

비공간 애트리뷰트를 포함한 공간 인덱스

오 병 우*

Spatial Index Including Non-Spatial Attributes

Byoung-Woo Oh*

이 연구는 금오공과대학교 대학 학술연구비로 지원되었음 (2021년)

요 약

디지털 트윈, 도시 계획, 시설물 관리 등과 같은 응용 프로그램들은 기하 좌표와 부가적인 비공간 애트리뷰트로 구성된 공간 데이터를 사용한다. 기존의 공간 인덱스는 기하 좌표를 검색하므로 비공간 애트리뷰트를 위한 별도의 필터링 과정이 필요하다. 본 논문에서는 비공간 애트리뷰트를 공간 인덱스에 통합하여 별도의 필터링 단계를 제거하고 검색 효율을 향상시키는 NR*-tree(Non-spatial and Spatial Region Star Tree)를 제안한다. 수치형 비공간 애트리뷰트는 MBR(Minimum Bounding Rectangle)의 차원을 확장하여 범위를 저장하고, 카테고리형 비공간 애트리뷰트는 단일 값을 MBR에 저장하여 리프 노드에서 필터링한다. 실험을 통해 NR*-tree는 기존의 R*-tree(Region Star Tree)에 비해 최대 86.8%의 검색 시간이 감소하는 결과를 얻었다. 검색 시 장축의 길이를 MBR에 포함시켜 검색에 활용함으로써 지도 가독성은 향상시키고 검색 성능은 더욱 높일 수 있었다.

Abstract

Applications such as digital twins, urban planning, and facility management use spatial data consisting of geometric coordinates and additional non-spatial attributes. Traditional spatial indexes search for geometric coordinates, requiring a separate filtering process for non-spatial attributes. This paper proposes NR*-tree(Non-spatial and Spatial Region Star Tree), integrating non-spatial attributes into a spatial index to eliminate the filtering step and enhance search efficiency. Numerical non-spatial attributes extend the dimensions of MBRs(Minimum Bounding Rectangles) to store ranges, while categorical non-spatial attributes store single values in MBRs for filtering at leaf nodes. Experiments show that the NR*-tree reduces search time by up to 86.8% compared to the conventional R*-tree (Region Star Tree). Including the length of the major axis in MBRs during searches enhances map readability and further improves search performance.

Keywords

spatial index, non-spatial attributes, R*-tree, NR*-tree

* 국립금오공과대학교 컴퓨터공학과 교수
- ORCID: <https://orcid.org/0000-0003-4700-9961>

• Received: Jan. 08, 2024, Revised: Jan. 23, 2024, Accepted: Jan. 26, 2024
• Corresponding Author: Byoung-Woo Oh
Dept. of Computer Engineering
Kumoh National Institute of Technology, South Korea
Tel.:+82-54-478-7531, Email: bwoh@kumoh.ac.kr

I. 서 론

공간 데이터베이스 시스템에서 공간 인덱스는 다양한 점, 선, 면 형태의 공간 데이터의 기하 데이터를 효과적으로 처리하는 데 중요한 역할을 담당한다. 공간 인덱스에서는 일반적으로 2차원 평면상의 x 와 y 좌표로 구성된 기하 좌표를 처리하므로 2차원 데이터를 저장하고 관리한다.

공간 데이터는 기하 좌표의 값뿐만 아니라 특성을 나타내는 비공간 애트리뷰트 값들도 가지고 있다. 예를 들어, POI(Point of Interest)를 위한 한 개의 공간 데이터는 2차원 평면상의 점 기하 좌표 (x, y) 값과 관심 지점의 비공간 애트리뷰트 “명칭”의 값인 문자열로 구성된다. 디지털 트윈, 도시 계획, 시설물 관리 등 다양한 응용 분야에서 공간 데이터를 실제 활용할 때는 기하 좌표뿐만 아니라 비공간 애트리뷰트도 중요한 역할을 담당한다.

본 논문에서는 기존의 공간 인덱스에서 기하 좌표를 가지고 검색한 뒤에 비공간 애트리뷰트로 필터링하는 과정을 통합하여, 공간 인덱스에 비공간 애트리뷰트를 포함해 검색함으로써 효율을 향상시킬 수 있는 방법을 제안한다.

본 논문의 구성은 다음과 같다. 2장은 관련연구로서 R*-tree(Region Star Tree)와 비공간 애트리뷰트를 활용한 연구를 살펴본다. 3장에서는 본 논문에서 제안하는 비공간 애트리뷰트를 포함한 공간 인덱스인 NR*-tree(Non-spatial and Spatial Region Star Tree)에 대해 기술한다. 4장에서는 제안한 NR*-tree의 성능 평가에 대한 실험 결과를 제시한다. 5장에서는 결론과 향후 연구 방향에 대하여 기술한다.

II. 관련 연구

2.1 R*-tree

R*-tree는 공간 데이터베이스에서 많이 사용되는 공간 인덱스 중 하나로서 R-tree의 성능을 향상시키기 위해 여러 가지 최적화 기법을 사용한다[1]. 점, 선, 면의 다양한 형태와 크기의 다차원 공간 데이터를 저장 및 관리할 수 있는 효율적인 인덱스 구조

이다. 공간 인덱스 R-tree 및 R*-tree는 n 개의 축으로 구성된 다차원 공간 데이터를 처리할 수 있도록 설계되었지만, 일반적으로 사용되는 공간 데이터는 x 축 및 y 축의 (x, y) 기하 좌표를 사용하므로 주로 2차원 데이터를 가지고 인덱스를 구축한다.

본 논문에서 제안하는 NR*-tree는 공간 좌표를 표현하는 기하 2차원 데이터뿐만 아니라 검색에 사용할 수 있는 비공간 데이터도 추가로 포함하여 인덱스를 구축한다.

2.2 비공간 애트리뷰트 활용

공간 데이터는 지도상에 출력되는 모양을 점, 선, 면 형태로 저장하기 위하여 기하 좌표의 값들을 가지고 있고, 해당 공간 데이터의 특성 값을 저장하기 위하여 비공간 애트리뷰트들을 가지고 있다. 비공간 애트리뷰트들은 공간 데이터를 분석하고 활용하는데 중요한 역할을 담당한다.

다양한 연구들에서 비공간 애트리뷰트를 활용한 인덱스 구조를 제안하고 있다. NSB-tree는 R-tree를 일반적인 비공간 애트리뷰트에 적용하여 제안한 구조이다[2]. Lattice-tree는 R-tree와 Lattice 구조를 혼합하여 공간 위치와 텍스트 키워드와 관계를 조직화하는 구조이다[3]. AR*-tree는 공간 인덱스의 MBR에 MBR의 면적을 포함하여 성능을 향상시킨 구조이다[4].

기하 좌표뿐만 아니라 비공간 애트리뷰트를 같이 분석하는 연구들도 다양하게 수행되고 있다[5][6]. 공간 질의에 비공간 애트리뷰트에 속하는 선호도를 사용하는 연구도 수행되고 있다. 공간적인 위치뿐만 아니라 비공간 애트리뷰트인 선호도에 기반하여 가중치를 할당하여 검색하는 연구[7]와, 선호도를 동맹적 선호도와 적대적 선호도로 나눠서 평가하여 검색하는 연구가 있다[8].

본 논문에서는 공간 인덱스의 질의 처리 성능을 향상시키기 위하여 비공간 애트리뷰트를 활용한 NR*-tree를 제안한다.

III. 공간 인덱스에 비공간 애트리뷰트 저장

본 장에서는 기존의 공간 인덱스 구조에 비공간 애트리뷰트를 포함하여 검색 효율을 높일 수 있는 인덱스 구조인 NR*-tree를 제안한다.

3.1 공간 데이터의 기하 좌표 및 MBR

공간 데이터는 기하 좌표의 값과 비공간 애트리뷰트의 값으로 구성된다. 공간 데이터의 기하 좌표 데이터는 타입에 따라서 점 (point), 선 (polyline), 면 (polygon)이 될 수 있다. 점 타입은 일반적으로 2차원 기하 좌표 한 개만 저장하므로 공간 데이터 중에서는 처리가 쉬운 편에 속한다. 반면에 선과 면 타입은 여러 개의 기하 좌표의 값들을 저장하는 가변 길이 데이터로서 구조가 복잡하여 처리하기에 어려움이 따른다.

공간 데이터의 복잡한 기하 좌표들을 처리하기 쉽게 단순화하기 위하여 해당 공간 데이터의 기하 좌표들을 모두 포함하는 가장 작은 사각형인 MBR (Minimum Bounding Rectangle)을 사용한다. 2차원 평면에서 사각형은 (minX, minY) 좌표와 (maxX, maxY) 좌표 두 개만을 사용하여 표현할 수 있다. 공간 인덱스에서는 복잡하고 자세한 공간 데이터의 기하 좌표들 대신에 단순화된 MBR을 사용하여 인덱스를 구축하고 검색에 사용한다.

3.2 비공간 애트리뷰트 저장 방법

기존 R*-tree는 공간 데이터의 MBR을 x축과 y축의 2차원 데이터로 구성하여, 공간 데이터의 효율적인 관리를 제공한다. 제안하는 NR*-tree에서 비공간 애트리뷰트를 MBR에 저장하는 방법은 크게 두 가지로 나뉜다.

3.2.1 수치형 애트리뷰트는 값의 범위 저장

수치형 비공간 애트리뷰트를 공간 인덱스에 포함하기 위해서는 MBR의 차원을 확장하여 범위를 저장한다. 수치형 데이터에 속하는 연속형 데이터와 이산형 데이터 모두 범위가 있으므로 MBR의 차원을 확장하여 범위를 저장한다.

MBR에 차원을 확장하여 비공간 애트리뷰트 값의 범위를 저장한다는 것은 다음을 의미한다. 인덱스를 구축할 때 트리 구조로 영역을 분할하는데, n개의 비공간 애트리뷰트 축도 공간 데이터의 2차원 축처럼 똑같이 고려되어 2 + n 차원의 MBR을 가지고 영역 분할 알고리즘이 적용된다는 의미이다. 또한, 검색할 때도 질의에서 주어진 2 + n 차원의 MBR을 가지고 검색 알고리즘이 적용되는 것을 의미한다.

3.2.2 카테고리형 애트리뷰트는 값 저장

카테고리 데이터는 범위가 큰 의미가 없으므로 비공간 애트리뷰트의 단일 값을 MBR에 저장한다. 공간 인덱스를 검색할 때 카테고리 데이터는 중간 노드에서는 검색에 관여하지 않고, 리프 노드에서만 질의에서 주어진 값과 일치하는지 비교하여 결과에 포함할지 여부를 결정한다.

3.3 MBR 구조

기존 R*-tree에서 2차원 평면상의 공간 데이터를 표현하는 MBR은 그림 1 (a)와 같고, 본 논문에서 제안하는 NR*-tree에서 n개의 수치형 비공간 애트리뷰트와 m개의 카테고리형 비공간 애트리뷰트를 추가로 포함하는 MBR은 그림 1 (b)와 같다.

MBR of R*-tree	MBR of NR*-tree
minX, minY maxX, maxY	min[2 + n] max[2 + n] value[m]

(a) 기존 R*-트리 (b) 제안된 NR*-트리
(a) Traditional R*-tree (b) Proposed NR*-tree

그림 1. MBR 구조
Fig. 1. MBR structure

R*-tree에서 2차원 평면상의 공간 데이터를 표현할 때, 본 논문에서 제안하는 NR*-tree에서 MBR의 min[0]과 min[1]은 각각 R*-tree의 minX와 minY에 해당하고, max[0]과 max[1]은 각각 R*-tree의 maxX, maxY에 해당한다.

R*-tree가 다차원 데이터를 처리할 수 있으므로, 제안하는 NR*-tree도 수치형 비공간 애트리뷰트의 범위를 가지고 동일한 알고리즘으로 삽입 및 삭제할 수 있다. m개의 카테고리형 비공간 애트리뷰트는 삽입 및 삭제 알고리즘에서는 사용되지 않는다.

본 논문에서 제안하는 NR*-tree는 기존의 R*-tree에 비공간 애트리뷰트를 포함하여 검색에 사용한다. 검색시 MBR에는 $2 + n$ 차원의 범위를 가지고 있더라도, 질의에서는 $2 + n$ 차원보다 적은 차원의 MBR을 줄 수 있다.

3.4 장축의 길이 활용

본 논문에서 제안하는 NR*-tree에 공간 데이터의 x축 길이와 y축 길이 중에서 길이가 더 긴 장축의 길이를 비공간 애트리뷰트로 MBR에 저장하여 검색에 활용할 수 있다. 장축의 길이는 지도를 출력하기 위해 특정 질의 영역에 속한 공간 데이터를 검색할 때 출력 화면 기준으로 충분히 작은 데이터를 필터링하는 데 사용된다. 충분히 작은 공간 데이터를 필터링함으로써 렌더링 시간은 낮추고 복잡도는 낮춰서 지도의 가독성을 높일 수 있다.

관련 연구에서 살펴본 [4] 연구에서는 공간 데이터의 MBR의 면적을 사용하여 필터링하여 렌더링의 성능을 향상시켰다. 수평선이나 수직선은 면적이 0인 문제점이 있고, 같은 면적을 가지고 있더라도 지도 출력에서는 길이가 더욱 중요한 의미를 가지고 있다. 예를 들어, 면적이 20으로 같을 때, { x축 길이, y축 길이 }가 각각 { 5, 4 }인 경우와 { 20, 1 }인 경우에 지도에서는 길이가 더 긴 { 20, 1 }인 공간 데이터가 출력시 중요도가 더욱 높다. 그러므로, 본 논문에서는 장축의 길이를 비공간 애트리뷰트로 MBR에 추가하여 렌더링 성능과 가독성을 높일 수 있는 방법을 제안한다.

IV. 실험 및 성능 평가

기존의 R*-tree와 본 논문에서 제안하는 비공간 애트리뷰트를 포함한 공간 인덱스의 성능을 비교한다. 실험은 실제 공간 데이터셋을 활용하여 실시되

었으며, 제안하는 인덱스가 기존의 R*-tree에 비해 어느 정도의 성능 향상을 제공하는지 정량적으로 평가하였다.

4.1 실험 데이터셋

실험에서 사용된 데이터셋은 국가공간정보포털에서 제공하는 1:5000 축척의 수치지형도 v2.0 데이터 중에서 서울에 해당하는 공간 데이터셋이다. 파일 이름은 “NGII_DTM_V2_5000_서울_강남구.zip”와 같이 서울의 25개 행정자치구에 대해 25개의 압축 파일들로 제공된다. 각 압축 파일에는 해당 구에 포함된 도엽들을 포함한다. 각 도엽은 수치지도 지형지물 표준코드들에 따라 Shape 파일 포맷으로 저장되어 있다. 서울 지역은 총 137개 도엽의 4,512개의 Shape 파일로 구성되며 전체 용량은 1.66GB이다. 전체 공간 데이터는 2,130,239개이다.

실험 데이터셋을 가지고 생성된 R*-tree의 인덱스 파일 크기는 89,758,260 bytes이며, NR*-tree의 인덱스 파일은 128,605,760 bytes이다. 제안하는 NR*-tree는 MBR에 비공간 애트리뷰트의 값들을 포함하고 있으므로 기존 R*-tree보다 43.3% 더 많은 저장 용량을 사용한다. 포함하는 비공간 애트리뷰트가 많아 질수록 저장 용량은 증가한다.

본 논문에서 제안하는 NR*-tree의 성능을 평가하기 위해서 공간 인덱스에 포함된 수치형 비공간 애트리뷰트는 MBR의 장축의 길이이고, 카테고리형 비공간 애트리뷰트는 수치지도 지형지물 표준코드이다. NR*-tree의 MBR은 { {minX, maxX}, {minY, maxY}, {minL, maxL}, valueCode } 구조(n=1, m=1)를 갖는다.

4.2 성능 평가

기존의 R*-tree와 본 논문에서 제안한 NR*-tree의 성능을 비교 평가하기 위하여, 서울 전체 수치지도에 대해 정사각형 질의 영역의 길이를 500m, 1Km, 2Km, 4Km, 8Km, 16Km, 32Km로 변경하면서 7개 질의에 대한 검색 시간을 측정한다.

질의 영역에 사용한 수치형 비공간 애트리뷰트는 장축의 길이의 최솟값으로서, 1,000 pixels * 1,000 pixels 이미지로 렌더링할 때 2 pixels에 해당하는 길이를 계산하여 지정한다. 질의 영역에 사용한 카테고리형 비공간 애트리뷰트 값은 교통의 “도로경계(미분류)”, “교량”, “교차로”, “입체교차부”, “인터체인지”, 모든 건물, 모든 시설, 수계의 “하천경계”, “호수/저수지”, 지형의 등고선 중에서 “계곡선”이다.

랜덤 함수를 사용하여 임의로 100개의 질의 영역을 생성하여 검색 시간을 측정하고 평균을 계산한다. 실험 환경은 AMD Ryzen9 3900X 12-Core 3.80GHz CPU, 32GB 메모리, Windows 10 Pro. 운영체제를 사용하였다. 표 1은 실험 결과를 보여준다.

표 1. 검색 시간에 대한 실험 결과 (ms)
Table 1. Experimental results for search time (ms)

query length	R*-tree	proposed NR*-tree
500m	0.585	0.413
1,000m	1.302	0.697
2,000m	4.325	1.783
4,000m	16.639	5.751
8,000m	71.345	18.912
16,000m	300.885	60.535
32,000m	929.236	122.929

NR*-tree는 R*-tree에 비해 500m에서는 70.6%, 32Km에서는 13.2%의 시간만으로 검색할 수 있어서 우수한 성능을 보여준다. 표 1의 실험 결과를 그래프로 표현하면 그림 2와 같다.

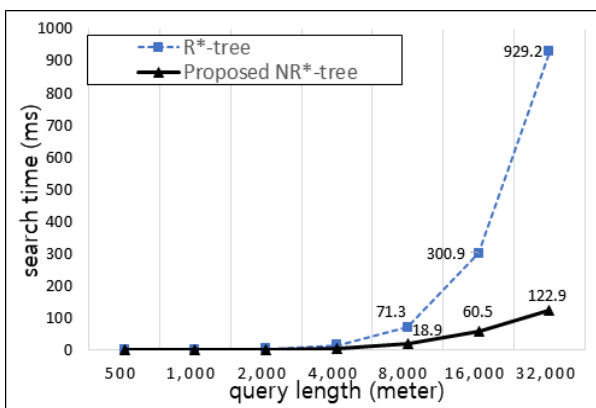


그림 2. 성능 평가 그래프
Fig. 2. Performance evaluation graph

그림 2에서 보는 바와 같이 기존의 R*-tree과 비교하여 제안한 NR*-tree의 성능은 질의 영역이 커질수록 더욱 우수한 성능을 발휘함을 알 수 있다.

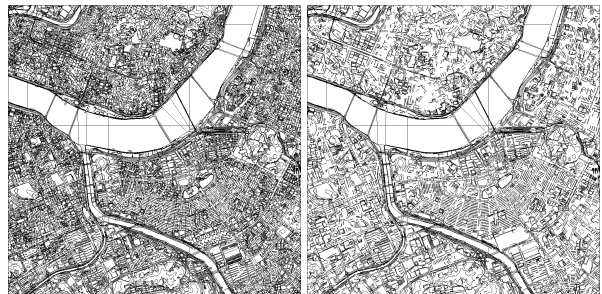
4.3 NR*-tree 장축 필터링

그림 3은 정사각형 질의 영역의 길이가 8Km인 경우에 장축의 길이를 적용하지 않을 때(a)와 화면 상에서 각각 2, 4, 8 pixels에 해당하는 장축의 길이를 적용했을 때의 1,000 pixels * 1,000 pixels 결과 이미지를 보여준다.

장축의 길이가 2 pixels보다 작은 MBR을 검색 결과에서 필터링한 그림 3 (b)와 4 pixels로 필터링한 그림 3 (c)는 장축의 길이를 적용하지 않은 그림 3 (a)에 비해 상세함이 부족하지 않고 오히려 가독성이 높은 결과를 보여준다. 검색에 소요되는 시간도 약 20% 정도씩 줄어든다. 8 pixels로 필터링한 그림 3 (d)는 다소 상세함이 부족하긴 하지만, 2 pixels로 필터링한 검색 시간보다 약 40% 정도 감소하므로 필요에 따라 사용할 수도 있다.



(a) 장축 길이 필터링 없음 (a) minPixels = 0
(b) 장축 길이 2픽셀 필터링 (b) minPixels = 2



(c) 장축 길이 4픽셀 필터링 (c) minPixels = 4
(d) 장축 길이 8픽셀 필터링 (d) minPixels = 8

그림 3. 장축 필터링 결과

Fig. 3. Major axis filtering results

V. 결 론

디지털 트윈, 시설물 관리, 지도 등에서 광범위하게 사용되는 공간 데이터는 모양을 표현하기 위한 기하 좌표의 값과 세부 특성을 저장하기 위한 비공간 애트리뷰트의 값으로 구성된다. 본 논문에서는 MBR로 단순화된 기하 좌표를 검색하기 위한 기존의 공간 인덱스에 비공간 애트리뷰트를 포함하여 검색 성능을 향상시키는 인덱스 구조인 NR*-tree를 제안하였다.

NR*-tree를 사용하면 공간 검색 이후에 비공간 애트리뷰트로 필터링하는 과정을 통합하여 한 번의 검색으로 수행할 수 있어서 검색 시간이 86.8%까지 감소하는 것을 실험을 통해 확인하였다. 그리고, 렌더링할 때 충분히 작은 공간 데이터를 필터링할 수 있도록 장축의 길이를 MBR에 포함하여 검색할 때 사용하여 검색 성능을 더욱 향상하고 지도의 가독성도 높일 수 있었다.

공간 인덱스에서 질의 영역에 대한 검색의 결과를 얻은 이후에, 화면에 출력하기 위한 렌더링 과정에서 필요한 비공간 애트리뷰트 값(예를 들면, 건물의 명칭, 등고선의 고도 등)을 효율적으로 접근하기 위한 구조 설계와 공간 인덱스와의 연계 방법에 대한 연구가 향후 연구 방향이 될 수 있다.

References

- [1] N. Beckmann, H. P. Kriegel, R. Schneider, and B. Seeger, "The R*-tree: An efficient and robust access method for points and rectangles", In Proc. of the 1990 ACM SIGMOD, pp. 322-331, May 1990. <https://doi.org/10.1145/93597.98741>.
- [2] S. Harikumar and A. Vinay, "NSB-TREE for an efficient multidimensional indexing in non-spatial databases", In 2013 IEEE RAICS, pp. 80-85, Dec. 2013. <https://doi.org/10.1109/RAICS.2013.6745451>.
- [3] A. Xu, Z. Zhang, X. Ma, Z. Zhang, and T. Xu, "Spatial Concept Query Based on Lattice-Tree", ISPRS International Journal of Geo-Information, Vol. 11, No. 5, pp. 312, May 2022. <https://doi.org/10.3390/ijgi11050312>.

- [4] B. W. Oh, "Efficient Spatial Index for Mobile Software", Spatial Information Research, Vol. 16, No. 1, pp. 113-127, Apr. 2008.
- [5] D. Murakami and D. A. Griffith, "Balancing spatial and non-spatial variation in varying coefficient modeling: A remedy for spurious correlation", Geographical Analysis, Vol. 55, No. 1, pp. 31-55, Nov. 2023. <https://doi.org/10.1111/gean.12310>.
- [6] Y. Li, J. A. He, S. Yu, H. Wang, and S. Ye, "Spatial structures of different-sized tree species in a secondary forest in the early succession stage", European Journal of Forest Research, Vol. 139, pp. 709-719, Apr. 2020. <https://doi.org/10.1007/s10342-020-01280-w>.
- [7] M. L. Yiu, H. Lu, N. Mamoulis, and M. Vaitis, "Ranking spatial data by quality preferences", IEEE Transactions on Knowledge and Data Engineering, Vol. 23, No. 3, pp. 433-446, Jul. 2010. <https://doi.org/10.1109/TKDE.2010.119>.
- [8] Q. Qu, S. Liu, B. Yang, and C. S. Jensen, "Integrating non-spatial preferences into spatial location queries", In Proc. of the 2014 SSDBM, pp. 1-12, Jun. 2014. <https://doi.org/10.1145/2618243.2618247>.

저자소개

오 병 우 (Byoung-Woo Oh)



1993년 2월 : 건국대학교
전자계산학과(공학사)

1995년 2월 : 건국대학교
전자계산학과(공학석사)

1999년 2월 : 건국대학교
전자계산학과(공학박사)

1999년 6월 ~ 2004년 2월 :

한국전자통신연구원 텔레매틱스연구단 선임 연구원
2004년 3월 ~ 현재 : 국립금오공과대학교 컴퓨터공학과
교수

관심분야 : 공간 데이터베이스, GIS, 위치 기반 서비스