

Bases de Datos y Aplicación de Técnicas de Avanzadas

Edilma Olinda Gagliardi, Maria Gisela Dorzán, Maria Teresa Taranilla,

Pablo Rafael Palmero y Carlos Andrés Casanova

Departamento de Informática
Facultad de Ciencias Físico Matemáticas y Naturales
Universidad Nacional de San Luis, Argentina
{oli, mgdorzan, tarani, prpalmero, cacasanova}@unsl.edu.ar

Gregorio Hernández Peñalver

Departamento de Matemática Aplicada
Facultad de Informática
Universidad Politécnica de Madrid, España
gregorio@fi.upm.es

RESUMEN

En el marco del proyecto *Tecnologías Avanzadas de Bases de Datos*, la línea de investigación *Bases de Datos Espaciales y Espacio Temporales* está orientada a vincular las disciplinas Bases de Datos, Geometría Computacional y Metaheurísticas. El objetivo general consiste en utilizar técnicas, métodos y herramientas para investigación de base y en la resolución de problemas de optimización en diversos dominios de aplicación.

Palabras clave: Bases de Datos, Geometría Computacional, Metaheurísticas, Bases de Datos Espaciales y Espacio Temporales.

CONTEXTO

El proyecto *Tecnologías Avanzadas de Bases de Datos* desarrolla actividades vinculadas al tratamiento de objetos de diversos tipos, estructurados y no estructurados que son de utilidad en diversos campos de aplicación, tales como computación móvil, sistemas de información geográfica, computación gráfica, robótica, diseño asistido por computadora, motores de búsqueda en internet, entre otras.

En el proyecto coexisten tres líneas de investigación, orientadas al desarrollo de nuevos modelos para administrar y recuperar

información almacenada en repositorios de datos no estructurados, donde los escenarios de exploración requieren modelos tales como las bases de datos de texto, bases de datos espaciales, espacio temporales, bases de datos de imágenes, bases de datos de sonidos, espacios métricos, entre otros.

Algunas aplicaciones de bases de datos espaciales y espacio temporales necesitan guardar y consultar información actual e histórica acerca de posiciones referenciadas espacialmente y cambios de forma que tuvieron los objetos de estudio en diferentes escenarios a lo largo del tiempo, por lo que son necesarios estos modelos de bases de datos.

En este contexto, resulta apropiado vincular las disciplinas Bases de Datos, Geometría Computacional y Metaheurísticas, a los efectos de utilizar métodos y herramientas para la resolución de problemas en diversos dominios de aplicación y para la resolución de problemas de complejidad NP orientados a optimización.

Asimismo, se hace necesaria la incorporación de métodos científicos, procesos y sistemas de descubrimiento de información oculta en grandes cantidades de datos estructurados y no estructurados, que permitan obtener información valiosa mediante análisis

avanzados que traten diferentes tipos de datos, a los efectos de contribuir en la toma de decisiones en un sistema.

Por otra parte, con el avance de las tecnologías, situaciones del mundo real ameritan el uso de las mismas, en pos de una mejora sustancial en cuestiones de calidad, gestión, economía, etc. A nivel mundial, las Tecnologías de la Información y la Comunicación (TIC) continúan introduciendo cambios políticos, económicos, sociales y culturales, entre muchos otros, convirtiéndose en procesos clave para cualquiera que quiera accionar sobre el presente y proyectar hacia el futuro.

Por lo expuesto, en la línea de investigación se vinculan temáticas que surgen de las disciplinas Bases de Datos, Geometría Computacional y Metaheurísticas, a fin de tratar con la optimización en problemas NP, o bien de poder trabajar en dominios de aplicación con soluciones integradas que hagan uso de técnicas y herramientas de estas disciplinas.

El trabajo de investigación se desarrolla en forma conjunta con investigadores afines de proyectos de esta Universidad, convenios entre organizaciones nacionales y provinciales con presencia en San Luis, como así también de universidades extranjeras mediante convenios de cooperación interinstitucional.

1. INTRODUCCIÓN

La optimización es una línea de investigación en Ciencias de la Computación, donde se procura encontrar la mejor solución posible a un problema dentro de un período de tiempo limitado. En el caso particular de problemas de optimización combinatoria, los hay NP-duros y los hay polinómicos; y no se puede garantizar encontrar la mejor solución en un tiempo razonable, para todas las instancias del problema.

En Geometría Computacional, la optimización de configuraciones geométricas respecto de ciertos criterios de calidad, pertenecen a esta clase de problemas, y pueden resolverse

utilizando métodos de aproximación, tales como las técnicas metaheurísticas [1].

Una metaheurística es un proceso de generación iterativo que guía la búsqueda de soluciones combinando inteligentemente diferentes conceptos de campos diversos como inteligencia artificial, evolución biológica, inteligencia colectiva, sistemas inmunes, entre otros [11].

Un objetivo de la línea es obtener soluciones aproximadas para problemas geométricos, para los cuales aún no se han encontrado algoritmos eficientes que los solucionen debido a su complejidad, mediante la aplicación de técnicas metaheurísticas. Las propuestas de estudio están orientadas a la optimización de diferentes configuraciones geométricas, tales como triangulaciones, pseudo triangulaciones, poligonizaciones, cuadrangulaciones, entre otras. Los criterios de calidad considerados son peso, dilación, número de apuñalamiento, número de guardias en problemas de vigilancia, área, perímetro, entre otros. Estos criterios inducen a la búsqueda de soluciones óptimas respecto de ellos, mediante la aplicación de técnicas metaheurísticas.

Algunos de los problemas de optimización estudiados son la Triangulación de Peso Mínimo (Minimum Weight Triangulation, MWT) y la Pseudo-Triangulación de Peso Mínimo (Minimum Weight Pseudo-Triangulation, MWPT), problemas de carácter NP-duro [9] [10]. La Triangulación de Dilación Mínima (Minimum Dilation Triangulation, MDT) es otro problema estudiado, donde la dilación mide la calidad de conexión entre puntos de la triangulación. Para este problema no se conoce un algoritmo que lo resuelva en tiempo polinomial y tampoco se ha demostrado que sea NP-duro.

Otro tópico de interés actualmente en análisis, refiere a problemas de vigilancia. Los problemas de vigilancia pueden interpretarse, muchas veces, como problemas de iluminación o vigilancia. Se han planteado numerosas variantes del problema, cuestionándose ¿qué se vigila? y ¿cómo se vigila? De esta forma, se trata con diversos objetos geométricos a vigilar y con diversas formas de vigilancia. En

particular, nos interesan como objeto geométrico de estudio las triangulaciones planas las cuales constituyen un entorno geométrico adecuado, en el que también tienen sentido las preguntas sobre vigilancia. Un guardia situado en un vértice (ó arista) de la triangulación vigila todos los triángulos incidentes al vértice (ó arista). Por tanto, dada una triangulación T , podemos considerar las siguientes preguntas ¿cuántos guardias, ubicados en vértices (ó aristas), se necesitan para vigilar todos los triángulos de T ?, y ¿dónde se deben ubicar? es decir, cuáles son los vértices (ó aristas) seleccionados para ubicar los guardias.

La utilización de estas configuraciones geométricas optimizadas respecto de algún criterio de calidad, resulta como soporte de estrategias en la resolución de problemas vinculados con bases de datos espaciales y espacio temporales. En este contexto, se propone el estudio y el desarrollo de herramientas para la visualización de aplicaciones vinculadas a las bases de datos mencionadas.

2. LÍNEA DE INVESTIGACIÓN

En la línea se investigan diversos dominios de aplicación de bases de datos espaciales y espacio temporales, con uso de técnicas y herramientas de apoyo en la resolución de problemas.

Entre los tópicos de estudio se pueden mencionar:

- Diseño y aplicación de índices espacio temporales, en diversos escenarios de movimiento (redes, espacios libres de obstáculos, etc.),
- Aplicación de la Geometría Computacional y su marco disciplinar, a los efectos de considerar aspectos propios de los problemas involucrados.
- Optimización de estructuras geométricas relacionadas con las bases de datos mencionadas, mediante la aplicación de metaheurísticas para la optimización de

problemas NP-duros en Geometría Computacional.

- Desarrollo de herramientas para la visualización de estructuras geométricas y aplicaciones vinculadas con bases de datos espacio temporales.
- Los procesos de descubrimiento para la recuperación y análisis de información oculta en grandes volúmenes de datos.

Como objetivos específicos en la línea de investigación se propone:

- Desarrollo de aplicaciones de Bases de Datos Espaciales y Espacio Temporales, con aplicación de herramientas de Geometría Computacional y la incorporación de procesos de descubrimiento de la información, utilizando métodos que permitan obtener información mediante análisis avanzados. Específicamente, el desarrollo de una plataforma prototipo, soporte para diversos eventos y sistemas de información con dominio de aplicación en el sector agropecuario, mediante el uso de las TIC. Las características principales están orientadas a la posibilidad de recolección de datos de diferentes fuentes, su almacenamiento compartido, la integración progresiva de diversas funcionalidades, la explotación y la visualización de la información. Dicha plataforma se propone disponible en la web, con accesibilidad mediante tecnología móvil (I+D+i).
- Estudio de la indexación espacio temporal sobre objetos en movimiento para diversos escenarios. Desarrollo de las estructuras de almacenamiento, algoritmos de consulta y evaluación experimental.
- Estudio de configuraciones geométricas de puntos en el plano considerando medidas de calidad, mediante la aplicación de técnicas metaheurísticas y diversas estrategias algorítmicas.
- Desarrollo de herramientas para la visualización de aplicaciones vinculadas a las bases de datos mencionadas.

3. RESULTADOS OBTENIDOS /ESPERADOS

En la resolución de problemas de optimización de triangulaciones y pseudo triangulaciones de peso mínimo se utilizaron diferentes técnicas metaheurísticas: Optimización basada en Colonia de Hormigas (Ant Colony Optimization, ACO) y Recocido Simulado (Simulated Annealing, SA), técnicas determinísticas Voraces (Greedy) y Triangulación de Delaunay. Se realizó la evaluación experimental de las técnicas metaheurísticas mencionadas para la búsqueda de triangulaciones y pseudo triangulaciones que cumplan la propiedad. Se diseñaron generadores de instancias de problema para ser utilizados en la evaluación experimental. Los resultados obtenidos para los problemas MWT y MWPT con la técnica ACO fueron publicados en [3] [7] [5]. Los resultados obtenidos con la técnica Recocido Simulado en [2] [4].

Para el problema Triangulación de Dilación Mínima las técnicas aplicadas fueron: Greedy, Local Search, Iterated Local Search, Simulated Annealing y Random Local Search. Para cada estrategia se propuso un conjunto de operadores adecuados. Se utilizó Optimización de Parámetros Secuencial (Sequential Parameter Optimization - SPO) para el ajuste de los parámetros requeridos por Simulated Annealing. Se realizó un análisis experimental en el cual se compararon dichos algoritmos con otras técnicas. Se crearon las instancias de prueba, ya que para estos problemas no se encontraron disponibles ningún tipo benchmark con el cual comparar los resultados. Se desarrolló un estudio estadístico aplicando diferentes test estadísticos y métodos de visualización [6].

Los resultados del tratamiento de estos problemas han dado lugar a tesis de doctorado.

También, se estudió el problema de obtener poligonizaciones de un conjunto de puntos con mínima área para un conjunto de puntos en el plano [13].

Por otra parte, se desarrolló una herramienta para la generación y visualización de

triangulaciones, pseudo triangulaciones y poligonizaciones de conjuntos de puntos en el plano [12]. Además, se implementó una aplicación en el ámbito de la Salud para el seguimiento de focos epidémicos utilizando base de datos espacios temporales y herramientas de Geometría Computacional [8]. El desarrollo de estas herramientas se plasmó en trabajos finales de Licenciatura en Ciencias de la Computación y Licenciatura en Sistemas.

Por otra parte, con el fin de promover una plataforma de investigación, intercambio y desarrollo de tecnologías de la información y la comunicación (TIC), se integró el Proyecto Campo Conectado. Las organizaciones socias se proponen realizar acciones vinculadas a la promoción de las TIC en el ámbito de la producción agropecuaria con un horizonte inicial a dos años (2017-2018). El objetivo principal es aportar a la cooperación científico tecnológica y a las prácticas sociales, productivas y comerciales de los actores de la producción agropecuaria del semiárido central argentino.

En el marco de Campo Conectado, se propone accionar en la gestación de herramientas y desarrollos tecnológicos aplicados a la gestión de la producción agropecuaria en sistemas reales de producción. Actualmente, se está trabajando en el desarrollo de una herramienta para el seguimiento espacio temporal de rodeos en establecimientos agropecuarios, mediante el uso de las TIC y de las Bases de Datos Espacio Temporales, con herramientas de Geometría Computacional.

4. FORMACIÓN DE RECURSOS HUMANOS

La formación del grupo de trabajo en la Universidad Nacional de San Luis, se consolida con actividades de cooperación mutua e intercambio recíproco de información científica, tecnología y desarrollo de nuevos conocimientos con investigadores locales y de otras universidades.

Entre las actividades más destacadas, se mencionan:

- i) Formación de recursos humanos plasmada en tesis doctorales, tesis de maestría y Licenciados en Ciencias de la Computación.
- ii) Realización de pasantías de investigación con docentes de otras universidades.
- iii) Actividades de formación académica, a través del dictado y realización de cursos de posgrado y de especialización
- iv) Actividades de divulgación científica, conferencias y publicaciones en congresos y revistas en el ámbito nacional e internacional.
- v) Integración del proyecto interinstitucional Campo Conectado.

La línea tiene como un objetivo continuar con las actividades integradoras relacionadas al presente proyecto, proponiendo actividades de formación académica, de formación de recursos humanos locales y de otras universidades nacionales, investigación, desarrollo, y otras actividades académico-científicas vinculantes.

5. BIBLIOGRAFÍA

- [1] de Berg, M., Cheong O., van Kreveld, M., Overmars, M., *Computational Geometry: Algorithms and Applications*. 3rd edition, Springer-Verlag, Heidelberg, 2008.
- [2] Dorzán M.G., Gagliardi E.O., Leguizamón M.G., Hernández Peñalver G. *Approaches for MWT and MWPT Problems*. XVII Congreso Argentino de Ciencias de la Computación 2011 (CACIC 2011), 2011
- [3] Dorzán M.G., Gagliardi E.O., Leguizamón M.G., Hernández Peñalver G. *Using ACO metaheuristic for MWT problem*. XXX International Conference of the Chilean Computer Science Society. ISBN 978-0-7695-4689-6. Chile. 2011
- [4] Dorzán M.G., Gagliardi E.O., Leguizamón M.G., Hernández Peñalver G. *Metaheuristic approaches for MWT and MWPT Problems*. XIV Encuentros de Geometría Computacional. Pág: 79-82. 2011
- [5] Dorzán M.G., Gagliardi E.O., Leguizamón M.G., Hernández Peñalver G. *Approximations on Minimum Weight Triangulations and Minimum Weight Pseudo-Triangulations using Ant Colony Optimization Metaheuristic*. *Fundamenta Informaticae*. ISSN: 0169-2968 (Print), 1875-8681 (Online). Volume 119, number 1, Pag: 1-27.
- [6] Dorzán, M. G; Leguizamón, M. G.; Mezura-Montes, E.; Hernández Peñalver G. *Approximated algorithms for the Minimum Dilation Triangulation Problem*. *Journal of Heuristics*. DOI 10.1007/s10732-014-9237-2. Print ISSN 1381-1231. Online ISSN 1572-9397. Publisher Springer US. 2014.
- [7] Gagliardi E. O., Dorzán M. G., Leguizamón M. G. y Hernández Peñalver. G.; *Approximations on Minimum Weight Pseudo-Triangulation problem using Ant Colony Optimization*. XXX International Conference of the Chilean Computer Science Society. ISBN 978-0-7695-4689-6. Chile. 2011
- [8] Guasch, M.M; Piergallini, M.R; Dorzán, M.G.; Gagliardi, E.O.; Taranilla, M.T.; *“Una herramienta para el análisis y seguimiento de focos epidémicos”* en Anales del 17º Concurso de Trabajos Estudiantiles en 43 JAIO. Pág.: 35-45 Universidad de Palermo, Argentina, 2014.
- [9] Gudmundsson J., Levcopoulos C.; *Minimum weight pseudo-triangulations*. *Computational Geometry. Theory and applications*. Elsevier Vol. 38- Pages 139-153, 2007.
- [10] Mulzer W., Rote G. *Minimum weight triangulation is NP-hard*. In *Proceedings of the 22nd Annual ACM Symposium on Computational Geometry*. 2006.
- [11] Michalewicz Z., Fogel D., *How to Solve It: Modern Heuristics*, Springer, 2004.
- [12] Palmero, P.R., Dorzán, M. G., Gagliardi E.O., *Una Herramienta para la Manipulación de Configuraciones Geométricas*, 42º Jornadas Argentinas de Informática e Investigación Operativa - 16º Concurso de trabajos de fin de carrera. Universidad Nacional de Córdoba, Argentina. 2013. ISSN: 1850-2946.
- [13] Taranilla, M. T.; Gagliardi, E. O.; Hernández Peñalver, G. *Approaching Minimum Area Polygonization*. Congreso Argentino de Ciencias de la Computación (CACIC 2011), Pág.: 31-40, 2011.