

# 효과적인 근사 k-최근접 분산 처리를 위한 질의 할당 기법

## Query Allocation Method for Efficient Distributed Processing of an Approximate k-Nearest Neighbor Query

최도진, 임종태, 북경수, 유재수†  
충북대학교

Do-Jiin Choi, Jong-Tae Lim, Kyoung-Soo Bok,  
Jae-Soo Yoo †  
Chungbuk National Univ.

### 요약

모바일 기기의 대중화 및 위치 인식 기술의 발달로 다양한 위치 기반 서비스가 제공되고 있다. 많은 위치 기반 서비스에서는 현재 위치에서 가장 가까운 k개의 아이템을 찾는 k-최근접 질의가 빈번하게 활용되고 있다. 본 논문에서는 효율적인 k-최근접 분산 질의 처리를 질의 할당 기법을 제안한다. 질의 처리 할당을 위해 질의 통계 값을 활용한 질의 모형을 정의하고 규칙 기반의 질의 할당을 수행한다. 성능 평가를 통해 제안하는 기법의 우수성을 보인다.

## I. 서론

스마트폰, 테블릿PC 등과 같은 모바일 기기의 대중화와 함께 위치 인식 기술의 발달로 다양한 위치 기반 서비스가 제공되고 있다. 이러한 위치 기반 서비스에서 현재 위치를 기반으로 가장 가까운 k개의 객체를 찾는 k-최근접 질의 처리가 빈번하게 활용되고 있다[1].

최근 위치 기반 서비스 사용자의 증가로 인해 대용량의 객체를 대상으로 k-최근접 질의 처리를 효과적으로 수행하기 위한 연구가 진행되고 있다[2-3]. 기존 k-최근접 질의 처리 방식은 분산 처리 방식과 근사 접근 방식으로 구분된다. 분산 처리 방식은 다수의 노드에 객체들의 정보를 저장하고 색인하여 질의 처리의 시간을 감소시킨다. 근사 k-최근접 질의 처리 방법은 k개의 정확한 검색 결과를 제공하지 않고 k개에 근접한 질의 결과를 생성하여 빠른 응답성을 보장한다. [3]에서는 빠른 응답성을 보장하기 위해 근사 접근법의 분산 처리 방법을 제안하였다. 그러나 각 노드에 저장된 객체 정보를 이용하여 정해진 규칙에 따라 질의를 분배하기 때문에 특정 노드에 부하가 집중되거나 분산 처리의 효율성이 감소한다.

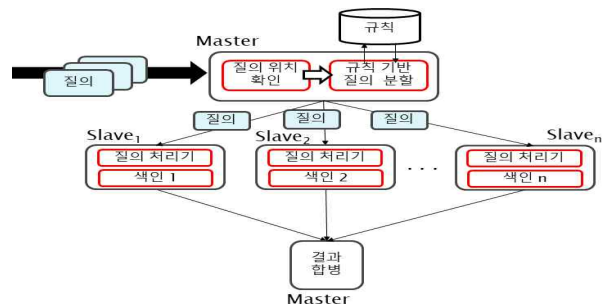
본 논문에서는 그리드 기반 색인을 이용하여 분산 근

사 k-최근접 질의 처리를 효과적으로 수행하기 위한 질의 할당 기법을 제안한다. 질의 할당 기법을 위해 분산 그리드 색인에서의 축 층들을 이용한 질의 할당 규칙을 정의한다. 이렇게 정의된 규칙에 맞는 질의 할당을 수행하기 위해 질의 결과의 통계 값을 활용하여 질의 모형을 생성하고, 질의 모형을 이용하여 질의 할당 규칙에 맞는 규칙에 따라 질의를 분산 할당한다.

## II. 제안하는 질의 할당 기법

### 1. 시스템 구조

분산 환경에서 각 노드가 처리해야 할 질의 처리량을 정하기 위해서는 질의 할당 규칙이 필요하다. 부하가 집중 되는 기존 질의 할당 규칙에서의 문제점을 해결하기 위해 질의와의 연관성이 부족한 노드에는 적은 질의를 할당하는 질의 할당 기법이 필요하다.



▶▶ 그림 1. 제안하는 시스템 구조도

제안하는 기법의 전체적인 구조는 그림 1과 같다. 질

† 교신 저자 : yjs@cbnu.ac.kr

본 연구는 2014년도 산업통상자원부의 재원으로 한국에너지기술연구원(KETEP)의 에너지인력양성사업(No. 20164030201330)과 2016년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원(No. 2016R1A2B3007527)과 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 [No.B0101-15-0266, 실시간 대규모 영상 데이터 이해-예측을 위한 고성능 비주얼 디스커버리 플랫폼 개발

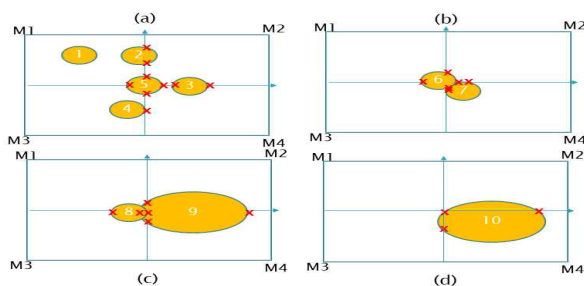
의 대상이 되는 데이터는 그리드 색인 형태로 분할되어 각 슬레이브에 저장되어있다. 질의는 마스터를 통해 입력이 되고 질의의 위치 정보와 분할된 색인정보를 바탕으로 규칙 기반의 질의 분할을 수행한다. 규칙은 분산 색인에서 질의 처리의 결과를 많이 생성하지 못할 것 같은 노드는 적은 질의량을 할당하여 질의 처리의 부하를 분산시킨다.

## 2. 규칙 기반 질의 분할

질의 분할을 위한 규칙은 그리드 분산 색인에서 질의가 분할된 x,y 축에 어떻게 충돌을 했는지에 따라 다르게 정의한다. 분할 규칙에 따라 질의 영역의 포함도에 따라서 질의량을 조절하여 각 노드에 질의 할당을 수행하여 노드의 부하를 분산시킬 수 있다. 규칙은 분할되는 x,y 축을 활용하여 총 10가지의 규칙을 정의한다. 규칙 정의를 위해 총 3개의 변수 base, advantage, numOfcrash를 사용한다. base는 각 슬레이브에서 실행되는 k를 정하기 위한 값이며 advantage는 질의 위치에 존재하는 슬레이브는 더 많은 k를 찾아야 하는 의미의 값이다. numOfcrash는 질의가 축에 충돌하는 횟수로 질의의 형태는 통계값을 기반으로 원 형태로 정의한다. 해당 값을 활용하여 질의 할당에 계산식에 사용한다.

질의 할당의 기본 규칙은 다음 수식 (1)과 같다. 질의 중심점에 해당하는 노드는 k값으로 advantage 값을 할당 받고 축 충돌이 발생한 경우 충돌 축의 인접 노드는 numOfcrash / base 값을 k에 곱하여 k-최근접 질의를 수행한다.

$$\begin{cases} \text{질의중심노드} : \text{advantage} * k \\ \text{축 충돌 발생} : \frac{\text{numOfcrash}}{\text{base}} * k \end{cases} \quad (1)$$



▶▶ 그림 2. 질의 할당 규칙 예제

그림 2는 질의 할당 규칙에 대한 예를 나타낸다. M1~4는 각 슬레이브의 명이며 x,y축을 기준으로 총 4분면으로 분할되어 있는 상황이다. 노란 원은 하나의 질의이며 축과 충돌되는 위치를 빨간 x선으로 표시하고 있다. (a)의 1번 질의는 M1에만 k-최근접을 수행하면 된다. 2번 질의에서 M1은 advantage 값인 1/2를 가지고 축 충돌이 발생했기 때문에 M1, M2는 각각 2/4 (numOfcrash/base) 값인 1/2를 추가적으로 할당 받는다.

질의 3은 중심점이 축 위에 있기 때문에 advantage 값이 주어지지 않고 M2, M4에 각각 1/2\*k를 할당한다. 4번 질의는 충돌이 1회이기 때문에 M3에는 3/4\*k, M4에는 1/4\*k값이 주어지며 5번 질의는 모든 노드가 1/2\*k를 할당 받는다.

## III. 성능 평가

성능 평가는 Intel core i5 2.7GHz, 8GB 메모리 환경에서 스톱[4]의 가상 분산 환경에서 기존 기법[3]과 실험을 통한 성능을 비교하였다. 데이터는 약 61만개의 실제 데이터 셋에서 질의를 100~500개까지 증가시키며 응답 시간의 차이를 비교하였다. k는 10으로 고정하였으며 base와 advantage 값은 4와 1/2 값을 사용하였다. 그림 3은 질의 수에 따른 질의 처리 응답 시간을 보여준다. 질의 수가 증가할수록 질의 처리 성능도 향상됨을 알 수 있다.



▶▶ 그림 3. 질의 수에 따른 응답 시간

## IV. 결론

본 논문에서는 그리드 색인 기반의 분산 근사 k-최근접 질의 처리에서 규칙 기반의 질의 할당 기법을 제안하였다. 제안하는 기법은 질의의 통계 값을 기반으로 질의의 형태를 정의하여 축의 충돌에 따라 다른 k 값을 할당 하였다. 성능 평가를 통해 제안하는 기법이 실제 환경에서도 적용 가능함을 입증하였다. 향후 기존 기법과의 정확성 비교를 통해 제안하는 기법의 우수성을 검증할 예정이다.

## ■ 참고 문헌 ■

- [1] <https://ko.foursquare.com>
- [2] Z. Yu, Y. Liu, X. Yu, and K. Q. Pu, "Scalable Distributed Processing of K Nearest Neighbor Queries Over Moving Objects," J. of Transactions on Knowledge and Data Engineering, Vol.27, No.5, pp.1383-1396, 2015.
- [3] DoJin Choi, JongTae Lim, KyoungSoo Bok, JaeSoo Yoo. "Distributed Approximate k-Nearest Neighbor Query Processing in Storm Environments.", 한국콘텐츠학회 ICCS 논문집, pp. 33-34, 2017.
- [4] <http://storm.apache.org>