

센서 네트워크에서 스카이라인 계산을 위한 계층적 관계를 유지하는 필터 선정 기법

김준수, 전주혁, 김명호
KAIST 전산학과

kimjs@dbserver.kaist.ac.kr, jhjeon@dbserver.kaist.ac.kr, mhkim@dbserver.kaist.ac.kr

Method to Select Hierarchical Filters for Skyline Computation in Sensor Networks

Jun Su Kim, Joo Hyuk Jeon, Myoung Ho Kim
Division of Computer Science, KAIST

1. 서론

현재까지 센서 네트워크에서 연속 스카이라인 질의를 처리하기 위한 다양한 기법이 제안되어 왔다. 센서 네트워크에서는 에너지를 효율적으로 사용해야 하기 때문에 가장 큰 에너지를 소모하는 데이터 전송량을 줄이는 것이 중요하다. 이를 위하여 필터링 기법이 사용되며 필터링 기법 중, 센서 노드들의 필터링 튜플들 사이에 계층적인 관계를 유지하는 기법은 데이터 전송량 측면에서 많은 장점을 가진다. 하지만, 스카이라인 질의 처리를 위한 기존의 필터링 튜플 선정 기법들은 계층적인 관계를 유지하면서도 필터링 효과가 뛰어난 필터링 튜플을 선정하는 것에 한계를 가진다. 본 논문에서는 이를 해결하여 계층적인 관계를 유지하면서도 필터링 효과가 뛰어난 필터링 튜플을 선정하는 기법을 제안하며 실험을 통해 그 효과를 검증한다.

2. 제안하는 필터링 튜플 선정 기법

MAX질의 처리에 대한 연구 [1]에서 센서 노드의 필터링 튜플 τ 와 부모 노드의 필터링 튜플 τ_p 사이에 $\tau \leq \tau_p$ 관계를 유지하는 기법이 제안되었다. 이 기법은 최종 결과에 포함되지 않는 불필요한 데이터가 베이스 스테이션까지 전송되는 것을 중도에 막을 수 있고, 데이터의 시간적 상관관계를 효과적으로 활용할 수 있어 데이터 전송량 측면에서 많은 이득을 가진다. 스카이라인 질의 처리시, 이러한 이득을 얻기 위해서는 필터링 튜플을 선정하는 기법이 특성1을 만족해야 한다. 특성1에서 \leq 은 지배관계를 나타내는 기호로써, 튜플 t_1 이 t_2 를 지배하는 경우 $t_1 \leq t_2$ 로 표기한다. 또한, 집합간의 지배 관계 $P \leq Q$ 는 모든 튜플 $q \in Q$ 에 대하여 $p \leq q$ 를 만족하는 튜플 $p \in P$ 가 존재한다는 것을 의미한다. CalFilter 함수는 필터링 선정기법을 의미하며, 이 함수의 결과가 필터링 튜플로 선정된다.

특성1. 집합 P, Q에 대하여

$P \leq Q$ 이면 $\text{CalFilter}(P) \leq \text{CalFilter}(Q)$ 이다

네트워크 내 프로세싱을 통해 스카이라인 질의를 처리하는 경우, 특성1은 센서 노드들에 할당된

필터링 튜플 사이에 계층관계를 유지하여 [1]에서 언급된 이득을 가질 수 있게 한다.

[2]에서 필터링 튜플 선정 기법으로 제안된 MinMax 연산자는 동일한 차원수를 갖는 튜플의 집합을 입력으로 하여 하나의 튜플을 출력하는 연산자로서, d차원 튜플의 집합이 $P = \{p_1, p_2, \dots, p_n\}$ 일 경우 다음과 같이 정의된다.

$$\max_i = \text{MAX}_{k=1}^d p_i[k] \quad (1 \leq i \leq n) \quad (1)$$

$$\text{minmax} = \text{MIN}_{i=1}^n \max_i \quad (2)$$

$$\text{MinMax}(P) = \underbrace{\{\text{minmax}, \text{minmax}, \dots, \text{minmax}\}}_d \quad (3)$$

MinMax 연산자를 통한 필터 선정 기법은 특성1을 만족하지만 그림 1처럼 균일하지 않은 데이터 분포에서는 필터링 효과가 떨어진다.

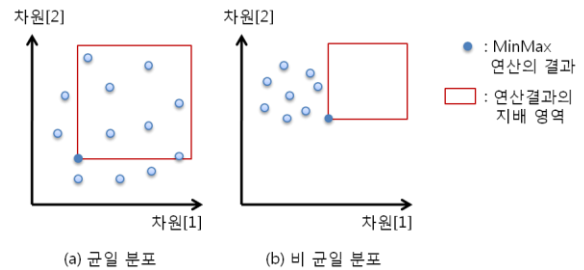


그림 1 MinMax 연산의 결과 비교

제안하는 기법은 MinMax 연산을 일반화한 방법으로써 “프로젝션”과 “최소선택” 두 단계로 구성된다. MinMax의 과정 중 과정(1)은 “프로젝션” 단계에 과정(2)는 “최소선택” 단계에 속한다고 볼 수 있으며 그림 2와 같이 나타낼 수 있다. 그림 2는 편의상 2차원 공간을 가정하였다.

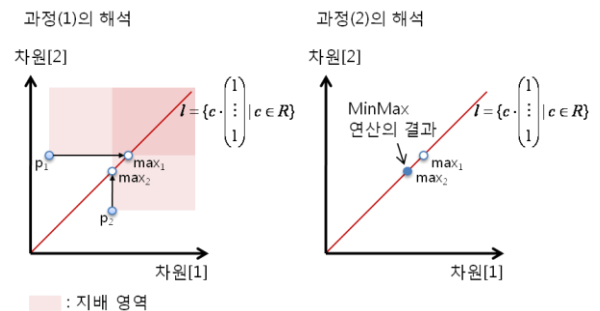


그림 2 MinMax 연산과정의 해석

본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT 연구센터 육성·지원사업(NIPA-2010-C1090-1031-0002)의 연구결과로 수행되었음.

튜플의 최대값이 포함되지 않은 차원의 값들을 대상 값이라고 정의 한 경우에, “프로젝션”단계는 대상 값이 최대값과 같아지도록 튜플들을 프로젝트 라인에 프로젝트 시킨다. 과정(1)은 프로젝트 라인 l 이 $l = \{c \cdot (1, \dots, 1) | c \in R\}$ 인 경우라고 볼 수 있다. 그림 2의 “과정(1)의 해석”은 과정(1)을 “프로젝션” 단계에 비추어 본 것이다.

“최소선택”단계는 프로젝트 된 점들 중에서 원점과 가장 가까운 점을 찾는 단계이다. 과정(2)는 이 단계와 동일한 의미를 가지며 그림 4의 “과정(2)의 해석”은 과정(2)를 “최소 선택” 단계에 비추어 본 것이다.

이러한 해석에 착안하여 본 논문에서는 데이터 분포를 반영할 수 있는 프로젝트 라인을 설정하여 해당 라인에 “프로젝션”단계와 “최소 선택”단계를 적용하는 필터링 튜플 선정 기법을 제안하고자 한다.

데이터 분포를 반영하는 프로젝트 라인은 아래의 휴리스틱을 이용하여 설정한다.

휴리스틱 프로젝트 라인에 의해 분할되는 각각의 영역이 동일한 수의 데이터를 가지도록 프로젝트 라인을 설정한다.

프로젝션 라인 $l = \{c \cdot (l[1], \dots, l[d]) | c \in R\}$ 이 휴리스틱을 통해 정해지면, “프로젝션” 단계에서는 정렬1을 통해 프로젝트 된 튜플들을 구한다. 이후로 오는 설명에서는 튜플의 집합을 P , P 에 속하는 임의의 튜플을 p , 이로부터 프로젝트 된 튜플을 q 라한다.

정리 1

$$q^{(\max)} = \max \left\{ \frac{p[1]}{l[1]}, \dots, \frac{p[d]}{l[d]} \right\} \cdot (l[1], \dots, l[d])$$

“최소 선택” 단계에서는 원점과 가장 가까운 튜플이 필터링 튜플로 선정되는데 프로젝트 된 튜플들의 라인을 나타내는 성분 $(l[1], \dots, l[d])$ 은 모두 동일하므로 최소의 $\max \left\{ \frac{p[1]}{l[1]}, \dots, \frac{p[d]}{l[d]} \right\}$ 를 계수 성분으로 갖는 튜플이 필터링 튜플로 선정된다. 제안한 기법이 특성1을 만족함을 증명하는 것은 지면관계상 생략하기로 하며 본 연구의 풀 버전을 통해 확인 할 수 있다.

그림 3은 기존의 기법[2]가 선정한 필터링 튜플과 제안한 기법이 선정한 필터링 튜플을 보여주며 데이터의 분포가 비 균일한 경우에도 제안한 기법이 선정한 필터링 튜플의 필터링 효과가 뛰어난을 알 수 있다.

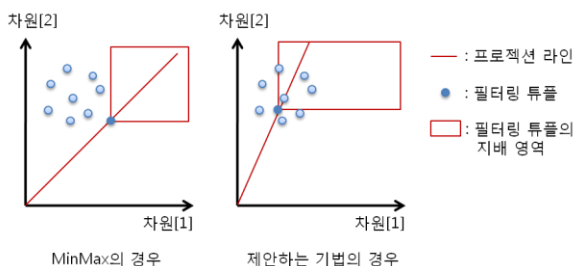


그림 3 비 균일 분포에서의 비교

제안하는 기법은 프로젝트 라인에 의해 분할되는 각 영역의 데이터 수가 임계값 T 이상의 차이를 보일 경우, 프로젝트 라인을 재설정한다.

3. 실험 결과 및 결론

그림 4는 [2]의 MinMax기법과 제안하는 기법의 필터링 효과를 비교한 것이다. 필터링 전의 스카이라인의 튜플 수를 Sky , 필터링한 튜플 수를 $Filter$, 프로젝트 라인의 재설정을 알리는데 쓰인 튜플 수를 $Reset$ 이라 했을 때, 두 기법 모두 $Sky - Filter + Reset$ 을 측정 하였다.

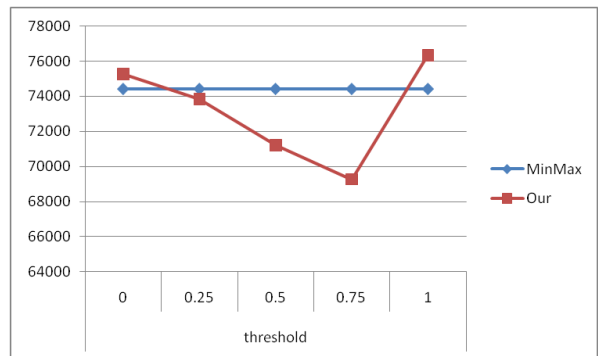


그림 4 경향성을 갖는 데이터 분포에서의 비교

프로젝션 라인의 재설정을 판단 하는 기준값으로 $\frac{\text{영역간 데이터 크기의 차}}{\text{센싱 된 데이터 크기}}$ 를 사용하였으며 샘플링을 통해 전체 데이터의 10%를 이용하여 기준값을 얻었다. 프로젝트 라인의 재설정은 기준값이 임계값 T 보다 큰 경우 이루어진다.

임계값이 0인 경우 영역간에 데이터 크기의 차이가 발생할 때마다 프로젝트 라인을 재설정 하기 때문에 프로젝트 라인의 재설정이 너무 빈번하게 발생하여 필터링 효과가 떨어졌다. 반대로 임계값이 1인 경우에는 모든 데이터가 하나의 영역에 몰려 있을 경우에만 프로젝트 라인의 재설정을 실시한다. 따라서 초기화 단계에서 설정된 프로젝트 라인이 계속 사용되어 필터링 효과가 떨어졌다. 하지만 이 두 가지의 극단적인 경우를 제외하고 제안하는 기법은 [2]의 MinMax 기법에 비해 뛰어난 필터링 효과를 보였다.

본 논문에서 제안한 기법은 센서 노드의 필터 사이에 계층적인 관계를 유지하여 [1]이 가지는 데이터 전송량 측면에서의 장점을 가지며 [2]의 기법에 비해 뛰어난 필터링 효과를 보였다.

참고 문헌

- [1] A. Silberstein, K. Munagala and J. Yang, “Energy-Efficient Monitoring of Extreme Values in Sensor Networks,” Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '06), June 2006.
- [2] H. Chen, S. Zhou and J. Guan. “Towards energy-efficient skyline monitoring in wireless sensor networks”, Proc. of EWSN, 2007.