

Geometría Computacional y Bases de Datos

Maria Gisela Dorzán, Susana Esquivel, Edilma Olinda Gagliardi,
Mario Guillermo Leguizamón, Pablo Palmero y Maria Teresa Taranilla

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

Gregorio Hernández Peñalver

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

Resumen

La línea de investigación *Geometría Computacional y Bases de Datos* del proyecto *Tecnologías Avanzadas de Bases de Datos*, vincula el estudio de las disciplinas Bases de Datos, Geometría Computacional y Metaheurísticas, utilizando los métodos y las herramientas provistas para la resolución de problemas orientados a optimización.

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

Contexto

El proyecto *Tecnologías Avanzadas de Bases de Datos* de la Universidad Nacional de San Luis, tiene como objetivo principal el estudio de bases de datos avanzadas, en donde se involucra el diseño y desarrollo de herramientas para administrar eficientemente datos no estructurados.

En diversas aplicaciones del campo de las Ciencias de la Computación son necesarios repositorios de datos no tradicionales, y aparecen nuevos modelos de bases de datos para buscar y administrar información en ellos. Como los problemas relacionados a la temática han aparecido en áreas muy diversas, las

soluciones también han surgido desde campos no relacionados. Algunos ejemplos son bases de datos de imágenes, huellas digitales o clips de audio, las cuales almacenan datos que son difíciles de estructurar para adecuarlos al concepto tradicional de búsqueda.

En este sentido, en el Proyecto *Tecnologías Avanzadas de Bases de Datos*, se desarrollan actividades vinculadas al tratamiento de objetos de diversos tipos, estructurados y no estructurados que son de utilidad en diversos campos de aplicación, por ejemplo, robótica, computación gráfica, visión artificial, sistemas de información geográfica, diseño asistido por computadora, computación móvil, motores de búsqueda en internet, entre otras, y que se relacionan en tales bases de datos.

En el proyecto existen 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.

Así, surge el estudio de modelos como las bases de datos espaciales y bases de datos espacio-temporales. También, se requiere la

construcción y manipulación de diferentes objetos y estructuras geométricas de utilidad en diversas áreas de aplicación.

En particular, las estructuras geométricas que se estudian deben cumplir con propiedades deseables, y algunos de los problemas relacionados con la optimización de las mismas son problemas NP-duros, por tanto en la búsqueda de soluciones aproximadas utilizamos técnicas metaheurísticas. Por lo expuesto, en la línea de investigación *Geometría Computacional y Bases de Datos* hemos vinculado temáticas que surgen de las disciplinas Bases de Datos, Geometría Computacional y Metaheurísticas.

El trabajo de investigación se desarrolla en forma conjunta con investigadores afines de proyectos locales y de universidades extranjeras (Universidad Politécnica de Madrid - España, Universidad Veracruzana - México, Universidad del Bío Bío- Chile, entre otras) mediante convenios de cooperación institucional. En este trabajo, presentamos los temas en estudio, junto con las propuestas más recientes de interés.

Introducción

La optimización es una línea de investigación en Ciencias de la Computación, donde se requiere 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, no escapan a esta clase de problemas, y pueden resolverse utilizando métodos de aproximación, tales como las técnicas metaheurísticas [BCKO08].

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 [MF04].

El objetivo principal de esta línea es proponer soluciones aproximadas para problemas de tipo geométrico para los cuales aún no se han encontrado algoritmos eficientes que los solucionen debido a su complejidad, mediante la aplicación de técnicas metaheurísticas.

En este sentido, las propuestas de estudio están orientadas a la optimización de configuraciones geométricas, tales como triangulaciones, pseudotriangulaciones, poligonizaciones, entre otras. Se consideran peso, dilación, perímetro, área, número de apuñalamiento mínimo, número de guardias en problemas de vigilancia, como ejemplos de criterios de calidad. Estos criterios nos inducen a buscar soluciones óptimas respecto de ellos, mediante la aplicación de técnicas metaheurísticas.

Entre los problemas de optimización estudiados están: la Triangulación de Peso Mínimo (*Minimum Weight Triangulation, MWT*) y la Pseudo-Triangulación de Peso Mínimo (*Minimum Weight Pseudo-Triangulation, MWPT*). Dichos problemas minimizan la suma de las longitudes de las aristas proveyendo así una medida de calidad para determinar cuán “buena” es la estructura. La complejidad del cálculo de MWT fue uno de los problemas abiertos más estudiados en Geometría Computacional hasta que Mulzer y Rote demostraron que la construcción de MWT es un problema NP-duro [MR06]. Por otra parte, la complejidad de la pseudo-triangulación de peso mínimo aún no está resuelta. Levcopoulos y Gudmundsson muestran una 12-aproximación de una pseudo-triangulación que puede ser calculada con complejidad de $O(n^3)$ [GL07]. Los autores dan una aproximación de $O(\log n w(\text{MST}))$ de una pseudo-triangulación de peso mínimo en un tiempo de $O(n \log n)$ donde $w(\text{MST})$ es el peso del Árbol de Expansión Mínimo (*Minimum Spanning Tree, MST*) el cual es un subconjunto de la estructura obtenida.

Otro problema estudiado es la Triangulación de Dilación Mínima (*Minimum Dilation Triangulation, MDT*) donde la dilación mide la

calidad de conexión entre puntos de la triangulación. La dilación entre un par de puntos, u y v , de una triangulación T se define como la razón entre el camino más corto entre u y v en T y la distancia euclídea entre dicho par de puntos. La máxima dilación entre todos los pares de puntos en T se llama la dilación de T ($\Delta(T)$). La mejor posible dilación de cualquier triangulación de un conjunto de puntos S se conoce como dilación de S ($\Delta(S)$). La complejidad del problema de encontrar $\Delta(S)$ para un conjunto de puntos S se desconoce y por lo tanto, nos centramos en el desarrollo de algoritmos aproximados para encontrar triangulaciones de alta calidad de dilación mínima.

Con respecto a poligonizaciones, estudiamos el problema de obtener poligonizaciones de área mínima/máxima para un conjunto de puntos en el plano. Este problema es NP-duro, según demostró Fekete [Fe00].

La utilización de estas configuraciones geométricas optimizadas respecto de algún criterio de calidad, resultan como soporte de estrategias en la resolución de problemas vinculados al área disciplinar de 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.

Línea de investigación

En la línea de investigación estudiamos el diseño y desarrollo de índices espacio-temporales, aplicables a diversos escenarios de movimiento (redes, espacios libres de obstáculos, etc.), considerando la geometría como una disciplina marco en la cual se formalizan aspectos propios de los problemas involucrados. En este contexto, nos proponemos el estudio de optimizaciones de estructuras geométricas que están relacionadas con las bases de datos ya mencionadas y al desarrollo de herramientas para la visualización de estructuras geométricas y aplicaciones vinculadas a las bases de datos espacio-temporales.

Como objetivos específicos de estudio en la línea de investigación podemos enumerar los siguientes:

i) Indexación espacio-temporal sobre objetos en movimiento para diversos escenarios. Desarrollo de las estructuras de almacenamiento, los algoritmos de consulta y la evaluación experimental, mostrando el desempeño de los distintos índices en aplicaciones de diferentes magnitudes respecto de la población de objetos en movimiento. Desarrollo de aplicaciones con herramientas de Geometría Computacional y Bases de Datos Espacio-Temporales. Desarrollo de una herramienta integral en el ámbito de la Salud para el seguimiento de focos epidémicos utilizando base de datos espacios-temporales.

ii) Estudio de configuraciones geométricas generales de puntos en el plano, que cumplan medidas de calidad como: peso, dilación, área, entre otras. Optimización de Triangulaciones, Pseudotriangulaciones y Poligonizaciones considerando medidas de calidad mínimas o máximas, aplicando técnicas metaheurísticas y diversas estrategias algorítmicas.

iii) Diseño y desarrollo de herramientas para la generación, visualización y manipulación de diferentes configuraciones geométricas de conjuntos de puntos en el plano.

Resultados Obtenidos

Para la resolución de problemas de optimización de Triangulaciones y Pseudotriangulaciones de Peso Mínimo aplicamos las 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 llevó a cabo el estudio, adecuación y evaluación experimental de las técnicas metaheurísticas mencionadas para la búsqueda de triangulaciones y pseudotriangulaciones que cumplan la propiedad. Los resultados obtenidos para los problemas MWT y MWPT utilizando la técnica metaheurística ACO fueron publicados en [DGLH11b] [GDLH11] [DGLH12]. Se llevó a cabo una

evaluación experimental y análisis de los resultados obtenidos con la técnica Recocido Simulado [DGLH11a] [DGLH11c].

Para el problema Triangulación de Dilación Mínima (Minimum Dilation Triangulation - MDT) todavía no se conoce un algoritmo que lo resuelva en tiempo polinomial y tampoco se ha demostrado que sea NP-duro. Para este problema se estudiaron, desarrollaron y aplicaron diferentes tipos de técnicas: Greedy, Local Search, Iterated Local Search, Simulated Annealing y Random Local Search proponiendo para cada estrategia un conjunto de operadores adecuados. Debido a la complejidad que implica la puesta a punto de los parámetros de técnicas metaheurísticas, 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, como por ejemplo, Delaunay. Se crearon las instancias de prueba, ya que para estos problemas no se encontraron disponibles ningún tipo benchmark con el cual comparar nuestros resultados. Las conclusiones fueron afirmadas desarrollando un estudio estadístico aplicando diferentes test estadísticos y métodos de visualización [DLMH14].

Los resultados obtenidos en el tratamiento de estos problemas han dado lugar a dos tesis de doctorado, defendidas en el presente año.

Por otra parte, se ha desarrollado una herramienta para la generación y visualización de triangulaciones pseudo-triangulaciones y poligonizaciones de conjuntos de puntos en el plano [DGPH10] [PDG13].

Además, se finalizó con la implementación de 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. Como trabajo futuro, específicamente para problemas sobre configuraciones geométricas, continuaremos con el estudio de los problemas de optimización considerando otras técnicas metaheurísticas, que puedan adecuarse para su resolución. Se comenzará con el estudio de

problemas relacionados con configuraciones geométricas, tales como cuadrangulaciones, pseudocuadrangulaciones. Se llevarán a cabo, la evaluación de las distintas técnicas para determinar su funcionalidad e impacto en la comunidad científica y el análisis que incluirá el correspondiente tratamiento estadístico, y comparativo con otros algoritmos de tipo aproximado o de tipo exacto.

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, período 2010-2013, se mencionan:

- i) Formación de recursos humanos plasmada dos Tesis Doctorales (UNSL), tres Licenciados en Ciencias de la Computación (UNSL), dos Licenciados en Sistemas (UNNOBA).
- ii) Obtención de becas de investigación tipo I y tipo II de CONICET.
- iii) realización y dirección de pasantías de investigación con docentes de otras universidades (UPM, UV, UNNOBA, entre otras)
- iv) actividades de formación académica, cursos de posgrado y de especialización,
- v) actividades de divulgación científica, conferencias y publicaciones en congresos y revistas en el ámbito nacional e internacional.

Nos proponemos 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.

Referencias

- [BFM97] Bäck T., Fogel D., Michalewicz Z. *Handbook of Evolutionary Computation*. IOP Publishing Ltd and Oxford University Press. 1997.

- [BCKO08] de Berg, M., Cheong O., van Kreveld, M., Overmars, M., *Computational Geometry: Algorithms and Applications*. 3rd edition, Springer-Verlag, Heidelberg, 2008.
- [DGPH10] Dorzán M.G., Gagliardi, E. O., Palmero P.R., Hernández Peñalver, G. *Una herramienta para la generación y visualización de Triangulaciones y Pseudotriangulaciones*. XVI Congreso Argentino de Ciencias de la Computación 2010 (CACIC 2010), 2010
- [DGLH11a] 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
- [DGLH11b] 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
- [DGLH11c] 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
- [DGLH12] 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, pp 1-27.
- [DLMH14] M. G. Dorzán, M. G. Leguizamón, Efrén Mezura-Montes, G. Hernández Peñalver *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. Enero 2014.
- [GDLH11] 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
- [GL07] Gudmundsson J., Levkopoulos C.; *Minimum weight pseudo-triangulations*. Computational Geometry. Theory and applications. Elsevier Vol. 38-pages 139-153, 2007.
- [MR06] Mulzer W., Rote G. *Minimum weight triangulation is NP-hard*. In Proceedings of the 22nd Annual ACM Symposium on Computational Geometry. 2006.
- [MF04] Michalewicz Z., Fogel D., *How to Solve It: Modern Heuristics*, Springer, 2004.
- [PDG13] Palmero, P.R., Dorzán, M. G., Gagliardi E.O., *Una Herramienta para la Manipulación de Configuraciones Geométricas*, 42° Jornadas Argentinas de Informática e Investigación Operativa - 16° Concurso de trabajos de fin de carrera. Universidad Nacional de Córdoba, Córdoba, Argentina. 2013. ISSN: 1850-2946.