Separabilidad Geométrica aplicada a las Búsquedas por Rangos con uso de Metaheurísticas

Edilma Olinda Gagliardi
Departamento de Informática
Facultad de Ciencias Físico, Matemáticas y Naturales
Universidad Nacional de San Luis, Argentina
oli@unsl.edu.ar
Fax: 54-2652-430224

Gregorio Hernández Peñalver
Departamento de Matemática Aplicada
Facultad de Informática
Universidad Politécnica de Madrid, España
gregorio@fi.upm.es
Fax: 34-91-3367426

Resumen

Para trabajo de investigación presentamos los aspectos teóricos y prácticos relevantes para las búsquedas por rangos en espacios de búsquedas geométricamente separables, con la propuesta de utilizar de herramientas no tradicionales como las metaheurísticas.

Palabras Claves: Búsquedas por rangos, Separabilidad geométrica, Metaheurísticas.

1. Introducción

El problema de las Búsquedas por Rangos en Bases de Datos puede ser tratado en el marco de la Geometría Computacional, dado que nos brinda nuevas posibilidades de diseño de algoritmos y estructuras de datos utilizando herramientas geométricas [AE98] [Aga97] [AM94] [BF79] [BY98] [Mat94] [Mel84].

Una de tales posibilidades es la aplicación de Separabilidad Geométrica que nos permite proponer nuevos algoritmos de partición del espacio de búsqueda por medio de la aplicación de diversos criterios de separabilidad (rectas, cuñas, bandas, entre otras) [AHMS00] [AHMS01] [DHMS01] [HMRS99] [HNRS01] [HNRS98] [HSS03] [Sea02]. La aplicación de tales separadores geométricos debe realizarse con relación a la instancia particular del espacio de búsqueda, lo cual hace presuponer, o, conocimientos de tal configuración, o bien, el uso de herramientas no tradicionales que ayuden a la selección de tales criterios de separabilidad.

Por tanto, en este trabajo presentamos una vinculación entre las búsquedas por rangos y la separabilidad geométrica, mostrando aspectos teóricos relevantes y proponiendo nuevas ideas para el diseño de estructuras de datos para las búsquedas por rangos en estado de investigación y desarrollo; y como complemento, proponemos trabajar con metaheurísticas para lograr selecciones adecuadas de separadores geométricos.

Primeramente, introducimos nociones relacionadas a la separabilidad geométrica y a las búsquedas por rangos, proponiendo una vinculación entre ambas temáticas. Posteriormente, brindamos una propuesta de investigación donde, explícitamente, para esta vinculación proponemos la utilización de metaheurísticas para la obtención de resultados satisfactorios en la elección de los separadores geométricos.

2. Separabilidad Geométrica

Innovando en nuevos métodos de partición, incorporamos las nociones separabilidad geométrica como posibles criterios de separación de objetos en el espacio. Los trabajos de separabilidad están orientados a dos o más conjuntos disjuntos de objetos geométricos, básicamente para puntos, bajo diversos criterios de separabilidad (bandas, cuñas, sectores, entre otros). Estos criterios de

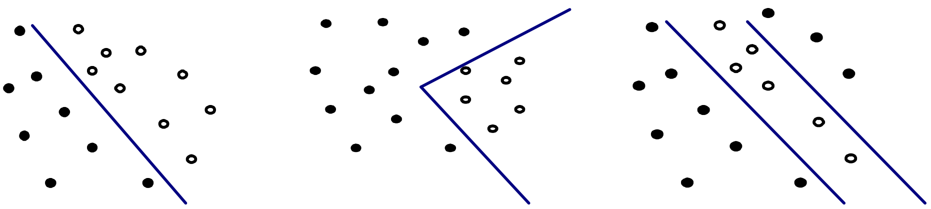
Proyecto Tecnologías Avanzadas de Bases de Datos 22/F314, Departamento de Informática, UNSL.
Proyecto AL2002-1010-2.43/ AL2003-1010-2.55/ AL2004-1010-2.53. Geometría Computacional, UPM.

separabilidad tienen aplicaciones interesantes, como por ejemplo el análisis de imágenes, clasificación de datos, etc. Siempre que sea necesario discriminar objetos en un espacio de trabajo, por algún atributo del mismo, los criterios de separabilidad juegan un papel importante.

Para modelar este problema, se dan dos conjuntos disjuntos P y Q de objetos en el plano, los que se referirán como objetos *rojos* y *azules*, respectivamente. Eventualmente, los objetos pueden ser: puntos, segmentos, polígonos o círculos. En cualquiera de los casos, n es el tamaño de la entrada. Entonces, dada una familia C de curvas en el plano, decimos que los conjuntos P y Q son *C-separables* si existe una curva $C_i \in C$, llamada *separador*, tal que cada componente conexa de R^2-C_i contiene solamente objetos, o bien de P, o bien de Q. Se dice que un conjunto finito S de superficies es un *separador* de conjuntos de objetos en R^d si las componentes conexas de R^d-S contienen objetos de un solo conjunto.

Si C es la familia de rectas en el plano, tenemos la *separabilidad lineal*. Dos conjuntos P y Q son *linealmente separables* si y sólo si sus cierres convexos no se intersecan. Dos conjuntos disjuntos P y Q de objetos en el plano son *separables por una cuña* si existe una cuña que contiene solamente todos los objetos de uno de los conjuntos. Aquí se estudia el problema de decidir la separabilidad por cuñas calculando la ubicación de los posibles ápices de cuñas; como una extensión al problema, se estudia hallar la cuña de ángulo mínimo (máximo). Dos conjuntos disjuntos P y Q de objetos en el plano son *separables por banda* si existe una banda, determinada por dos rectas paralelas, que contiene solamente todos los objetos de uno de los conjuntos. Aquí se estudia el problema de decidir la separabilidad por bandas calculando el conjunto de intervalos de pendientes de bandas separadoras; como una extensión al problema, se estudia hallar la banda mínima (máxima).

Los algoritmos para resolver los problemas de decisión y optimización propuestos anteriormente se ejecutan en tiempo $O(n \log n)$; salvo en el caso de la separación lineal, que toma tiempo $O(n)$ [Sea02].



Separabilidad líneal Separabilidad por Cuña Separabilidad por Banda

Figura 1

Los principales resultados sobre separabilidad con bandas y cuñas, con diferentes versiones sobre combinaciones y propiedades de ellas, son algoritmos eficientes para decidir la existencia de los separadores y calcular soluciones factibles. Algunos de estos resultados pueden ser extendidos en el plano al uso de otros objetos geométricos que no sean puntos y al espacio tridimensional [HSS03] en donde sólo existen algunos algoritmos para resolver algunos de los problemas vistos.

En este marco, también queda abierto el estudio sobre nuevos criterios de separabilidad. Es decir, no sólo usar planos o semiplanos, sino otros objetos como pirámides dobles (bipirámides), cilindros, cónicas, cónicas dobles, etc. En todos los casos resueltos, la complejidad crece polinomialmente respecto del tamaño de entrada, por lo cual los algoritmos propuestos caen en la clase de los algoritmos polinomiales.

3. Búsquedas por Rangos

Trasladándonos a las búsquedas por rangos, comenzamos por presentar algunos aspectos teóricos. Así, se define

- Un *Espacio de Rangos* como un sistema Ω = (U, F), donde U es un conjunto de objetos geométricos y F es una familia de subconjuntos de U. Los elementos de F son llamados *Rangos* de Ω.

  Por ejemplo: $\Omega_1 = (\mathbf{R^d}, \{ h/ h$ es un semiespacio en $\mathbf{R^d} \})$; $\Omega_2 = (\mathbf{R^d}, \{ h/ h$ es una bola en $\mathbf{R^d} \})$, etc.

- Dados un espacio de rangos Ω = (U, F), S un subconjunto de objetos de U y R un rango de F, consultar los objetos geométricos que están en $S \cap R$. En este caso, a R suele llamárselo *rango de consulta* (*query range*).

El problema real para el cual tiene sentido este estudio, es que el conjunto de objetos geométricos, S, es dado con anterioridad y luego, repetidas veces, variando el rango R, se consulta por la intersección. Entonces, no es suficiente con diseñar un buen algoritmo sino que también se debe diseñar una estructura de datos apropiada para almacenar el conjunto S, de forma tal que cada consulta por un rango R, pueda ser respondida eficientemente.

La mayoría de las estructuras de datos para búsquedas o consultas por rango, son construidas de forma recursiva, dividiendo el espacio de objetos geométricos en varias regiones, con propiedades geométricas deseables sobre ellas. La búsqueda por rangos, esencialmente, consiste en buscar los objetos geométricos que contiene una determinada región (rango) del espacio.

Los análisis de las complejidades tanto en tiempo de construcción de la estructura de datos, en espacio que ocupa y en tiempos de respuestas de las consultas, depende de la dimensión del espacio, de la cardinalidad de *S*, de la estrategia de partición de *S* y del tamaño de la respuesta de la consulta.

4. La vinculación y el uso de metaheurísticas

Como hemos visto acerca de separabilidad geométrica, tenemos conjuntos disjuntos de objetos en el plano, los cuales queremos separar acorde a algún criterio de separabilidad. Los conjuntos son disjuntos debido a alguna propiedad específica de los objetos y de su posición en el espacio. Nos interesan particularmente las descripciones de las curvas que determinan las regiones que contienen tales conjuntos disjuntos, dado que ellas constituyen los separadores geométricos. Con esta información podemos obtener un esquema de partición, en cuyo caso estamos en condiciones de crear una estructura de datos apropiada para acceder a los puntos. En la figura 2, podemos visualizar estas ideas:

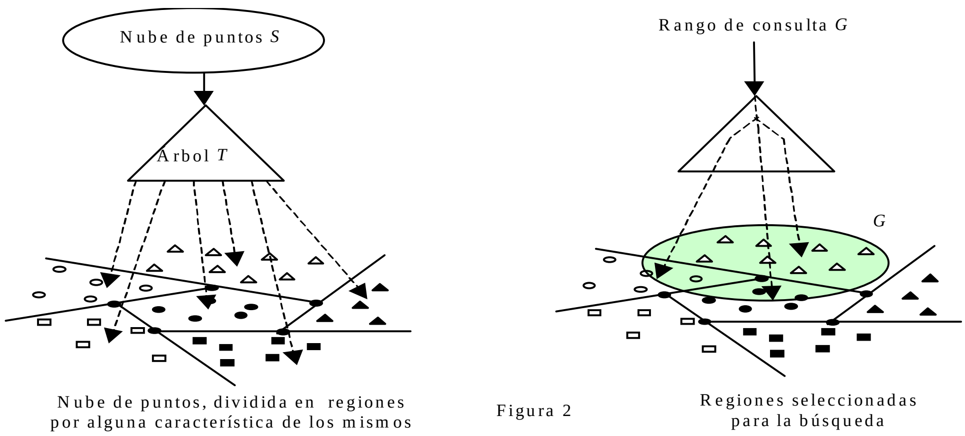


Figura 2

Las consultas se pueden realizar en función de las características dadas de los puntos, o bien, por rangos. En el primer caso, por ejemplo, podemos solicitar las regiones correspondientes a los puntos rojos, pudiendo obtener como respuesta una o varias regiones; o también, consultar cuáles son los diferentes colores existentes. En caso de los rangos, podemos utilizar un rango de consulta que abarque parte de una región o varias regiones. Por ejemplo, un rango de consulta dado por un semiplano *H*, definido por una recta *h*, puede corresponderse con partes de regiones, con regiones completas, y posiblemente (es deseable), dejar excluidas regiones íntegras.

Para concluir con estas ideas, hemos creado un *árbol de partición*, obtenido por medio de la aplicación de diversos criterios de separabilidad geométrica. Nos planteamos qué cosas podrían continuar haciéndose, y vemos que se abre un abanico de posibilidades. Por cada hijo de la raíz podemos analizar qué posibilidades de separación existen y basándose en ello, volvemos a aplicar algunas de las variantes vistas previamente. Repitiendo proceso con cada hijo, estamos en condiciones de conformar los diversos niveles del árbol. Observemos que la clasificación en cada nivel debe realizarse en función de características diferentes de la nube de puntos [GH02a] [GH02b] [GH03a] [GH03b].

Considerar cuáles son los criterios de separabilidad más propicios para la nube de puntos resulta una incógnita a resolver y que dependerá, en la mayoría de los casos, de la instancia de la nube de puntos a tratar. Así, encontrar la secuencia óptima (o aceptable) de separadores geométricos puede resultar un problema difícil, lo cual nos puede llevar a la aplicación de distintos tipos de búsquedas en un espacio considerable de soluciones. Para resolver este problema proponemos el uso de metaheurísticas.

Entonces, supongamos que llamamos *ES* a nuestro espacio de soluciones, conformado por todas las secuencias posibles de separadores de cualquier longitud. Nuestro objetivo consiste en hallar un $x \in ES$ que resulte adecuado; es decir, que represente una secuencia de separadores geométricos adecuada respecto de alguna métrica en particular como por ejemplo longitud de la secuencia, complejidad final de la secuencia, entre otras. Ello permitiría obtener una partición del espacio de rangos que permita crear, finalmente, el árbol de partición como estructura de datos apropiada para acceder a la nube de puntos.

De acuerdo a lo anterior, podemos ver este problema como de optimización combinatoria, donde se pueden tratar los siguientes problemas computacionales: i) caracterización del espacio de búsqueda en cuanto a la posible existencia de soluciones no factibles; ii) optimización: encontrar una solución óptima; iii) evaluación del costo de las soluciones consideradas.

Como métodos de búsqueda en este espacio de soluciones, se pueden considerar diferentes clases. Nosotros en particular proponemos las metaheurísticas, dado que permiten resolver problemas de optimización de interés práctico y, además, estos enfoques incluyen muchas variaciones, basadas en enfoques de la naturaleza misma, otros en evolución biológica, neurofisiología, y comportamientos biológicos, entre otros. Algunos ejemplos de estos son: *simulated annealing* (SA), *tabu search* (TS), *algoritmos evolutivos* (AE), *optimización basada en colonias de hormigas* (OCH), etc.

5. Conclusiones y visión de futuro

Este trabajo de investigación constituye un proyecto de tesis de postgrado y se enmarca dentro de la línea de investigación Geometría Computacional y Bases de Datos, perteneciente al Proyecto Tecnologías Avanzadas de Bases de Datos 22/F314, Departamento de Informática, UNSL. Además, el grupo de trabajo en Geometría Computacional de la UNSL, mantiene un proyecto de investigación conjunto con la UPM, Proyecto AL2002-1010-2.43/ AL2003-1010-2.55/ AL2004-1010-2.53. Geometría Computacional, con el objetivo principal de consolidar la línea de trabajo en la UNSL, aportando nuevos enfoques y técnicas algorítmicas a las líneas de investigación ya establecidas en su Departamento de Informática. Asimismo, dentro del Departamento de

Informática se está trabajando con el apoyo de integrantes del Grupo LIDIC, de la UNSL, quienes tienen experiencia en el desarrollo y trabajo de Metaheurísticas.

6. Referencias bibliográficas

[AE98] Agarwal, P.; Erickson, J. *Geometric range searching and its relatives*; Advances in Discrete and Comput. Geom. (B. Chazelle,J. Goodman, and R. Pollack, eds.), American Mathematical Society, Providence, 1998.

[Aga97] Agarwal, P.; *Range searching* CRC Handbook of Computational Geometry (J. Goodmand and J. O'Rourke, eds.).

[AHMS00] Arkin, E.M.; Hurtado, F.; Mitchell,S.B.; Seara, C.; Skiena, S.S.; *Some Separability Problems in the Plane*; 16th European Workshop on Computational Geometry, Eilat, Israel, 2000.

[AHMS01] Arkin, E.M.; Hurtado, F.; Mitchell,S.B.; Seara, C.; *Some Lower Bounds on Geometric Separability Problems*; 11th Fall Workshop on Computational Geometry, Polytechnic University, Brooklyn, NY, 2001.

[AM94] Agarwal, P.; Matousek,J.; *On range searching with semialgebraic sets;* Discrete Comput. Geom., 11: 393-418, 1994.

[BF79] Bentley, J.L.; Friedman, J.H; *Data Structures for range searching*; ACM Comput. Surv. 11:397-409;1979.

[BY98] Boissonnat, J.D.; Yvinec, M. *Algorithmic Geometry*, Cambridge University Press, 1998.

[DHMS01] Devillers, O; Hurtado, F.; Mitchell,S.B.; Seara, C.; *Separating Several Point Sets in the Plane*; 13th Canadian Conference on Computational Geometry, 2001.

[GH02a] Gagliardi, E.O.; Hernández Peñalver, G.; *Geometría Computacional y Bases de datos: Búsquedas por Rangos y Separabilidad Geométrica.* I Workshop de Bases de Datos, Jornadas Chilenas de Ciencias de la Computación, Copiapó, Chile, 4-9 de Noviembre de 2002.

[GH02b] Gagliardi, E. O.; Hernández Peñalver, G.; *Un enfoque propuesto para las Búsquedas por Rangos con Separabilidad Geométrica*. VIII Congreso Argentino de Ciencias de la Computación 2002 (CACIC 2002). Capital Federal, 15-18 de Octubre de 2002.

[GH03a] Gagliardi, E.O.; Hernández Peñalver, G.; *Las búsquedas por rangos en espacios de búsquedas geométricamente separables.* II Workshop de Bases de Datos, Jornadas Chilenas de Ciencias de la Computación, Chillán, Chile, 3-7 de Noviembre de 2003.

[GH03b] Gagliardi, E. O.; Hernández Peñalver, G.; *Separabilidad Geométrica aplicada a las Búsquedas por Rangos*; Workshop de Investigadores en Ciencias de la Computación 2003, WICC 2003. Tándil, Buenos Aires, 22-23 de Mayo de 2003.

[HMRS99] Hurtado, F.; Mora, M.; Ramos, P.A.; Seara, C.; *Separación de Objetos en el Plano por Doble Cuña y por Θ-Poligonal*; VIII Encuentros de Geometría Computacional, Castelló, España, 1999.

[HNRS01] Hurtado, F.; Noy, M.; Ramos, P.A.; Seara, C.; *Separating Objects in the Plane with Wedges and Strips*; Discrete Applied mathematics, Vol 109, pp 109-138, 2001.

[HNRS98] Hurtado, F.; Noy, M.; Ramos, P.A.; Seara, C.; *Separating Objects in the Plane with Wedges and Strips*; 110th Canadian Conference on computational Geometry, Montreal, Canadá, 1998.

[HSS03] Hurtado, F.; Seara, C.; Sethia, S.; *Red-Blue separability problems ein 3D;* to appear in the 19th European Workshop on Computational Geometry, Bonn, Germany, March 24-26, 2003.

[Mat94] Matousek,J.; *Geometric Range searching*; ACM Comput. Survey, 26:421-461;1994.

[Mel84] Melhorn, K. *Multidimensional Searching and Computational Geometry* Springer Verlag 1984.

[Sea02] Seara, C; *On geometric separability*; Tesis Doctoral 2002, Univ. Politécnica de Cataluña.