

데이터 방송 환경에서 Sweep SFC를 활용한 효율적인 스카이라인 질의 처리 기법*

하중우^o 최재호 이정현 이상근

고려대학교 정보통신대학

{okcomputer, redcolor25, jhbslpd, yalphy}@korea.ac.kr

Efficient Skyline Query Processing Method utilizing Sweep SFC in Data Broadcast Environment

JongWoo Ha^o Jae-Ho Choi Jung-Hyun Lee SangKeun Lee

College of Information and Communications, Korea University

1. 서론

최근 무선 통신 기술과 모바일 기기의 비약적인 발전에 힘입어 데이터 방송 서비스가 널리 활용되고 있다. 대표적인 예인 T-DMB의 경우 멀티미디어 방송을 비롯하여 뉴스, 날씨, 교통 정보, 주식 등의 데이터 방송 서비스를 제공하고 있다. 이러한 데이터 방송 서비스에서 서버는 데이터 집합을 가공하여 방송 프로그램을 제작한 후 방송 채널을 통하여 푸시(push)하는 방식으로 모바일 사용자에게 데이터를 전송한다. 이때 방송 채널의 특성으로 인하여 다수의 사용자가 동시 다발적으로 데이터를 수신할 수 있게 된다.

본 논문에서는 데이터 방송 환경에서 모바일 사용자가 스카이라인 질의를 점진적으로 처리하는 문제를 다룬다. 스카이라인 질의는 복수 개의 속성을 가지는 데이터 집합에서 다른 데이터로부터 지배되지 않는 데이터 집합을 반환하는 질의이다[1]. 여기서, 데이터 x 의 모든 속성이 다른 데이터 y 보다 더 작으면 x 가 y 를 지배한다고 정의한다. 스카이라인 질의는 복수 개의 기준으로 의사 결정을 수행하는 응용 분야(중고차 매매, 호텔 예약 등)에서 널리 활용될 수 있기 때문에 최근 활발한 연구가 진행되고 있다.

스카이라인은 데이터 집합에 대하여 정의되기 때문에, 데이터 방송 환경에서 모바일 사용자가 스카이라인 질의를 처리하기 위해서는 데이터 집합 전체를 다운로드 해야 한다. 하지만 이러한 방식은 모바일 기기의 배터리를 빨리 소모시키기 때문에 비효율적이다. 이러한 문제를 해결하기 위하여, 본 논문에서는 분산 공간 인덱스(DSI: distributed spatial index)[2]에 Sweep SFC(space-filling curve)[3]를 적용한 점진적인 질의 처리 기법을 제안한다. DSI를 적용함으로써 질의 처리 과정에서 발견된 데이터로부터 지배되는 다른 데이터들을 다른 데이터들을 다운로드 하지 않을 수 있기 때문에 튜닝 시간(tuning time) 성능을 개선할 수 있다. 더불어, 사용자가 스카이라인의 부분 집합을 질의할 경우 Sweep SFC의 특성을 활용하여 질의 처리를 점진적으로 수행할 수 있기 때문에, 튜닝 시간 및 접근 시간(tuning time) 성능을 비약적으로 개선할 수 있다.

2. 본론

기존의 DSI에서는 다차원의 데이터를 1차원의 포인트로 인코딩 할 때 Hilbert 커브[4]를 적용하였다. 이는 다차원의 공간에서 근처에 위치한 데이터들이 비슷한 인코딩 값을 가진다는 특성이 있기 때문에 위치 기반 서비스 질의를 처리할 때 유리한 측면이 있다. 하지만 Sweep SFC는 Hilbert 커브와 달리 데이터의 속성 값에 따라 인코딩이 이루어지기 때문에 스카이라인 질의를 점진적으로 처리할 때 유리하게 활용할 수 있는 특성이 있다. 첫 번째로, 특정 데이터 x 보다 Sweep SFC 값이 큰 데이터는 x 를 지배할 수 없다. 두 번째로, x 보다 작은 Sweep SFC 값을 가지는 모든 데이터 중에서 x 를 지배하는 데이터가 없으면 x 는 스카이라인에 포함된다. 이러한 두 가지 특성에 따라서, 질의 처리 과정에서 x 보다 Sweep SFC 값이 작은 데이터가 모두 발견되었음에도 불구하고 x 를 지배하는 데이터가 없다면 x 가 스카이라인에 포함된다는 것을 점진적으로 확인할 수 있다. 이러한 점은 기존의 DSI에서 적용한

* 이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임(No. 2009-0077925)

Hilbert 커브에서는 발견할 수 없는 특성이다.

이러한 특성을 활용하여, 제안 기법에서 서버는 다차원의 데이터를 Sweep SFC로 인코딩 한 후 DSI 기법[2]을 적용하여 방송 프로그램을 제작한다. 방송 프로그램은 인덱스 테이블과 데이터 패킷의 두 종류로 구분되며, 각 인덱스 테이블은 데이터의 Sweep SFC 인코딩 값과 포인터로 구성되어 있다. 인코딩 값을 통하여 데이터 집합에서 다른 데이터에 지배되는 데이터를 확인할 수 있게 된다. 포인터는 그에 상응하는 데이터가 방송되는 시간을 기록한 것으로서, 모바일 사용자는 이를 확인하여 질의 처리 과정에 불필요한 데이터를 다운로드 하지 않을 수 있게 된다. 또한 하나의 인덱스 테이블에는 여러 개의 테이블 엔트리를 포함하고 있으며, 각각의 엔트리 i (i 는 0부터 시작하여 1씩 증가하는 양의 정수)는 현재 방송되는 데이터로부터 2^i 만큼 떨어진 데이터의 인코딩 값과 포인터를 가진다.

Sweep SFC를 적용한 DSI를 활용하여 제작된 방송 프로그램을 바탕으로, 모바일 사용자는 그 일부분을 다운로드 함으로써 스카이라인 질의를 점진적으로 처리할 수 있다. 매번 인덱스 테이블을 다운로드 하면서 모바일 사용자는 데이터 집합 구성의 일부분을 확인할 수 있다. 그 과정에서 현재 발견된 데이터로부터 지배되는 데이터를 다운로드 하지 않음으로써 튜닝 시간 성능을 개선할 수 있다. 또한 특정 데이터 x 보다 작은 인코딩 값을 가지는 모든 데이터가 발견 되었음에도 불구하고 x 를 지배하는 데이터가 없으면 x 는 스카이라인에 포함된다는 사실을 질의 처리 중간에 파악할 수 있다. 따라서 이러한 데이터들을 점진적으로 반환함으로써, 스카이라인의 부분 집합을 질의할 경우 질의 처리를 더 빨리 종료할 수 있다. 더불어 이에 따른 튜닝 시간 및 접근 시간을 줄일 수 있게 된다.

본 논문에서 제안한 기법의 성능을 측정하기 위하여 CSIM for Java에 기반을 둔 시뮬레이터를 구현하였다. 시뮬레이터에서 방송 서버는 데이터 집합으로부터 방송 프로그램을 제작한 후 패킷 단위로 방송 프로그램을 전송한다. 모바일 사용자는 방송되는 데이터를 접근하는 방식으로 스카이라인 질의를 처리한다. 데이터의 각 차원이 서로 독립적인 independent 데이터 집합을 생성하여 실험 데이터로 활용하였다. 각 데이터 집합은 10,000개의 데이터로 구성되어 있다. 일반적으로 데이터의 차원이 커질수록 의사 결정의 기준이 늘어나는 효과가 있기 때문에 스카이라인의 크기도 함께 커진다. 제안 기법(Sweep 기법)의 성능을 비교하기 위한 비교 대상은 총 2개이며 각각 Naïve와 Hilbert로 명명하였다. Naïve는 DSI를 적용하지 않고 서버에서 데이터만을 전송하였을 때의 성능을 파악하기 위한 것으로, 모바일 사용자가 스카이라인 질의를 처리하기 위해서는 방송 프로그램 전체를 튜닝 및 접근해야 한다. Hilbert는 원래의 DSI에서 채택한 Hilbert 커브를 바탕으로 데이터를 인덱싱 하여 방송하였을 때의 성능을 측정해보기 위한 것이다.

앞서 살펴본 바와 같이 Hilbert 커브를 활용하여 점진적인 질의 처리를 수행하는 것이 어렵기 때문에, Hilbert 기법을 활용할 경우 스카이라인의 부분집합을 반환하는 질의 처리 성능이 전체 스카이라인을 반환하는 질의 처리 성능과 동일한 것을 확인할 수 있다. 이와는 달리 Sweep 기법의 경우 질의 처리 과정에서 특정 데이터가 스카이라인인지 여부를 판단할 수 있기 때문에 점진적인 질의 처리가 가능하다. 따라서 실험 결과, Sweep 기법을 적용할 경우 Naïve와 비교하였을 때 2차원 데이터 집합에서 튜닝 시간 및 접근 시간 성능이 각각 94%, 44% 향상됨을 확인하였다.

3. 결론

본 논문에서 데이터 방송 환경에서 모바일 사용자가 효율적으로 스카이라인 질의를 처리하기 위한 기법을 제안하였다. 제안 기법은 기존의 효율적인 인덱싱 기법인 DSI에 스카이라인 질의를 점진적으로 처리하는데 유리한 특성을 가진 Sweep SFC를 적용하였다. 실험 결과를 통하여 제안 기법이 스카이라인 질의 처리의 튜닝 시간 및 접근 시간 성능을 비약적으로 개선하는 것을 확인하였다.

참고문헌

- [1] S. Börzsönyi, D. Kossmann, and K. Stocker, "The skyline operator," in *ICDE*, 2001, pp. 421–430.
- [2] W.-C. Lee and B. Zheng, "Dsi: A fully distributed spatial index for location-based wireless broadcast services," in *ICDCS*, 2005, pp. 349–358.
- [3] M. F. Mokbel, W. G. Aref, and I. Kamel, "Analysis of multi-dimensional space-filling curves," *Geoinformatica*, vol. 7, no. 3, pp. 179–209, 2003.
- [4] C. Gotsman and M. Lindenbaum, "On the metric properties of discrete space-filling curves." *IEEE Trans Image Process*, vol. 5, no. 5, pp. 794–797, 1996.