

제약을 가진 최소근접을 찾는 이동질의 효율적인 수행

반재훈

경남정보대학 인터넷응용계열
e-mail:chban@kit.ac.kr

A Efficient Query Processing of Constrained Nearest Neighbor Search for Moving Query Point

Chae-Hoon Ban

Subdivision of Internet Application,
Kyungnam College of Information & Technology

요 약

This paper addresses the problem of finding a constrained nearest neighbor for moving query point (we call it CNNMP). The Nearest neighbor problem is classified by existence of a constrained region, the number of query result and movement of query point and target. The problem assumes that the query point is not static, as 1-nearest neighbor problem, but varies its position over time to the constrained region. The parameters as NC, NCMBR, CQR and QL for the algorithm are also presented. We suggest the query optimization algorithm in consideration of topological relationship among them

1. 서론

이동 컴퓨팅 환경이 급속하게 발전하면서 언제, 어디서나 사용자에게 필요한 컴퓨팅 환경을 제공하기 위해 휴대폰 또는 PDA와 같은 이동체(Moving Object)에서 제공되는 서비스에 대한 연구가 활발히 진행 중에 있다. 이러한 서비스 중에서 이동체의 위치와 관련된 서비스를 위치기반서비스(Location Based Service)라 하며 물류 및 수송관리, 교통정보, 디지털 전장과 같은 응용에서 제공된다. 위치기반서비스의 핵심 기술은 위치기반질의 처리로서 전통적인 질의처리와는 달리 질의를 요청하는 위치가 매개변수이며 시간에 따라 연속적으로 이동하므로 위치정보를 효과적으로 관리하고 검색하는 것이 필요하다.

본 논문은 시간에 따라 질의 요청 위치가 이동하는 이동체(이하 이동 질의점)를 대상으로 최소근

접질의(Nearest Neighbor Query) 처리 문제를 다룬다. 먼저, 이동 질의점에서의 제약을 가진 최소근접을 찾는 문제를 정의한다. 또한 질의처리를 위해 새로운 개념들을 정의하고 그들의 위상관계를 고려한 질의최적화 알고리즘을 제시한다. 본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 소개하고, 3장에서는 위치기반질의 중 하나인 이동 질의점에서의 최소근접을 찾는 문제를 정의한다. 4장에서는 제시하는 알고리즘에 사용되는 여러 가지 용어를 정의하며 5장에서는 제약을 가진 최소근접을 찾는 이동 질의 효율적인 수행을 위한 알고리즘을 제시한다. 마지막으로 6장에서 결론 및 향후 연구를 기술한다.

2. 관련연구

[1]은 공간데이터베이스의 인덱스인 R-Tree에서의 최소근접질의 처리하기 위한 알고리즘을 제시하였다. 질의점에서 가장 근접한 객체를 찾기 위해 R-Tree에서 실 객체의 근사치인 최소경계사각형

본 논문은 2002년도 경남정보대학 특별연구비(신진교수연구장려)에 의해 지원되었음

(MBR: Minimum Boundary Region)과 질의점 간의 MINDIST와 MINMAXDIST를 정의하고 이 개념을 이용한 가지치기(pruning) 알고리즘을 제시하였다.

[2]는 공간데이터베이스에서 제약을 가진 최소근접을 찾는 질의에 관한 연구이다. [1]의 연구는 최소근접질의 도메인이 공간데이터베이스 전체이나 이 논문의 도메인은 사용자가 정하는 임의의 영역이다. R-tree를 기반으로 하였으며 2가지 종류의 알고리즘을 제시한다. 먼저 2단계로 구성된 알고리즘인 Incremental Nearest Neighbor Search, Nearest Neighbor Search with Range Query 알고리즘을 제시하였으며 보다 향상된 단일 단계로 구성된 Integrated Method 알고리즘을 제시하였다.

[3]은 [1]과 같이 최소근접질의에 관한 연구이다. 그러나 질의점이 고정인 아닌 시간에 따라 이동하는 점이다. 이 논문에서는 이전의 질의 값을 이용하여 현재의 질의를 처리하는 Fixed Upper Bound, Lazy Search, Pre-fetching Search, Dual Buffer Search 알고리즘을 제시하여 질의 최적화를 수행한다.

3. 제약을 가진 최소근접을 찾는 이동 질의

최소근접질의는 질의점에서 가장 근접한 대상을 찾는 질의로 본 논문에서는 이 문제를 (1)질의영역의 제한 유무, (2)질의결과 수 및 (3)질의점과 질의대상의 이동 유무에 따라 분류한다. 질의영역의 제한은 질의 대상 도메인이 공간데이터베이스의 데이터 전체인지 아니면 사용자가 정하는 특정 영역인지를 의미한다. 질의결과 수의 1 또는 k이며 이 의미는 결과가 하나인지 아니면 2 이상인 k개인지를 의미한다. 질의점과 질의대상의 이동은 질의를 요청하는 위치와 그 대상이 이동하는지 아니면 고정인지를 의미한다.

[1]은 질의영역의 제한이 없으며 질의결과 수의 1 또는 k를 모두 다룬다. 또한 질의점과 질의대상은 고정이다. [2]는 질의영역의 제한이 있으며 질의결과 수의 1 또는 k를 모두 다룬다. 또한 질의점과 질의대상은 고정이다. [3]은 질의영역의 제한이 없으며 질의결과 수의 1 또는 k를 모두 다룬다. 또한 질의점은 이동이며 질의대상은 고정이다. 본 논문은 질의영역의 제한이 있으며 질의결과 수의 1, 질의점은 이동이며 질의대상은 고정을 대상으로 한다. 따라서 기존의 연구들과는 그 환경이 다르다.

그림1은 본 논문에서 제시하는 제약을 가진 최소근접을 찾는 이동 질의에 대한 예이다. 예를 들어

질의점이 q_{t-3} 에서 q_t 로 이동하였다고 가정하자. 이 경우에 질의점 q_{t-3} 과 q_{t-2} 에서의 최소근접은 d이다. 또한 질의점 q_{t-1} 과 q_t 에서의 최소근접은 a이다. 반면에 사용자의 제한 영역이 CQR로 주어지는 경우 질의점 q_{t-3} , q_{t-2} , q_{t-1} 에서의 최소근접은 c이며 질의점 q_t 에서는 b이다. 본 논문의 연구대상은 후자와 같이 질의점이 이동하고 질의 대상의 제한 영역이 주어지는 경우이다. 여기서 주목할 점은 질의점이 q_{t-3} 에서 q_{t-1} 까지 이동하더라도 그 결과는 같다는 것이다. 즉, 질의 수행을 하지 않더라도 각 객체의 위상 관계에 따라 그 결과가 동일하다는 것이다.

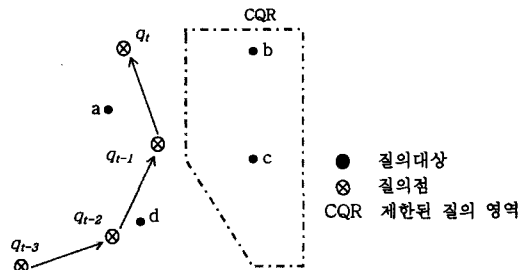


그림1 제약을 가진 최소근접을 찾는 이동 질의의 예

4. 용어 정의

4.1 NC (Nearest neighbor Cell)

점으로 구성된 집합 DB에 하나의 점을 p라고 하고 d를 점과 점 사이의 거리를 측정하는 함수라고 하자. 이때 NC(p)는 다른 점 p'보다도 점 p에서 가까운 모든 점 객체로 구성된 집합이다. 또한 모든 NC의 합이 바로 NN-Diagram이다. 이러한 NC의 특징은 NC(p) 안에서 모든 점의 최소 근접은 바로 p이다. 그림3에서 점은 NC를 나타내며 면은 NN-Diagram을 나타낸다.

$$p : \text{any point } p \in DB$$

$$d : \text{distance function } d : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$$

$$NN\text{-cell}(p) := \{x \in DB \mid \forall (p' \in DB - \{p\}) : d(x, p) \leq d(x, p')\}$$

$$NN\text{-Diagram}(DB) := \{NN\text{-Cell}(p) \mid p \in DB\}$$

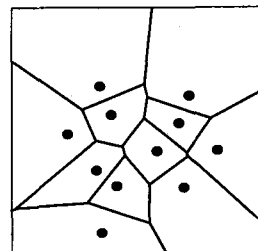


그림 2 점객체에 대한 NC와 NN-Diagram의 예

4.2 NCMBR(Nearest neighbor Cell's Minimum Boundary Region)

NCMBR은 NC를 포함하는 최소경계사각형(Minimum Boundary Box: 이하 MBR)이다. 최소경계사각형이라 어떤 도형을 포함하는 사각형 중에서 그 넓이가 가장 작은 사각형을 말한다. 그림 4의 왼쪽은 NN-Diagram에서의 NCMBR이다. 각 NC를 포함하는 최소 영역의 사각형이 바로 NCMBR이며 이러한 NCMBR만을 나타낸 그림이 오른쪽 그림이다.

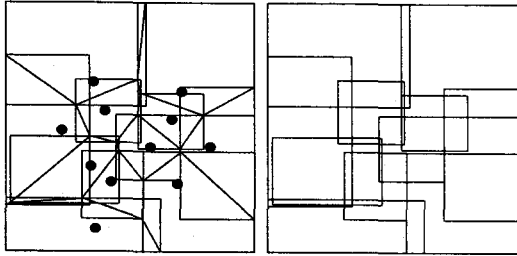


그림4 NN-Diagram에서의 NCMBR

4.3 CQR(Constraint Query Region)

공간데이터베이스의 특정 응용에서 질의의 도메인을 제한하는 경우가 필요하다. 이러한 제한은 영역인 경우가 대부분이며 질의 수행을 위한 제한된 영역을 바로 CQR로 정의한다. 그림 5는 CQR의 예이다.

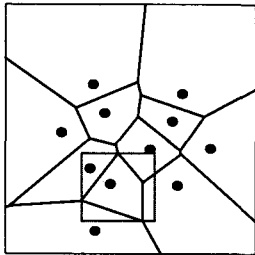


그림5 점 객체에 대한 제한된 질의 영역 CQR

4.4 QL(Query Line)

q_t 를 질의가 발생한 시간 t 에서의 질의점이라고 하자. 이 질의점은 시간 t 가 변하면서 그 위치도 변할 수 있다. 이전 질의 시간 $t-1$ 에서의 질의점 q_{t-1} 을 시작점으로하고 현재 질의 시간 t 에서의 질의점 q_t 를 끝점으로하여 선분을 그을 수 있으며 이 선분을 QL로 정의한다. 즉 QL은 이전 질의점과 현재 질의점을 양끝점으로하는 선분이다. 그림 6은 QL에 대한 예이다.

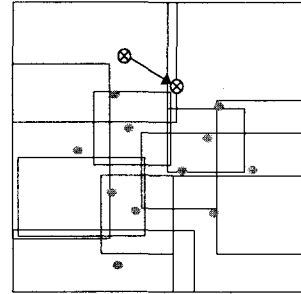


그림6 이전 질의점과 현재 질의점으로 구성된 QL

5. 제약울 가진 최소근접을 찾는 이동 질의 알고리즘

제약을 가진 최소근접을 찾는 이동 질의의 수행은 질의가 발생할 때마다 모든 객체를 대상으로 제약 영역 내에 존재하는 객체를 찾고, 다시 그 대상으로 질의점과 대상 객체의 거리를 계산하여 그 최소를 획득하는 방법으로 수행될 수 있다. 보다 나은 방법은 앞에서 정의한 NC를 이용해 인덱스를 구축하고 제약 영역과 겹치는 NC의 실 객체를 대상으로 거리를 계산하여 최소를 획득하는 것이다. 물론 [2]에서 제시하는 방법을 적용할 수 있다. 그러나 질의가 발생할 때마다 질의 수행을 해야 하는 단점이 발생한다. 본 논문에서는 위상 관계를 고려하여 이전 질의의 결과를 이용해 빠르게 질의를 수행하는 질의 최적화 알고리즘을 제시한다.

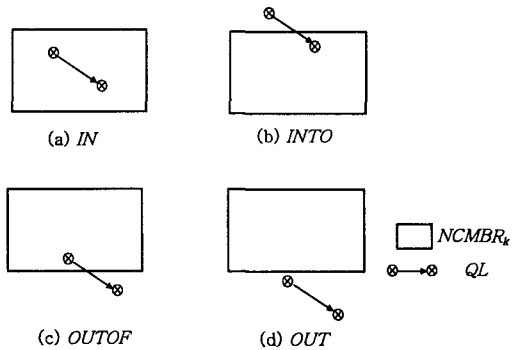


그림7 CQR과 QL의 가능한 위상 관계

먼저 $NCMBR_k$ 를 질의점 q_{t-1} 에서 최소근접인 점 k 의 $NCMBR$ 이라고 정의하자. 정의된 $NCMBR_k$ 와 QL의 위상관계에 따라 그림 7과 같이 질의 유형은 4가지로 분류된다. 7-(a)는 QL이 $NCMBR_k$ 에 완전히 포함된 상태로 본 논문에서는 IN이라고 정의한다. 7-(b)는 QL의 시작점은 $NCMBR_k$ 의 외부에 포함되고 끝점은 내부에 포함되는 상태로 INTO라 정

의한다. 마찬가지로 7-(c)는 OUTOF, 7-(d)는 OUT 이라고 정의한다.

다음은 각 위상관계에 따른 4가지 경우에 대하여 이전 질의 결과를 이용하여 제약을 가진 최소근접을 찾는 이동 질의의 최적화 알고리즘은 다음과 같다.

Case : IN

$$\text{if } QL \cap \bigcap_{i=1, i \neq k} NCMBR_i = \emptyset$$

$$\text{then } CNNMP(q, DB) = p$$

else 재계산

Case : INTO

$$\text{if } (NCMBR_k \cap QL) \cap \bigcap_{i=1, i \neq k} NCMBR_i = \emptyset$$

$$\text{then } CNNMP(q, DB) = p$$

else 재계산

Case : OUT

Case : OUTOF

재계산

IN은 이전 질의점과 현재 질의점으로 구성된 QL 이 NCMBR_k에 포함되어 있고 다른 NCMBR들과 겹치지 않는 경우로 질의의 결과가 이전과 동일하다. 만약 다른 NCMBR과 겹친다면 다른 객체의 NC에 포함되어질 수 있으며 그 경우에 질의 결과는 그 NC의 실 객체가 되어 이전의 질의 결과와 달라질 수 있다. 따라서 이 경우에는 기존 알고리즘에 따라 재계산을 수행한다.

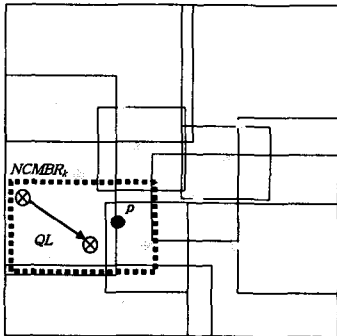


그림7 IN의 알고리즘과 그 예

INTO는 QL의 끝점이 NCMBR_k에 포함되어 있고 시작점은 포함되어 있지 않으며 NCMBR_k에 포함된 QL의 부분이 다른 NCMBR과 겹치지 않는 경우로 질의의 결과는 이전과 동일하다. 만약 QL의 부분이

다른 NCMBR과 겹친다면 다른 객체의 NC에 포함 되어질 수 있으며 그 경우에 질의 결과는 그 NC의 실 객체가 되어 이전의 질의 결과와 달라질 수 있다. 따라서 이 경우에는 기존 알고리즘에 따라 재계산을 수행한다.

마지막으로 OUTOF와 OUT의 경우는 재계산을 수행한다. OUTOF의 경우는 QL의 시작점이 NCMBR_k의 내부에 포함되고 끝점이 NCMBR_k의 외부에 포함되는 경우이다. 이 의미는 현재의 질의점이 다른 NCMBR에 포함되어 있다는 말이며 그 경우에는 질의의 결과가 달라질 수 있다는 의미이므로 이전의 질의 결과를 이용할 수 없다. 따라서 앞에서 언급한 방법으로 재계산을 수행한다. 마찬가지로 OUT의 경우도 현재의 질의점이 다른 NCMBR에 포함되어 있기 때문에 질의의 결과가 달라질 수 있으므로 재계산을 수행한다.

6. 결론 및 향후 연구

본 논문에서는 시간에 따라 이동하는 이동 질의 점을 대상으로 제약을 가진 최소근접을 찾는 질의 최적화 방안을 제시하였다. 먼저, 최소근접을 찾는 문제를 질의영역의 제한 유무, 질의결과 수 및 질의점과 질의대상의 이동 유무에 따라 분류하고 알고리즘에서 사용되는 여러 개념인 NC, NCMBR, CQR 및 QL을 정의하였다. 마지막으로 제시된 여러 정의들 간의 위상관계를 고려하여 질의 최적화 알고리즘을 제시하였다.

앞으로 실제 구현 및 실험을 통하여 알고리즘의 성능을 평가하고자 한다. 또한 보다 많은 위상 관계를 도출하여 알고리즘을 개선하며 질의 결과의 수를 k로 확장하고 질의 대상과 질의 영역이 동적으로 변하는 환경에서 적용되는 알고리즘의 연구가 필요하다.

참고문헌

- [1] N. Roussopoulos, S. Kelley, F. Vincent, "Nearest Neighbor Queries", *ACM SIGMOD International Conference in Management of Data*, 1995
- [2] H. Ferhatosmanoglu, I. Stanoi, D. Agrawal, A. E. Abbadi, "Constrained Nearest Neighbor Queries", *Symposium on Spatial and Temporal Databases(SSTD)*, 2001
- [3] Z. Song, N. Roussopoulos, "K-Nearest Neighbor Search for Moving Query Point", *Symposium on Spatial and Temporal Databases(SSTD)*, 2001