

보르노이 다이어그램 상의 효율적인 이웃 MBR 연산 기법

An Efficient Method for Finding the Neighbor MBRs on Voronoi Diagram

박용훈, 이진주, 임종태, 최길성*, 유재수
충북대학교, 동아방송예술대학*

Yonghun Park, Jinju Lee, Jongtae Lim,
Kilseong Choi*, Jaesoo Yoo
Chungbuk National University,
Dong-Ah Institute of Media and Arts*

요약

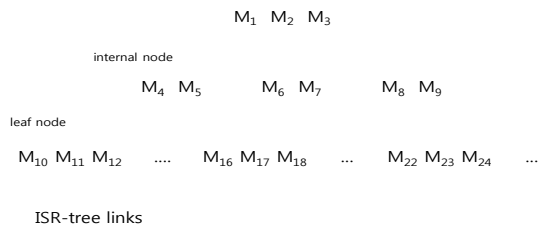
이동객체의 공간 데이터를 색인하기 위해 검색성능이 뛰어난 R-tree구조가 많이 활용된다. 최근 R-tree를 B+-tree처럼 인접한 단말노드 간의 연결을 통해 질의 처리를 수행하는 ISR-tree와 ISG-index가 제안되었다. 이 기법들은 MBR (Minimum Boundary Rectangle) 간의 인접한 이웃 노드를 결정하기 위해 보르노이 다이어그램(Voronoi Diagram)을 이용한다. MBR을 대상으로 하는 보르노이 다이어그램은 매우 복잡한 연산과정을 거친다. 본 논문에서는 점을 대상으로 하는 보르노이 다이어그램 연산을 활용한 인접한 이웃 MBR을 연산하는 기법을 제안한다. 각 MBR의 꼭지점들을 기준으로 보르노이 다이어그램을 만들 경우, 인접한 MBR의 꼭지점들의 보르노이 셀이 항상 인접한 것을 알아내었고, 이를 활용한다. 제안하는 기법의 우수성을 증명하기 위해 기존의 기법과 비교하여 성능평가를 수행하였다.

I. 소개

최근 위치 인식 기술과 휴대 장치의 발달로 인해 이동하는 객체를 기반으로 하는 위치 기반 서비스(Location Based Service, LBS)에 관심이 점점 증가하고 있고 그에 관련된 연구들이 활발하게 진행되고 있다. 대규모의 확장성있는 위치 기반 서비스를 제공하기 위해 소프트웨어 측면에서는 각 객체들에 대한 정보를 빠르게 검색하는 색인 기법이 필요하다. 공간 좌표와 같은 위치 데이터를 위한 공간 색인 기법으로서 R-tree가 많이 사용되어 왔고, 사용 목적에 따라 R-tree를 개선한 기법들이 제안되었다[1].

최근에는 R-tree상에서 B+-tree와 같이 인접한 노드 정보를 활용한 질의 처리를 지원하는 ISR-tree와 ISG-index가 제안되었다[2][3]. 이 기법에서는 단말노드에서 인접한 이웃 단말노드 정보를 유지한다. k-NN 질의 처리 시, 질의 점과 가장 근접한 단말노드만 찾으면 더 이상 중간노드를 탐색하지 않고 질의를 수행한다. 그리고 단말노드의 이웃 노드를 결정하기 위해 단말노드 MBR간의 보르노이 다이어그램을 구성하고 이때 임의의 두 MBR의 보르노이 셀이 인접한 경우 서로를 이웃으로 결정한다. 그림 1은 ISR-tree의 구조적 특징을 보여준다. 단말노드를 제외하면 R-tree와 동일한 구조를 가진다. 하지만 단말노드에서는 이웃 정보를 추가적으로 유지하여 질의 처리 시 중간노드의 탐색 횟수를 줄인다.

* 이 논문은 2010년 교육과학기술부(지역거점연구단육성사업/충북BIT연구중심대학육성사업단)와 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업의 결과임.(No. 2009-0089128)



▶▶ 그림 1. ISR-tree의 구조적 특징

MBR을 대상으로 하는 보르노이 다이어그램은 매우 복잡한 연산과정을 거친다. 점 데이터를 대상으로 하는 보르노이 다이어그램은 점 간의 경계선을 만드는데 한 개의 직선이 필요하다. 하지만 MBR 데이터를 대상으로 하는 보르노이 다이어그램은 MBR 간의 경계선을 만드는데 일반적으로 5개의 직선과 4개의 곡선이 필요하다.

본 논문에서는 효율적으로 MBR 간의 이웃을 결정하는 기법을 제안한다. MBR들의 꼭지점을 기준으로 보르노이 다이어그램을 구성했을 때 한 MBR의 4개의 꼭지점들의 인접한 점들의 집합은 그 MBR의 이웃 MBR들의 꼭지점을 적어도 1개 이상 포함한다. 제안하는 기법은 MBR 간의 경계를 그리지 않고 이웃 MBR 간의 이웃 관계를 연산한다.

본 논문의 구성은 아래와 같다. 2장에서는 제안하는 이웃 MBR 연산 기법을 기술한다. 3장에서는 성능평가를 통해 제안하는 기법의 우수성을 보여준다. 마지막으로 4장에서는 결론을 기술한다.

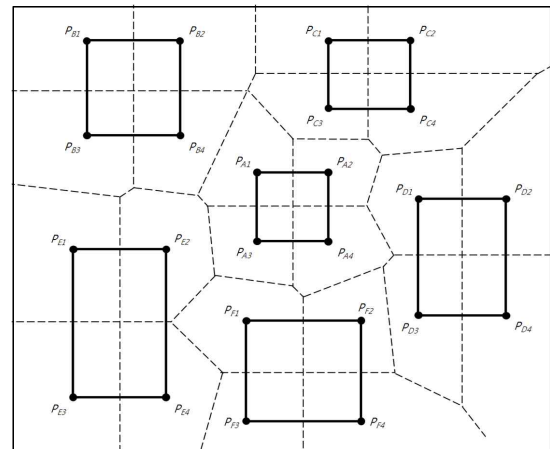
II. 제안하는 이웃 MBR 연산 기법

본 논문에서는 효율적으로 MBR 간의 이웃을 결정하는 기법을 제안한다. MBR의 꼭지점들을 기준으로 보르노이 다이어그램을 만들 경우, 인접한 MBR에 속한 꼭지점의 보르노이 셀은 항상 인접한 것을 알아내었다. 따라서 MBR의 꼭지점들에게 MBR을 식별할 수 있는 식별자를 부여하고, 꼭지점들 간에 보르노이 다이어그램 연산을 수행한다. 그 결과, 인접한 MBR의 꼭지점의 보르노이 셀은 항상 인접하기 때문에, 점과 점의 보르노이 다이어그램 결과에 포함된다. 나온 결과들의 식별자를 이용하여, 이웃하는 MBR들을 결정한다.

2차원 MBR 간의 경계선을 그리고, 보르노이 셀을 만들어 이웃을 결정하는 과정은 많은 방정식에 대한 연산

을 필요로 하기 때문에 비용 소모가 크다. 하지만 제안하는 기법은 점과 점의 보르노이 다이어그램 연산만을 수행하고, 꼭지점에 부여된 식별자를 이용하여 MBR 간의 이웃을 결정하기 때문에 적은 비용이 든다.

그림 2는 MBR의 꼭지점들을 기준으로 보르노이 다이어그램을 작성한 그림이다. 2차원 공간에 MBR이 존재하고, 각 MBR의 꼭지점은 P_{A1} , P_{B1} 과 같이 MBR을 식별할 수 있는 식별자를 가진다. MBR 간의 이웃을 결정하기 위해서 꼭지점들을 기준으로 그림 2와 같이 점과 점의 보르노이 다이어그램을 작성한다. 그 결과, 꼭지점 P_{A1} 의 인접한 보르노이 셀의 점은 $\langle P_{B2}, P_{B4}, P_{C3}, P_{A2}, P_{A3}, P_{A4}, P_{E2} \rangle$ 이다. 이 결과로부터 P_{A1} 의 이웃에는 MBR $\langle B, C, E \rangle$ 가 존재한다는 것을 확인한다. 같은 방법으로 P_{A2} , P_{A3} , P_{A4} 의 이웃에 존재하는 MBR들을 탐색하면, MBR A 의 이웃을 모두 확인할 수 있다.

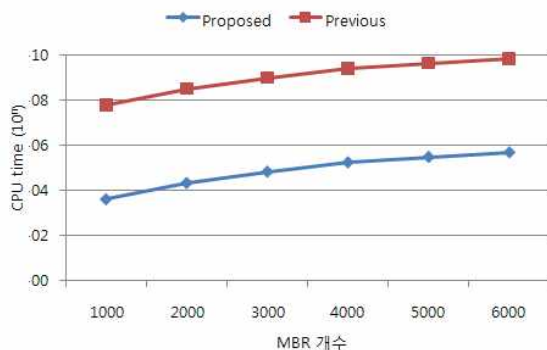


▶▶ 그림 2. 꼭지점 기준의 보르노이 다이어그램

MBR의 꼭지점을 이용한 이웃 MBR 연산 기법의 결과는 False Positive를 갖는다. 선을 무시하고 점을 기반으로 하기 때문에 MBR의 겹침이 발생할 경우에 False Positive가 발생한다. 주어진 결과에 대한 추가적인 재평가 기법이 필요하다. 재평가 기법은 보르노이 다이어그램 상의 이웃 꼭지점을 MBR 단위로 묶어서 임의의 MBR의 변에 의해 한 MBR의 모든 결과 점이 모두 가려지는 경우를 찾아서 결과 MBR에서 제거한다. 제안하는 기법은 재평가를 통해 정확한 이웃 노드를 연산한다.

III. 성능평가

본 논문에서 제안하는 효율적인 이웃 MBR 연산 기법의 성능을 기존의 기법과 비교 평가하였다. 평가 환경은 Intel Core2 Duo CPU 2.53GHz, 2GB RAM 그리고 Windows XP 32bit로 구성하였다. MBR의 개수에 따라 이웃 MBR 연산 시간을 측정하였고, MBR의 개수는 1000에서 6000까지 변경하였다. 성능 평가결과 제안하는 기법은 기존 기법보다 보다 10^4 배 이상 향상된 처리 성능을 보였다. 그림 3은 제안하는 기법과 기존 기법의 성능평가 결과를 보여준다. CPU time은 지수 단위로 표현하였다.



▶▶ 그림 3. MBR 개수에 따른 연산 시간 비교

IV. 결론

본 논문에서 제안하는 기법은 효율적으로 MBR 간의 이웃을 결정하는 기법을 제안하였다. 2차원 MBR 간의 경계선을 그리고, 보르노이 셀을 만들어 이웃을 결정하는 과정은 많은 방정식에 대한 연산을 필요로 하기 때문에 많은 비용이 소모된다. 제안하는 기법은 점과 점의 보르노이 다이어그램 연산만을 수행하고, 꼭지점에 부여된 식별자를 이용하여 MBR 간의 이웃을 결정하기 때문에 적은 비용이 든다. 그리고 재평가 기법을 추가적으로 수행하여 False Positive 문제를 해결하였다. 성능 평가 결과 기존의 기법 보다 10^4 배 이상 빠르게 처리되는 것을 확인하였다. 향후 연구로는 MBR이 동적으로 변할 경우 효율적인 처리 기법을 연구 할 것이다.

■ 참고 문헌 ■

- [1] T. Sellis, N. Roussopoulos, and C. Faloutsos, "The R+-Tree: A dynamic index for multi-dimensional objects", In Proc. intl. conf. Very Large Data Bases, pp. 507-518, 1987.
- [2] Y. Park, D. Seo, H. Park and J. Yoo, "An Index Structure for Efficient k-NN Query Processing in Location Based Services", In Proc. intl. conf. Ubiquitous Information Technologies & Applications, pp. 165-170, 2009.
- [3] Y. Park, D. Seo, J. Lim, J. Lee, M. Kim, W. Bao, C. Ryu and J. Yoo, "A New Spatial Index Structure for Efficient Query Processing in Location Based Services", In Proc. intl. Workshop on Ubiquitous and Mobile Computing, 2010.