operations Research and Data Mining, *European Journal of Operational Research*, Vol. 206, pp. 1-10.
- [36] J. Han, W. Lu and B. Ooi, 1993. Discovery of general knowledge in large spatial databases. *Proceedings of the Far East Workshop on Geo- graphic Information Systems*, pp. 275-289.
- [37] M. Shan and L. Wei, 2010. Algorithms for discovery of spatial co-orientation patterns from images, *Expert Systems with Applications*, Vol. 37, pp. 5795-5802.
- [38] W. Chen and Y. Du, 2009. Using neural networks and data mining techniques for the financial distress prediction model, *Expert Systems with Applications*, Vol. 36, pp. 4075-4086.
- [39] D. Chau, E. Ngaia and L. Xiub, 2009. Application of data mining techniques in customer relationship management: A literature review and classification, *Expert Systems with Applications*, Vol. 36, pp. 2592-2602.
- [40] T. Chena, X. Ma, K. Xieb and L. Wanga, 2005. Efficient discovery of multilevel spatial association rules using partitions, *Information*

- and Software Technology, Vol. 47, pp. 829-840.
- [41] J. Leathwick, A. Lehmann and J. McC. Overton, 2003. GRASP: generalized regression analysis and spatial prediction, *Ecological Modelling*, Vol. 160, pp. 165-183.
 - [42] E. Buia, B. Henderson and K. Viergever, 2006. Knowledge discovery from models of soil properties developed through data mining, *Ecological Modelling*, Vol. 191, pp. 431-446.
 - [43] W. Chang, C. Lee and S. Liao, 2008. Mining marketing maps for business alliances, *Expert Systems with Applications*, Vol. 35, pp. 1338-1350.
 - [44] F. Castro, L. Garcia, T. Korting and M. Sobral, 2008. GeoDMA A novel system for spatial data mining, 2008 IEEE International Conference on Data Mining Workshops, pp 975-978.
 - [45] J. Dikes, M. Kraak and M. Maceachren, 2007. Exploring Geovisualization, International Cartographic Association, Elsevier.
 - [46] J. M Kraak. and F. Oberling, 1996. *Cartography: Visualization of Spatial Data*, Longman, Essex, U.K.
 - [47] D. Cheung, W. Lian and S. Yiu, 2005. An efficient algorithm for finding dense regions for mining quantitative association rules, *Computers & Mathematics with Applications*, Vol. 50, pp. 471-490.
 - [48] E. Renshaw and A. Särkkä, 2006. The analysis of marked point patterns evolving through space and time, *Computational Statistics & Data Analysis*, Vol. 51, pp. 1698-1718.
 - [49] D. Kontos and V. Megalooikonomou, 2005. Fast and effective characterization for classification and similarity searches of 2D and 3D spatial region data, *Pattern Recognition*, Vol. 38, pp. 1831-1846.
 - [50] J. Fong, S. Huang and H. Wonga, 2003. Continuous and incremental data mining association rules using frame metadata model, *Knowledge-Based Systems*, Vol. 16, pp. 91-100.
 - [51] Y. Chen, W. Ip and A. Lee, 2009. Mining frequent trajectory patterns in spatial-temporal databases, *Information Sciences*, Vol. 179, pp. 2218-2231.

125 AÑOS
1887 - 2012
Ingeniería con Trabajo y Rectitud



**UNIVERSIDAD
NACIONAL
DE COLOMBIA**
SEDE MEDELLÍN
FACULTAD DE MINAS