

모바일 환경에서 고품질 추천을 위한 스카이라인 확장 기법 분석

김지현, 김 명

이화여자대학교 컴퓨터공학과

e-mail: jhrosa@ewhain.net, mkim@ewha.ac.kr

An Analysis of Skyline Extension Schemes for High Quality Recommendations in Mobile Environments

Jihyun Kim, Myung Kim

Dept. of Computer Science and Engineering, Ewha Womans University

요 약

의사 결정을 위한 추천 시스템에 유용하게 쓰이는 스카이라인 계산은 데이터간의 비교에 상당한 시간이 소요되기 때문에 추천 서비스를 실시간에 하기 위해서는 스카이라인을 사전에 계산해 두는 것이 일반적이다. 그러나 모바일 환경에서는 사용자의 위치에 따라 스카이라인이 변하기 때문에, 사전에 계산해 놓은 스카이라인을 조정하는 다양한 연구가 진행되고 있다. 본 연구에서는 이동하는 사용자에게 품질 높은 데이터를 추천해 주기 위해 스카이라인을 확장하는 기법들을 제안하고 이들을 분석해 보고자 한다.

1. 서론

휴대용 모바일 기기의 급속한 발전으로 다차원 대용량 데이터 집합을 대상으로 하여, 사용자들의 의사결정에 도움을 주는 다양한 정보 서비스들이 제공되고 있다. 그 중에서 의사결정을 위한 추천 시스템에 활용도가 높은 연산으로 스카이라인 (skyline) 을 들 수 있다.

스카이라인[1]은 다차원 데이터 집합의 원소들 중에서 집합내의 다른 원소에게 그 속성 값들이 지배당하지 않는 원소들로 구성된 부분집합을 말한다. 예를 들어, 해변가에 있는 집들을 {가격, 해변으로부터의 거리, 전망} 속성으로 정의할 때 가격이 저렴하고, 해변으로부터의 거리가 가까우며 전망이 좋은 집들의 집합이 스카이라인에 속한다.

스카이라인 계산에는 집합내의 데이터간의 비교에 많은 시간이 소요되기 때문에 데이터의 속성이 변하지 않는 경우는 사전에 스카이라인을 계산해 두고 추천 서비스를 하게 된다. 그러나, 질의 객체 (또는, 사용자) 가 이동하는 모바일 상황에서는 이와 같이 계산된 스카이라인을 그대로 사용하기 어렵다. 예를 들면, 사용자가 분당 지역에서 맛집 추천을 받기를 원하는데 추천할 맛집들의 대부분이 부산 지역에 몰려 있을 수 있기 때문이다.

최근 들어 모바일 환경을 고려하여 스카이라인을 계산하는 연구에 많은 관심이 모이고 있다. 주요 연구 방향으로는 변하지 않는 정적 속성들로만 스카이라인을 계산해 두고, 사용자가 이동하는 상황에서는 그 사용자에게 거리가 가까운 데이터를 실시간으로 계산해서 추가로 추천해 주는 것이다. 이

와 같은 방향의 기법들은 사용자에게 정적 속성과 동적 (거리) 속성을 고려한 스카이라인을 제공하지만, 스카이라인이 아니면서 추가로 제공되는 데이터의 정적 속성 품질은 높지 않을 수 있다는 단점이 있다. 예를 들어, 분당 지역에서 사용자가 맛집 추천을 원할 때 품질은 좋지 않지만 사용자에게 단지 가깝게 위치한다는 이유로 추천될 수 있기 때문이다.

본 연구에서는 이러한 점을 고려하여 실제 상황에서 사용자에게 품질 높은 데이터를 추천해주고자 스카이라인 개념을 확장하여 다양하게 정의하고, 이들의 장단점을 분석해 보고자 한다. 본 논문은 다음과 같은 구성된다. 2절에서는 기존 연구를 소개하고, 3절에서는 다양한 확장 스카이라인 계산 기법을 소개하며, 4절에서 이 기법들을 분석하고 성능 평가를 하며, 5절에서 결론을 맺는다.

2. 관련 연구

모바일 환경에서 스카이라인을 사용하여 정보추천 서비스에 활용하려는 연구로 [2~7]을 들 수 있다. [2]에서는 사전에 구해 놓은 스카이라인 원소들과 이동하는 질의 객체 사이에 관심 영역을 설정하는 기법을 소개하고, [3]의 기법은 도로 상에 차도를 가정하여 미리 계산된 객체들의 최단 영역을 이용하여 스카이라인 원소들 중 가까운 점을 반환하는 다양한 기법들을 소개하고 있다. 또한 동적 속성을 고려한 이동 환경에서의 스카이라인 계산 기법도 연구되고 있는데 [4]는 이동환경에서의 스카이라인 계산시, 시간을 고려한 이동 객체의 패턴에 따른 연속적 스카이라인

질의 처리를 소개하고, [5]에서는 동적 자료 구조를 사용한 그리드 기반 기법을 이용하여 이동객체를 위한 스카이라인 계산 기법을 소개하고 있다. 이동 환경에서 스카이라인을 계산하는 기존의 연구들은 정적 차원에 대한 스카이라인 영역이나 패턴을 전처리[7]한 후 그 결과를 바탕으로 후보 스카이라인으로 선정하여 이동 객체의 질의에 맞게 최종 스카이라인을 구한다. 이러한 연구는 질의 객체가 이동하는 상황에서는 정적 차원 기준의 전처리된 스카이라인 점들의 품질을 보장하기 어렵다. 이에 본 연구는 모바일 환경에서의 스카이라인의 품질 좋은 데이터를 확보하기 위한 다양한 스카이라인 확장 기법을 제안하고자 한다.

3. 스카이라인 확장 기법

우선 정적 환경에서의 스카이라인 구조를 살펴보자. 예를 들어, 그림 1(a)에 2차원 데이터가 가우스 분포 형태 (원 모양) 로 영역 A 와 같이 주어졌다고 하자. 이 데이터 집합의 스카이라인은 개념적으로 볼 때, 원 둘레의 짧은 선, m , 과 같은 형태가 된다. 즉, 사용자의 요구조건에 최적인 원점에서 가까우면서 다른 데이터에 지배받지 않는 데이터들로 구성된다. 그러나 사용자가 이동하는 상황에서는 이와 같이 구해진 스카이라인이 사용자의 주변에서 멀리 떨어져 있을 수가 있다. 본 연구에서는 이러한 스카이라인 이외에 품질이 우수한 데이터를 추가로 추천하는 기법들을 제안하고 비교해 보고자 한다.

확장 1: 밴드 형태의 확장

다차원 데이터 집합 A 로부터 스카이라인을 구하고 이를 $S = \{s_1, s_2, \dots, s_k\}$ 라고 하자. 이제 스카이라인에 속하지 않는 데이터 집합, $A-S$, 의 각 원소 p 에 대해, p 와 거리가 가장 가까운 스카이라인 원소 s_i 를 찾고, p 와 s_i 의 거리가 사전에 정한 한계치 (T , threshold) 이내에 들면 p 를 확장 스카이라인 ES 에 포함시킨다. ES 에는 S 도 포함된다.

확장 2: 품질을 고려한 밴드 형태의 확장

다차원 데이터 집합 A 로부터 스카이라인을 구하고 이를 $S = \{s_1, s_2, \dots, s_k\}$ 라고 하자. 이제 스카이라인에 속하지 않는 데이터 집합, $A-S$, 의 각 원소 p 에 대해, p 를 지배하는 모든 스카이라인 점들을 찾고, 그 중에서 원점에 가장 가까운 점을 s_i 라고 하자. 원소 p 와 s_i 사이의 거리가 한계치 T 이내에 들면 p 를 확장 스카이라인에 포함시킨다. ES 에는 S 도 포함된다.

확장 3: 그리드와 경계 스무딩을 통한 확장

다차원 데이터 집합 A 를 동적 속성인 위치 (가로, 세로 좌표)를 기준으로 그리드 형태로 그림 1(b)와 같이 분할한다. 그 다음에는 각 파티션, $P_{i,j}$, 별로 독립적으로 스카이라인 $S_{i,j}$ 를 구한다. 그 다음에는 파티션 $P_{i,j}$ 의 중심을 기준으로 하고, $P_{i,j}$ 의 내부 대각선을 지름으로 하는 원 C 를 그림

1(b)와 같이 그렸을 때, 이 원과 주변 파티션과 겹치는 부분이 각 거리 차원 (가로, 세로) 별로 2개씩 생긴다 (그림에서 D 로 표시한 음영이 있는 부분임). 이 겹치는 부분을 북, 동, 남, 서 방향으로 각각 $V_{i,j,N}$, $V_{i,j,E}$, $V_{i,j,S}$, $V_{i,j,W}$ 라고 하자. 다음 단계에서는 $V_{i,j,N}$ 에 속한 스카이라인 점들과 $V_{i-1,j,S}$ 에 속한 스카이라인 점들을 모아 다시 스카이라인을 구한다. 마찬가지로 $V_{i,j,E} \cup V_{i,j+1,W}$, $V_{i,j,S} \cup V_{i+1,j,N}$, $V_{i,j,W} \cup V_{i,j-1,E}$ 각각에 대해서도 각 겹치는 부분에 속한 스카이라인 점들 중에서 다시 스카이라인을 구한다. 구해진 모든 스카이라인 점들을 합하여 최종 확장 스카이라인 ES 로 정한다.

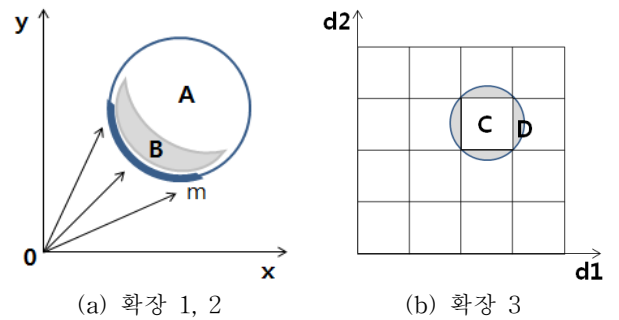


그림 1. 스카이라인 확장 기법

4. 스카이라인 확장 기법 분석 및 평가

앞서 소개한 확장 스카이라인에 대해 분석해 보기로 한다. 우선 정적 환경에서 이 기법들을 살펴보자. [확장 1]과 [확장 2] [6]의 기본 개념을 도식화 한 것이 그림 1 (a)에 있다. 그림에서 전체 데이터 집합은 A 로 표현할 수 있고, 스카이라인은 원주의 짧은 선인 m 으로 표현할 수 있으며, 확장 스카이라인에는 m 과 스카이라인에 근접하여 한계치 안에 드는 영역 B 의 원소들이 추가된다.

여기에서 [확장 1]의 경우는 스카이라인에 속한 모든 점들과 거리가 한계치인 T 이내에 드는 점들로 품질이 비교적 우수하다고 볼 수 있다. 이를 좀 더 발전시킨 것이 [확장 2] 기법이다. 이는 [확장 1]에서 구한 점들 중에서 스카이라인에는 가깝지만 사용자의 요구조건인 원점으로부터의 거리가 멀리 떨어져 있는 점들을 제거하는 효과를 준다. 즉, [확장 1]과 [확장 2] 기법은 스카이라인보다는 품질이 떨어지지만 스카이라인에 근접한 점들을 확보해 주는 역할을 한다. 그러나, 이렇게 추가로 확보된 점들이 이동하는 사용자의 주변에 많이 분포해 있다는 보장이 되는 것은 아니다.

이동중인 사용자가 특정 위치에 있을 때 그 사용자에게 품질이 보장되는 점들을 추가로 실시간에 계산해 주는 것이 쉽지 않다. 따라서 [확장 3]에서는 사용자의 위치에 관계없이 관심 대상 지역에 품질이 우수한 데이터를 골고루 분포시키고 추천하려는 시도이다. 각 그리드 내부에서 스카이라인을 구하여 후보 스카이라인으로 골라놓고 이후 인접한 음영 부분의 스카이라인 원소들과 합병을 통해 최종 스카이라

인으로 결정하는 것이다. 이 경우 그리드 내부의 후보 스카이라인으로 선택된 스카이라인 원소들은 품질이 우수한 원소들로 구성되어 있으며 인접한 부분의 스카이라인 원소들까지 보정되므로 고품질의 스카이라인 원소들이 선택되게 된다.

확장 2와 3에 대해 성능평가를 해 보았다. 실험은 2.40GHz 인텔 i5 CPU와 4.0GB 메모리를 장착한 PC 에서 하였고 .NET 환경에서 C 언어로 프로그래밍을 하였다. 상관 분포를 갖는 10M 개의 4차원 데이터로 실험한 결과 확장 2의 경우 스카이라인 계산에 1.369초 걸렸고, 이 중에서 확장 스카이라인 계산에 드는 시간은 0.432초였다. 확장 스카이라인의 크기는 한계치와 관련 있으며, 실험에서는 스카이라인에 7배 가량 되도록 한계치를 정하였다. 확장 스카이라인에 속한 데이터가 크더라도 스카이라인 계산과 연계하여 확장 스카이라인이 계산되므로, 추가 비용이 무시할만한 정도로 작다는 것을 알 수 있다. 확장 3의 경우는 각 파티션 별로 스카이라인을 구하게 되어 데이터간 비교회수가 크게 감소하게 되며, 확장 스카이라인은 스카이라인이 구해진 후에 추가로 계산되는 것이므로 계산 시간이 크게 늘지 않는다. 실험에서는 확장 스카이라인 계산에 총 1.967초가 걸렸고, 확장 스카이라인의 크기는 스카이라인 크기의 4배 가량 되었다. 그러나 이 경우는 관심 대상 지역에 추천할 데이터가 고르게 분포한다는 점이 확장 2와 다른 점이다.

5. 결론

본 연구에서는 모바일 환경에서의 사용자 주변에서 최적인 품질 좋은 데이터를 추천해 주는데 쓰일 수 있는 확장 스카이라인 계산 기법들을 소개하고 분석하였다. 확장 스카이라인은 스카이라인 계산 과정 또는 그 이후에 이루어지는 것으로 데이터가 비교회수가 크지 않으므로, 계산 시간 복잡도가 매우 낮아 실용적으로 사용될 수 있다는 것을 알 수 있다. 확장 스카이라인은 모바일 상황에서 다양한 응용에 활용될 수 있는 장점을 갖는다.

참고문헌

- [1] S. Borzsonyi, D. Kossmann, and K. Stocker, "The Skyline Operator," *In ICDE*, pp.421~430, 2001.
- [2] 나경석, 김진호, 안대혁, 박영배, "연속적인 스카이라인 질의를 위한 효율적인 영역 결정기법," *정보과학회 데이터베이스 연구*, 제22권 제2호, 2006.
- [3] 유재수, 장수민, "도로망에서 연속적인 스카이라인 질의 처리를 위한 효율적인 전처리기법," *정보과학회 논문지*, 제36권, 제4호, pp. 314~320, 2009.
- [4] Z. Hung, H. Lu, B. Ooi and A.Tung, "Continuous skyline queries for moving objects," *IEEE TKDE*, Vol. 18. pp. 1645~1658, 2006.
- [5] K-L. Tian, L. Wang, P. Zou, Y. Jia, and A. Li., "Continuous monitoring of skyline query over highly dynamic moving objects," *In MobiDE*, pp. 59~66, 2007.

- [6] 김지현, 김명, "다차원 데이터 분석을 위한 효율적인 확장 스카이라인 계산 알고리즘," *한국데이터베이스연구회 논문지*, 제27권 제3호, pp.67~80, 2011.