

일반적인 다각형 모양의 질의 윈도우를 이용한 공간 선택 질의의 정제 전략¹

A Refinement Strategy for Spatial Selection Queries with Generally Shaped Query Window

유준범(Junbeom Yoo) 정진완(Chin-Wan Chung)
한국과학기술원 전자전산학과
jbyoo@salmosa.kaist.ac.kr chungcw@islab.kaist.ac.kr

요 약

공간 선택 질의에 사용되는 질의 윈도우는 직사각형이 주로 사용된다. 하지만, 최근에는 GIS 등과 같은 응용 프로그램들이 성능 향상으로 인해 보다 다양한 종류의 응용이 등장하고 있으므로, 직사각형뿐만 아니라 임의의 다각형 형태의 질의 윈도우에도 적합한 정제 단계 수행 전략에 대해 고려해 볼 필요가 있다. 이러한 전략으로는 기존의 공간 조인에서와 같이 plane-sweep 알고리즘을 이용하는 방법이 일반적이다. 하지만, 공간 데이터와 질의 윈도우의 특성을 관찰해보면, 일반적으로 질의 윈도우가 공간 데이터보다 훨씬 간단한 모양으로 구성되어 있음을 알 수 있으므로, 본 논문에서는 이러한 상황에 보다 적합한 정제 단계 수행 방법을 제시하고 있으며, 실험을 통하여 제시한 방법의 우수성을 입증하고 있다.

1. 서 론

최근에 공간 데이터(spatial data)는 GIS(Geographic Information System) 및 공간 데이터베이스(spatial database) 등에서 널리 사용되고 있으며, 그 사용 또한 증가 추세에 있다. 공간 데이터 즉, 공간 객체는 매우 복잡한 구조로 이루어져 있으며 크기 또한 다양할 뿐만 아니라 그 양도 매우 크다. 따라서, 일반적인 데이터보다는 질의 즉, 공간 질의(spatial query)의 수행이 복잡하며 또한 수행하는데 많은 시간이 요청된다[1]. 이러한 공간 질의를 보다 효과적으로 수행하기 위한 연구로는 질의 수행 시간을 줄이려는 연구인 [2]의 여과 및 정제 전략이 대표적이다. 특히 정제 단계는 객체의 실제 모양을 이용하는 정교한 검사 알고리즘이 요구되므로 여과 단계보다는 많은 시간이 요구된다[1,3].

공간 선택 질의에 대한 연구는 질의 윈도우가 직사각형 모양인 경우에 대해서로 한정되어 있다. 이는 실제 응용에서 직사각형 질의 윈도우를 주로 사용해왔기 때문이다. 하지만, 요즘은 GIS와 같은 응용 프로그램들이 기존보다 더 많은 공간 데이터를 처리할 수 있게 됨에 따라 질의 윈도우의 모양이 직사각형이 아닌 일반적인 다각형 모양으로 간주하는 것이 보다 바람직하며 [3], 또한 질의 윈도우는 공간 데이터에 비해서 훨씬 간단한 모양이라고 가정해도 무리가 없다. 현재 다각형 모양의 질의 윈도우를 사용하는 공간 선택 질의의 정제 단계 수행을 위한 접근 방법으로는 plane-sweep이 사용되고

있지만[3], 이와 같은 상황에서는 복잡하고 정교한 알고리즘을 사용하는 것에 개선에 여지가 있으므로, 본 연구에서는 이러한 상황에 적합한 정제 단계 수행 알고리즘을 제안하고 있다.

본 논문의 구성은 다음과 같다. 2장에서는 공간 선택 질의의 수행과 관련된 기초적인 연구들을 간단히 기술하고 있으며, 3장에서는 본 논문에서 제안하는 PBST와 PEC에 대해서 설명하고 있다. 4장에서는 기존의 방법과의 비교실험 결과를 보이고 있으며, 5장에서 결론을 맺고 있다.

2. 관련 연구

이 장에서는 본 연구의 기저로서 공간 선택 질의를 수행하기 위한 기본적인 과정들과 progressive approximation에 대해서 설명하고 있다.

2.1 공간 선택 질의의 수행 과정

하나의 공간 객체는 수 개의 점에서부터 수 천 개의 점으로 구성될 수 있으며 각각의 크기도 매우 다양하므로, 이러한 공간 데이터를 효과적으로 처리하기 위해서는 R^* -tree와 같은 SAM을 사용해야 한다. R^* -tree는 인덱스를 구축하는데 있어서 MBR(Minimum Bounding Rectangle)과 같은 데이터의 conservative approximation을 이용하며, 이들을 이용하는 공간 선택 질의의 수행 과정은 다음의 두 단계로 나누어 진행될 수 있다.[2]

¹ 본 연구는 한국과학재단의 목적기초연구(과제번호 99-2-315-001-3) 지원으로 수행됨.

여과 단계: 질의 윈도우와 겹치는 MBR을 갖는 객체들을 후보로서 찾는다.

정제 단계: 이러한 후보 객체들이 실제로 질의 윈도우와 겹치는지를 실제 모양을 이용해서 검사한다.

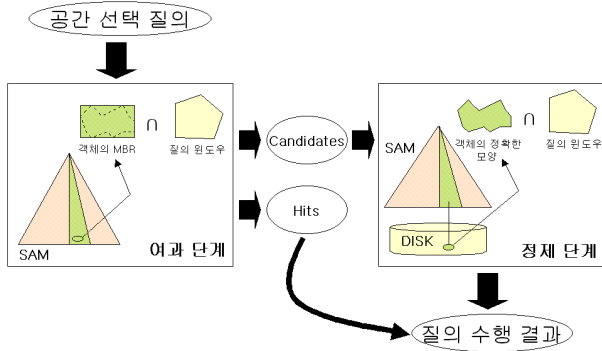


그림 1: 공간 선택 질의의 수행 과정

위의 과정을 설명하고 있는 그림 1에서 알 수 있듯이, 정제 단계는 여과 단계보다 더 많은 시간이 소요되므로 [3], 이러한 정제 단계에서 소요되는 비용을 줄이는 것이 중요하다.

2.2 Progressive Approximation

한 공간 객체의 Progressive approximation은 그 객체에 의해 최대를 둘러싸는 도형을 의미한다. 만약 한 객체가 다른 객체의 progressive approximation이 겹친다면, 두 객체도 역시 겹친다고 결론지을 수 있다. 따라서, 이 방법은 여과 단계로부터 검출된 지원자 객체들로부터 true-hit을 미리 검사하는데 간접적으로 사용될 수 있다. 하지만, 미리 저장 공간의 사용과 계산하는 비용이 크기 때문에 비교적 간단한 MEC(Maximum Enclosed Circle)와 MER(Maximum Enclosed Rectangle)만이 주로 사용된다 [1].

3. 정제 단계 수행 전략

3.1 기존의 방법: Plane-Sweep을 사용한다.

Plane-Sweep 알고리즘 두 다각형을 구성하는 선분들 간의 겹침 관계를 $O(n \log n)$ 의 시간 내에 조사할 수 있는 알고리즘이며, 이를 이용하는 정제 단계 수행 과정은 다음과 같다.

- (1) 검사 영역을 제한한다.
- (2) 모든 선분들을 정렬한다.
- (3) Plane-sweep 알고리즘을 수행한다.

이러한 plane-sweep 알고리즘을 적용하기 전에 (1)의 검사 영역 제한의 작업이 선행되어야 하는데, 이는 그림 2에서와 같이 질의 윈도우와 공간 객체는 반드시 두 다각형의 MBR이 겹치는 부분에서만 겹칠 수 있기 때문이

다 [1].

분석: n_1 과 n_2 가 각각 질의 윈도우와 공간 객체를 구성하는 선분(점)의 개수라고 할 때, 이 방법의 전체 수행 시간은 $O((n_1+n_2)\log(n_1+n_2))$ 이라 할 수 있다. 그러나 일반적으로 질의 윈도우는 공간 데이터에 비교했을 때 매우 간단한 모양이므로, 위의 복잡도(complexity)식에서 n_2 의 크기가 n_1 보다 훨씬 작으므로 n_2 의 영향을 무시할 수 있다. 그러므로, 이러한 상황에 보다 적합한 다른 정제 단계 수행 알고리즘 특히 nested-loop과 같은 간단한 순차 검색을 응용한 알고리즘을 고려해볼 여지가 있다.

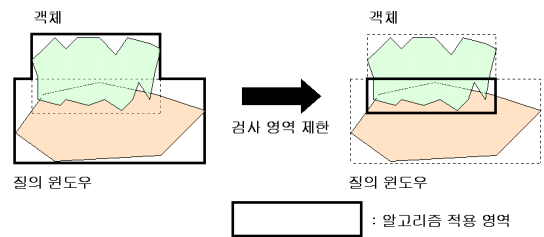


그림 2: 검사 영역 제한

3.2 PEC (Partial Enclosed Circle) Approximation

본 논문에서는 2.2에서 설명한 여과 단계에서 사용되는 progressive approximation을 정제 단계 과정에 적용하여 추가적인 오버헤드는 줄이면서 기능은 충분히 할 수 있게 하는 PEC(Partial Enclosed Circle)을 제안하고 있다.

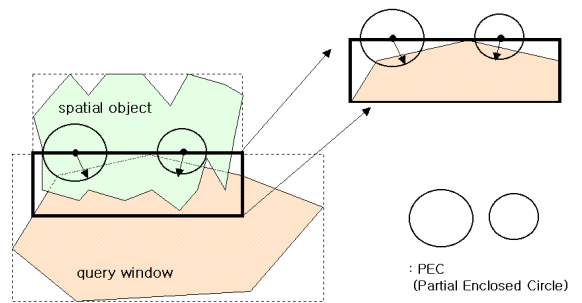


그림 3: PEC progressive approximation

PEC는 그림 3에서 알 수 있듯이 PEC의 중심과 반지름을 이용해서 질의 윈도우의 점과 선분에 대한 거리 검사가 성공한다면, 더 이상의 복잡한 검사 작업 없이 간단하게 true-hit을 검출해 낼 수 있다.

3.3 새로운 정제 전략 방법: PBST를 사용한다.

PBST (Partition Based Sequential Testing)는 PEC approximation에 기초한 순차 검색 기법이다. 앞서 언급

한 바와 같이 PEC는 progressive approximation이므로 실패한 작업만 plane-sweep 검사와 같은 다음 과정이 요구된다. 만약 PEC 검사가 실패하면, 우리는 PEC가 있지 않는 다른 영역에서 질의 윈도우와 겹치는 현상이 발생할 것임을 예측할 수 있으므로 이런 영역들을 먼저 순차 검색하는 것이 보다 효율적이다. PBST의 전반적인 수행 과정은 다음과 같으며, 그림 4는 두 방법의 차이를 보여주고 있다.

- (1) 검사 영역을 제한한다.
- (2) PEC approximation 검사를 한다.
- (3) 실패하면, PEC의 영역이 아닌 부분에 있는 선분들에 대해서 순차적으로 겹침을 검사한다.
- (4) 실패하면, PEC의 영역 내에 있는 선분들에 대해서 순차적으로 검사한다.

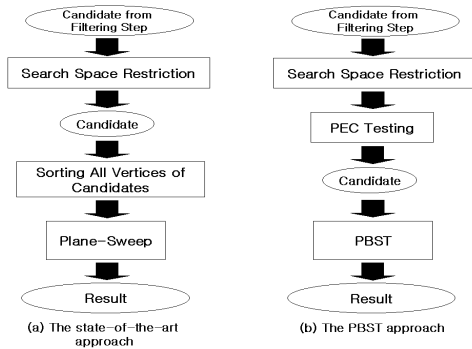
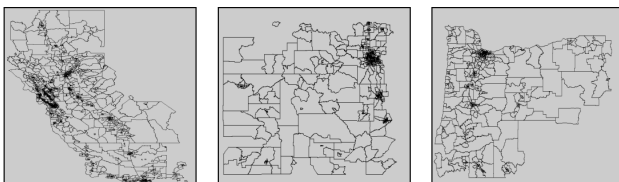


그림 4: 두 방법의 수행 과정의 비교

4. 실험 평가

4.1 실험 Setup

본 논문에서는 실험 데이터로서 TIGER/Line Files[4]에서 캘리포니아 지역의 CTBNA 데이터를 사용하였으며, 이로부터 3 개의 test data set을 구성하였다 또한, 4, 8, 12, 16, 20개의 점으로 구성된 100개의 임의의 질의 윈도우들로 구성된 test query set을 이용하였다. 그림 5는 실험에 사용한 데이터에 대한 정보를 요약하고 있다.



테스트 데이터 Set	데이터 수 (개)	평균점의 수 (개)	데이터의 분포(%)		최저점의 수(개)	최고점의 수(개)
			100개 미만	100개 이상		
Set 1	2878	168	75.12	24.88	4	4586
Set 2	838	215	70.52	29.48	4	2809
Set 3	689	380	47.61	52.39	5	5382

그림 5: 실험 데이터의 특성 요약

4.2 성능 분석

그림 6은 test data set 2에 대해 기존의 방법과 새로 제안하는 방법의 전체 성능을 비교 분석한 결과를 나타낸다. 전체적으로 PBST 방법을 사용하는 새로이 제안하는 방법이 plane-sweep 기술을 사용하는 기존의 방법에 비해서 약 20% 정도의 성능 향상을 모임을 알 수 있다.

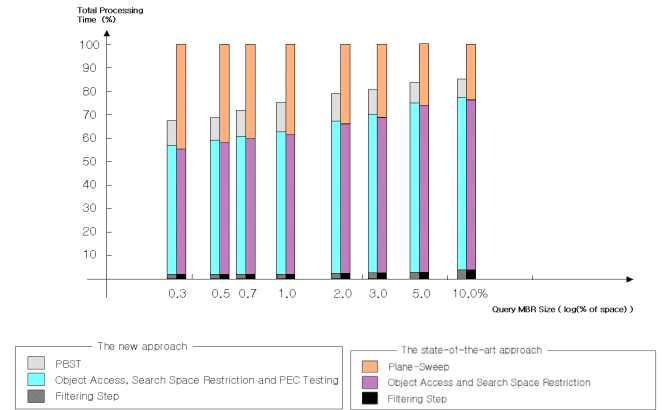


그림 6: 전체 성능 분석

5. 결론

최근에는 GIS 등과 같은 응용 프로그램들이 성능 향상과 여러 다양한 종류의 응용이 많이 등장함에 따라 직사각형뿐만 아니라 임의의 다각형 형태의 질의 윈도우에도 적합한 정제 단계 수행 전략에 대해 고려해 볼 필요가 있다. 따라서, 본 연구에서는 일반적인 다각형 형태의 공간 선택 질의를 위한 정제 단계 수행 방법으로서 PEC에 기반하는 PBST 방법을 제안하고 있다. 실 세계에서 사용되는 지도 데이터를 이용해서 비교 평가한 결과, 일반적으로 볼 때 PBST를 사용하는 새로운 방법이, 질의 윈도우를 구성하는 점의 개수가 약 20개 이하인 경우에는, plane-sweep을 사용하는 기존의 방법보다 약 20% 정도의 성능 향상을 보이고 있음을 알 수 있다.

참고 문헌

- [1] T. Brinkhoff, H.P. Kriegel, R. R. Schnieder, B. Seeger, "Multi-Step Processing of Spatial Joins," ACM SIGMOD 1994
- [2] J. A. Orenstein, "A Comparison of Spatial Query Processing Techniques for Native and Parameter Space," ACM SIGMOD 1990
- [3] A. Abounaga, J. F. Naughton, "Accurate Estimation of the Cost of Spatial Selections," IEEE ICDE 2000
- [4] Bureau of the Census, "TIGER/Line Files, 1995: Technical Documentation," 1996