Búsqueda por Rangos:
Geometría Computacional y Bases de Datos

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

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

Introducción

La Geometría Computacional se ocupa de resolver problemas geométricos de modo constructivo. Se interesa por demostrar la existencia de la solución de un problema y por encontrar los algoritmos y estructuras de datos eficientes, medidos respecto de su complejidad temporal y espacial respectivamente. Por lo tanto, podemos decir que esta disciplina forma parte de la teoría del diseño y análisis de algoritmos y estructuras de datos.

En ocasiones, la Geometría brinda soluciones más eficientes en problemas que no parecen geométricos. Descubrir que los datos de un problema verifican propiedades geométricas sirve para poder aplicar alguna técnica algorítmica o alguna estructura de datos especial, que nos permite describir una solución óptima.

Respecto de las bases de datos, un problema que se presenta a menudo es el estudio de los rangos y las consultas por rangos, denominado *Búsqueda por Rangos*. Este problema tratado desde una perspectiva geométrica nos permite diseñar y analizar los algoritmos y estructuras de datos utilizadas con herramientas propias de la Geometría Computacional.

Búsqueda por Rangos

La *Búsqueda por Rangos* es uno de los problemas centrales en Geometría Computacional, no sólo por la variedad de aplicaciones que posee sino porque, además, una gran cantidad de problemas geométricos pueden resolverse observándolos como problemas relacionados a las búsquedas por rango.

Un *Espacio de Rangos* es un sistema $\Omega = (U, F)$, 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 $\mathbf{R}^d \})$

El sistema Ω es llamado *Espacio de Rango Finito* si el conjunto U es finito.

Un problema de búsqueda por rangos puede expresarse del siguiente modo:

> Dados un espacio de rangos $\Omega = (U, F)$, 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 frente a cada consulta por rango R, ésta 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 de objetos geométricos.

Observando los rangos de consultas, podemos hacer una clasificación en cuatro tipos clásicos de búsqueda por rangos:

- *Búsqueda por Rangos Ortogonales (BRO):* los rangos son d-rectángulos, cada uno de la forma $\Pi_{i=1..d} [a_i, b_i]$ $a_i, b_i \in \mathbf{R}^d$, con lados paralelos a los ejes de coordenadas del espacio subyacente.

- *Búsqueda por Rangos Semi-Espaciales (BRSp):* los rangos conforman el conjunto de todos los semiespacios cerrados de $\mathbf{R}^d$. Por ejemplo, en d=2 tenemos semiplanos.
- *Búsqueda por Rangos Simpliciales (BRSx):* los rangos son los simplices. En d=2 tenemos triángulos, en d=3 tetraedros, etc.
- *Búsqueda por Rangos Semi-Algebraicos (BRSa):* los rangos están definidos por conjuntos semialgebraicos. Por ejemplo, el conjunto de todas las paraboloides en $\mathbf{R}^3$.

Dado un rango de consulta, puede interesarnos efectuar alguna de las siguientes consultas sobre él:

- *Consultas de Recuento (range counting)*: se pregunta por la cardinalidad (o *peso* -ver más adelante-) del conjunto.
- *Consultas de Reporte (range reporting)*: se pide cuáles son los objetos geométricos que están contenidos en el rango.
- *Consultas Booleanas (range emptiness)*: se verifica si la intersección es vacía o no.
- *Consultas Especiales*: se pregunta por los objetos de $S \cap R$ que satisfacen alguna determinada propiedad.

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 particionamiento de *S* y del tamaño de la respuesta de la consulta.

A continuación expondremos algunos casos de un problema relacionado al problema de búsquedas por rangos, denominado *Búsqueda de Intersección*.

Búsqueda de Intersección

Dados un espacio de objetos geométricos de $\mathbf{R}^\mathbf{d}$, un conjunto de objetos *S* de $\mathbf{R}^\mathbf{d}$, un semigrupo $(\boldsymbol{D},+)$ y una función de peso $w{:}S \rightarrow \boldsymbol{D}$, queremos crear una estructura de datos para *S*, de modo que para un objeto de consulta *X*, podamos calcular la siguiente suma $\Sigma_{p \cap X \neq \phi}\, w(p)$, donde la suma se toma sobre todos los objetos $p \in S$ que intersectan a *X*.[1]

Por ejemplo, si los objetos $p \in S$ son puntos (en $\mathbf{R}^\mathbf{d}$), entonces estamos en un caso especial de búsquedas por rango. Independientemente del tipo de objeto geométrico que trate *X*, consultaremos por los puntos que intersectan a (están contenidos en) *X*.

Como antes, tenemos las variantes de consultas en el problema de intersección: recuento y reporte de una intersección, detección de una intersección distinta de vacío, etc.

Los siguientes son algunos ejemplos de problemas relacionados a la búsqueda de intersección:

- *Búsqueda de intersección para rectángulo:* dado un conjunto de polígonos *S* en R^2, se busca construir una estructura de datos para *S*, tal que todos los objetos de *S* que intersectan un rectángulo de consulta *X*, puedan ser reportados eficientemente. Este problema es conocido como *Windowing query*.
- *Consultas de Ray-Shooting:* dado un conjunto de objetos *S* en R^d, se busca construir una estructura de datos para *S*, tal que el primer[2] objeto de *S* que sea cruzado por una semirecta de consulta *X*, pueda ser reportado eficientemente.
- *Consultas de los vecinos más cercanos:* dado un conjunto de puntos *S* en R^d, se busca construir una estructura de datos para *S*, tal que todos los puntos de *S* más cercanos[3] a un punto de consulta *X*, puedan ser reportados eficientemente.

La búsqueda de intersección es uno de los problema más estudiados en Geometría Computacional, dado que surge de aplicaciones tales como sistemas de información geográfica, reconocimiento de patrones, recuperación de la información, biología molecular, computación gráfica, etc.

Aspectos geométricos de las bases de datos

Las bases de datos tienen sus aplicaciones más comunes en espacios de datos convencionales, referidos a números y cadenas de caracteres, llamados *atributos con dominios atómicos*.

[1] Obsérvese la generalización de *S* respecto de su conformación.
[2] Considérese algún orden sobre los objetos referidos, respecto de alguno de los ejes de coordenadas del espacio subyacente.
[3] En este caso debe considerarse un espacio métrico con alguna función de distancia definida sobre el espacio subyacente.

Un *dato geométrico* describe un objeto perteneciente al espacio geométrico, que posee características geométricas. Puede ser un punto, una recta, un polígono, una superficie, un volumen y así sucesivamente, podemos mencionar objetos de dimensión mayor.

Los datos geométricos generalmente van acompañados de datos convencionales y pueden ser *discretos* o *continuos*. En el caso discreto (un punto en un espacio d-dimensional, un instante en el tiempo), pueden ser modelados con atributos atómicos haciendo uso de modelos de bases de datos tradicionales: Relacional, Red, etc. Mientras que cuando son continuos (polígonos, superficies, intervalos de tiempo, etc.), cubren una región del espacio, y necesitan un tratamiento diferente respecto de las estructuras de datos utilizadas para su almacenamiento y de las técnicas algoritmicas empleadas.

La integración y/o combinación de ambos tipos de datos, ha dado lugar nuevos modelos de bases de datos tales como las *bases de datos espaciales,* las *bases de datos temporales,* y en algunos casos, unificando ambos modelos se tienen las *bases de datos espacio-temporales;* o, simplemente, algunos autores se refieren a *bases de datos geométricas*, como una generalización mayor.

Con esto podemos tomar una ligera noción de la necesidad de tener una herramienta teórica de base que nos permita modelar datos geométricos, realizar operaciones sobre ellos, definir un lenguaje de consultas especial, etc. En ello, la Geometría Computacional cumple un rol distintivo.

Objetivos

El objetivo de esta propuesta es realizar un trabajo de investigación, en el que mostremos el estado del arte del tema expuesto, presentando los aspectos teóricos y prácticos relevantes para las búsquedas por rangos, y que ello signifique un aporte a la comunidad científica, mostrando los tópicos principales del tema, los problemas abiertos o aún no resueltos y las referencias bibliográficas para acudir en ampliación o búsqueda de detalles.

Como objetivos subyacentes a esta investigación, se busca consolidar y alimentar una línea de estudio, que brinde un puente a trabajos de investigación de mayor profundidad, que permitan posteriormente cubrir temas de tesis doctorales o de maestría.

Conclusiones

La Geometría Computacional es una disciplina propia de la Matemáticas, la cual brinda un marco teórico y formal para el diseño de estructuras y análisis de algoritmos requeridos para dar soluciones a problemas que surgen en áreas de la Computación, tales como las bases de datos, recuperación de la información, computación gráfica, robótica, VLSI, etc. Las soluciones a tales problemas son halladas reduciendo los mismos a problemas propios de la Geometría Computacional. En nuestro trabajo nos proponemos observar las técnicas algoritmicas utilizadas, las estructuras de datos que surgen de la Geometría Computacional apropiadas para el manejo de datos geométricos y los algoritmos propuestos para resolver problemas relacionados a búsquedas por rangos, con sus respectivos análisis y complejidades.

Referencias Bibliográficas

- Aho, A.V.; Hopcroft, J. E.; Ullman, J. *The desing and analysis of computer algorithms*, Addison-Wesley Series in Computer Science and Information Processing, 1974.
- Berg, Kreveld, Overmars, Schwarzkopf. *Computational Geometry: algorithms and applications*, Springer Verlag, BH 1997
- Boissonnat, J.D.; Yvinec, M. *Algorithmic Geometry*, Cambridge University Press, 1998.
- Gonnet G.;Baeza Yates, R. *Handbook of Algorithms and Data Structures*
- Goodman, J.; O'Rourke, J. *Handbook of Discrete and Computational Geometry.* CRC Press 1997.
- Günther, O. *Efficient Structures For Geometric Data Management*. Springer Verlag 1988.
- Kurt Melhorn *Multidimensional Searching and Computational Geometry* Springer Verlag 1984
- Pach, J. *New trends in Discrete and Gomputational Geometry.* Springer Verlag NY 1993.
- Preparata, F.; Shamos,M. *Computational Geometry: an introduction,* /Springer Verlag, NY 1985
- Sack, J.R.; Urrutia, J.. *Handbook of Computational Geometry.* Elsevier Science B.V. 2000.
- Schneider, M. *Spatial Data Types For Database Systems* (corresponde a su tesis doctoral) 1995.