

이동방송 환경에서의 효율적인 NN 탐색 기법*

이명수^o 이상근
고려대학교 모바일데이터공학 연구실
(lms9711, yalphy)@korea.ac.kr

An Efficient Searching Method for Nearest Neighbor in Mobile Broadcast Environments

Myong-Soo Lee^o Sang Keun Lee
Dept. of Computer Science, Korea University

요 약

무선 방송 방식은 부족한 대역폭의 효율적인 활용과 채널을 듣는 모든 사용자를 지원할 수 있다는 효율성 측면에서 각광받고 있다. 위치기반 서비스 중에서도 효율적인 방송기법을 이용하기 위한 연구 및 가장 기본적인 질의 중 하나인 NN 질의를 효율적으로 수행하기 위한 연구가 이루어져 왔다. 그러나 기존의 연구된 기법들은 NN 탐색 시 하나 이상의 방송주기를 필요로 하여 긴 접근 시간을 가진다는 단점이 있다. 이러한 단점은 모바일 환경에서 비효율적으로 자원을 사용한다는 문제를 발생시킨다. 이에 따라 본 논문에서는 한층 효율적인 자원 사용을 위해서 무선 기기에서 무선 방송 채널을 통해 NN 탐색을 수행할 수 있는 새로운 기법을 제안하고자 한다. 기존의 기법들에 비해서 접근 시간과 튜닝 시간을 줄임으로써 본 논문에서는 효율적으로 자원을 사용하고자 한다. 또한, 실험을 통해 본 논문에서 제안한 기법이 기존의 기법보다 향상된 성능을 보이는 것을 증명한다.

1. 서 론

개인 디지털 장비 활용이 증가하고 무선 통신 기술이 발전하면서 모바일 환경에서 사용자들에게 위치에 기반을 둔 위치기반 서비스(LBS: Location Based Services)를 제공하기 위한 노력들이 활발히 진행되고 있다[1]. 그중에서도 사용자가 서버에 질의를 요청하는 온-디맨드(On-demand) 방식에 비해 부족한 대역폭의 효율적인 활용과, 채널을 듣는 모든 사용자를 지원할 수 있는 무선 방송 방식이 크게 대두되었다[2].

NN(Nearest Neighbor) 탐색은 위치기반 서비스의 중요한 부분을 차지한다. “가장 근접한 호텔을 찾아라.”와 같은 NN 탐색은 많은 연구가 진행되어왔다[3]. 대부분의 연구가 공간 데이터베이스 내에서 해결되어 왔으나, 최근 무선 방송 환경에서의 NN 탐색을 수행하기 위한 기법들이 연구되고 있다. 본 논문에서는 제한된 자원을 가진 무선기기에서 무선 방송 채널을 통해 NN 탐색을 수행할 수 있는 새로운 기법을 제안하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서 현재까지 제안된 색인 기법에 대해 살펴보고, 3장에서는 기존의 색인 기법을 개선한 기법을 제안한다. 4장에서는 제안한 기법의 성능을 평가한다. 마지막으로 5장에서는 결론 및 향후 연구 방향을 제시한다.

2. 관련 연구

지금까지 무선 방송 환경에서의 NN 탐색을 수행하기 위한 많은 색인들이 제안되었다[4]. 무선 방송 환경에서의 색인 방법은 기존의 온-디맨드 방식과는 다른 색인이 요구된다. 즉, 기존의 온-디맨드 방식에서는 사용자가 서버에 공간 정보를 요청하면 서버에서 공간 데이터베이스 내의 색인을 통해 탐색을 하지만, 무선 방송 환경에서는 방송 채널 내의 색인을 통해 탐

색을 한다. 따라서 기존 온-디맨드 방식에서 가능했던 백트래킹(Backtracking)이 불가능함으로 새로운 색인이 필요하게 되었다. 이러한 색인으로서 순차 탐색을 이용한 R-tree, D-tree, Grid-partition이 제안되었다.

2.1 R-tree

R-tree는 공간 데이터베이스 내에서의 탐색을 위해 가장 많이 쓰이는 색인이다[5]. 초기에 무선 방송 환경에서의 NN 탐색을 위해 기존의 R-tree를 변형하여 순차적 탐색이 가능하게 하는 기법이 제안되었다. 기존의 R-tree에서 행해지던 브랜치-앤-바운드(Branch-and-Bound) 탐색 방식은 무선 방송환경에서 백트래킹 시 다음 방송주기까지 기다려야 하므로, 접근시간이 크게 증가한다. 이에 기존의 브랜치-앤-바운드 탐색방식을 순차 탐색 방식으로 바꾸어 접근시간을 줄이고자 하였다.

2.2 D-tree

R-tree가 NN 탐색을 위해 객체를 색인화한데 반해, D-tree[6]는 보로노이 다이어그램(Voronoi Diagram)[7]을 이용하여 미리 계산된 솔루션 공간을 색인화한다. 보로노이 다이어그램은 특정 객체가 NN 질의의 결과가 될 수 있는 지역을 하나의 보로노이 셀로 표현하는 그래프이다. D-tree는 보로노이 셀을 하나의 공간객체로 간주하여 전체 공간을 보로노이 셀의 경계를 형성하는 분할을 기반으로 보로노이 셀들을 색인화한다. 이때, 순환적으로 각각의 분할된 공간을 다시 동등한 수의 셀을 가진 두 개의 공간으로 재분할한다.

2.3 Grid-partition

Grid-partition은 D-tree와 같이 보로노이 셀을 이용지만, 전체공간을 격자로 나눔으로써 탐색 공간을 줄이고자 한 기법이다[8]. 이 기법은 우선 전체 공간을 보로노이 셀로 구성한 후, 격자로 탐색 공간을 분할한다. 각각의 격자 셀에서는 격자 셀 내부의 질의점의 잠재적 NN에 대한 모든 객체를 색인화한다. 즉, 격자 셀과 보로노이 셀이 겹쳐지는 모든 객체가 색인화

* 본 결과물은 정보통신부의 정보통신 기초기술연구지원사업 (정보통신연구진흥원)으로 수행한 연구결과입니다.

된다. 격자 크기가 모두 동일한 고정된 격자 분할 기법의 예는 그림 1과 같다.

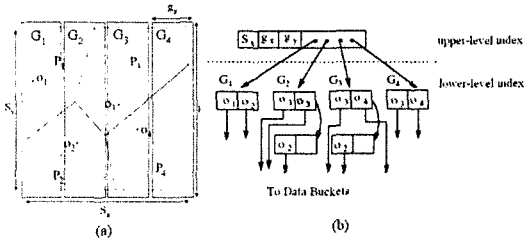


그림 1 고정된 격자 분할 기법의 예

그림 1(a)는 전체 공간을 4개의 동일한 격자로 분할하는 예를 보여준다. 각 격자는 해당 격자와 겹치는 보로노이 셀을 가지는 객체를 모두 색인화한다. 그림 1(b)는 색인 구조를 나타내는데, 상위 레벨의 색인은 질의점과 그에 해당하는 격자를 찾아준다. 하위 레벨의 색인은 각 격자 내의 모든 객체에 대한 접근을 가능하게 한다. 하위 레벨의 색인은 전송하는 패킷 크기별로 분할하여 순차적으로 전송하게 된다.

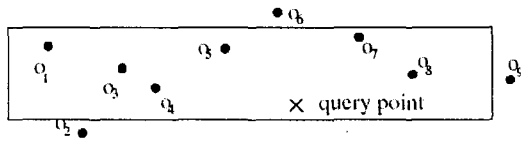


그림 2 격자 예제

Grid-partition 기법은 격자 내에서의 NN을 탐색할 때 해당 격자 내의 객체에 대해서 각각 질의점과의 거리를 측정 비교한다. 격자 내의 객체와 질의점이 그림 2와 같을 때, 탐색과정은 표 1과 같다.

표 1 NN 탐색 예제

단계	객체	dis_sd	cur_dis	min_dis
1	O ₆	1	7	7
2	O ₇	4	7	7
3	O ₅	4	6	6
4	O ₈	7.5	stop	.
5	O ₄	8	stop	.

NN 탐색은 다음과 같이 행해진다. 우선 각 객체는 한 축에 의해서 정렬된다. 정렬된 순서는 O₁, O₂, O₃, O₄, O₅, O₆, O₇, O₈, O₉이다. 질의가 수행되고 해당 질의에 대한 격자가 결정되면, 해당 격자 내에서 그 격자와 정렬한 축에서의 거리(dis_sd)가 작은 순서대로 객체를 검색해 나가면서 질의점과의 거리(cur_dis)를 측정하여 최소 거리(min_dis)를 가지는 객체를 찾는다. 이때, dis_sd가 min_dis보다 크게 되면 현재의 최소거리를 가지는 객체가 NN이 되므로 탐색을 멈추게 된다. 이와 같은 탐색 기법을 통해 모든 객체를 검색하지 않고 NN 탐색을 수행할 수 있다.

3. 제안 기법

3.1 기존 Grid-partition의 문제점

기존에 제안된 Grid-partition은 격자 내의 NN 탐색 시 격

자 내의 객체를 질의점을 기준으로 앞뒤로 이동하며 탐색한다. 이때, 뒤에 위치한 객체를 검색한 후 앞의 객체를 탐색하려 할 때, 이미 방송 채널 상에서 지나가버린 패킷에 포함된 객체일 경우 해당 방송주기 내에서는 탐색이 불가능하고 다음 주기가 지 기다려야 한다.

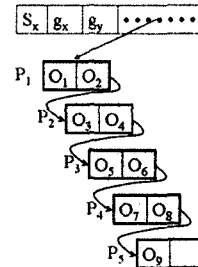


그림 3 '그림2'의 색인 구조

그림 2의 색인 구조가 그림 3과 같다면, 표 1과 같은 NN 탐색 과정에서 패킷의 순서는 P₃, P₄, P₃, P₄, P₂이다. 이때 P₄에서 P₃로 탐색해야 할 때, P₃는 P₄ 이전에 전송되므로 해당 방송주기 내에서는 탐색이 불가능하다. 따라서 다음 주기에 해당 패킷을 탐색해야 한다. P₄에서 P₂로 탐색해야 할 때도 같은 상황이 발생하므로 위의 예제에서는 NN 탐색을 위해 3번의 방송주기가 지나야 하므로 긴 접근 시간을 요구한다.

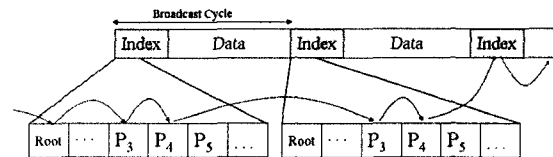


그림 4 NN 탐색 과정

그림 4는 표1의 예제에 대한 탐색 과정을 보여준다. 사용자는 세 번의 방송 주기의 색인을 읽은 후에 실제 데이터에 접근할 수 있다.

3.2 제안 기법

기존의 Grid-partition의 긴 접근 시간을 요구하는 문제점을 보완하기 위해서 본 논문에서는 새로운 NN 탐색 기법을 제안한다. 기존의 Grid-partition에서는 질의점을 기준으로 한축에 대한 거리가 가까운 객체 순으로 탐색을 수행한 반면, 본 논문에서는 정렬된 순서대로 탐색을 수행하여 앞 패킷의 객체를 탐색하지 않아 접근 시간이 짧아진다. 그림 2에 대한 본 기법의 탐색 예제는 표 2와 같다.

표 2 제안된 기법의 탐색 예제

단계	객체	dis_sd	cur_dis	min_dis
1	O ₁	-12	13	13
2	O ₂	-11	12	12
3	O ₃	-10	11	11
4	O ₄	-8	8.5	8.5
5	O ₅	-4	6	6
6	O ₆	-1	7	6
7	O ₇	4	7	6
8	O ₈	7.5	stop	.

(* (-)는 객체의 위치가 질의점과 비교하여 왼쪽에 위치함을 나타냄)

위의 표를 보면 객체는 정렬된 순서대로 차례대로 탐색된다. 탐색은 정렬된 축에서의 객체가 질의점보다 크고, 정렬된 축에서의 거리가 현재 최소거리보다 클 때($dis_sd > min_dis$), 탐색은 수행을 멈추고 현재의 최소거리를 가지는 객체를 NN으로 선정하게 된다.

4. 실험 및 평가

제안한 기법에 대한 성능을 평가하기 위해서, R-tree와 기존의 Grid-partition 기법을 제안한 기법과 비교하였다. 방송기법에서의 색인 기술의 주요 성능 측정 요소는 결과값을 얻는 시간인 접근 시간(Access Time)과 사용자가 활동 모드로 전역할 수 소모하게 되는 시간인 튜닝 시간(Tuning Time)이므로 이 두 가지 요소를 측정하였다. 데이터 집합은 각각 5922개의 객체를 가지고 있으며, 균등한 분포를 지닌 Uniform Data와 R-tree 포탈[9] 웹사이트에 있는 그리스의 도시와 마을의 위치인 Skewed Data를 사용하였다.

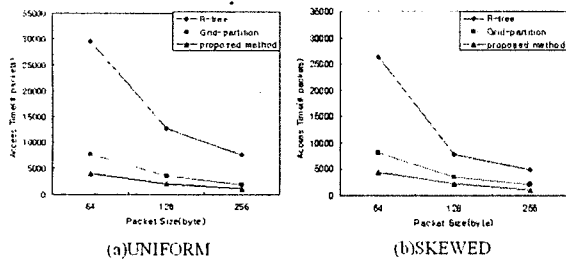


그림 5 접근 시간 측정

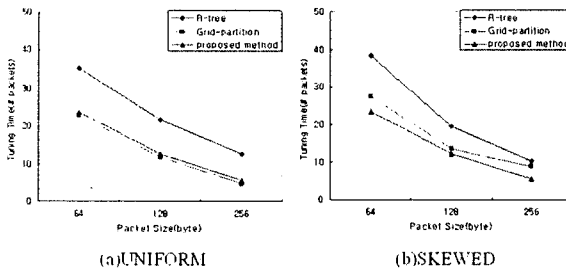


그림 6 튜닝 시간 측정

그림 5는 패킷의 크기에 따른 세 가지 기법의 접근 시간을 측정한 그래프이다. 그림에서 보는 바와 같이 R-tree가 가장 긴 접근 시간을 보였고, 제안한 기법이 패킷의 크기에 상관없이 가장 짧은 접근 시간을 나타냈다. 기존의 Grid-partition과 비교하여 평균 43%정도 감소하였다. 패킷의 크기가 64byte일 때 접근 시간이 가장 크게 감소하였으며, 패킷의 크기가 커질수록 그 감소하는 폭은 줄어든다. 제안한 기법의 경우 하나의 방송주기 이내에 탐색을 끝낼 수 있었다. Grid-partition 기법의 경우 NN 탐색을 할 경우 하나의 격자 내에서 이미 방송한 패킷을 찾을 경우가 발생하여 접근 시간이 하나 이상의 방송주기를 가진다. 특히 패킷의 크기가 작을 때, 그런 경우가 많이 발생하기 때문에 그 격차는 더 커진다.

그림 6은 패킷의 크기에 따른 세 가지 기법의 튜닝 시간을 측정한 그래프이다. 튜닝 시간 역시 R-tree가 가장 긴 시간을 보였고 uniform data의 경우엔 제안한 방법과 Grid-partition이 거의 비슷하였으나, skewed data의 경우엔 제안한 방법이

평균 20% 향상되었다. 이 경우 실제 검색하고자 하는 객체를 읽기 위해 격자 내의 이전의 패킷을 모두 읽어야 하므로, 주기가 길어질수록 시간은 크게 증가한다. 제안한 기법의 경우 검색해야하는 객체는 더 많지만, 앞에서 설명한 Grid-partition의 경우보다 읽는 패킷의 수가 적게 된다. 두 기법 사이의 격차는 패킷이 커짐에 따라서 감소한다.

5. 결론 및 향후 과제

본 논문에서는 이동 방송 환경에서 NN 탐색을 효과적으로 수행하기 위한 기법을 제안하였다. 제안 기법에서는 기존의 Grid-partition이 수행하는 NN 탐색 방법을 변경하여 접근 시간을 줄이고자 하였다. 기존의 Grid-partition의 경우에는 하나의 격자 내에서 탐색 시 해당 방송주기 이내에 탐색을 수행할 수 없는 경우가 발생하였지만, 본 제안 기법의 경우 해당 방송주기 내에 탐색을 수행할 수 있어 접근시간에 있어서 평균 43%의 성능 향상을 가져왔다.

향후 연구 과제로는 본 기법을 NN 탐색이 아닌 보다 다양한 탐색 질의에 적용하고자 한다. 즉, NN 탐색 이외의 k-NN 탐색, 여러 개의 질의점으로부터 NN을 탐색하는 GNN 탐색 등을 본 기법에 적용하고자 한다.

참고문헌

- [1] M-S. Chen, K-L. Wu, and S. Yu. Optimizing index allocation for sequential data broadcasting in wireless mobile computing. *IEEE Transactions on Knowledge and Data Engineering*, 15(1), pages:161-173, 2003.
- [2] T. Imielinski, S. Viswanathan, and B.R. Badrinath. Data on Air-Organization and Access. *IEEE Transactions on Knowledge and Data Engineering*, 9(3), pages:353-372, 1997.
- [3] N. Roussopoulos, S. Kelley, and F. Vincent. Nearest neighbor queries. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages:71-79, 1995.
- [4] J. Xu, B. Zheng, W.-C. Lee, and D. L. Lee. Energy efficient index for querying location-dependent data in mobile broadcast environments. In *Proceedings of the 19th IEEE International Conference on Data Engineering*, pages:239-250, 2003.
- [5] A. Guttman. R-trees: A dynamic index structure for spatial searching. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pages:47-54, 1984.
- [6] D. L. Lee, W.-C. Lee, J. Xu, and B. Zheng. Data management in location dependent information services. *IEEE Pervasive Computing*, 1(3), pages:65-72, 2002.
- [7] M. Berg, M. Kreveld, M. Overmars, and O. Schwarzkopf. *Computational Geometry: Algorithms and Applications*. Springer-Verlag, 1996.
- [8] B. Zheng, J. Xu, W.-C. Lee, D. L. Lee. Energy-Conserving Air Indexes for Nearest Neighbor Search. In *Proceedings of the 9th International Conference on Extending Database Technology*, pages:48-66, 2004.
- [9] Spatial Dataset. "http://www.rtreeportal.org/datasets/spatial/greece/cities_loc.zip"