

이동 컴퓨팅 환경에서 위치종속 데이터를 위한 영역 구성

(Organizing Data Regions for Location Dependent Data in Mobile Computing Environments)

유 제 혁 [†] 황 종 선 ^{**}
(JeHyok Ryu) (Chong-Sun Hwang)

요 약 이동 컴퓨팅 환경에서는 이동 클라이언트의 위치에 따라서 질의에 대한 결과 값이 달라지는 상황이 발생하는데 그러한 결과의 데이터를 위치종속 데이터(LDD: Location Dependent Data)라 한다. 따라서 그러한 질의는 지리적인 거리를 고려하여 처리되어야하며 또한 데이터의 관계성 등을 함께 고려하여 처리 효율을 높일 수 있다. 그러나 위치종속 질의에 나타나는 거리를 평가하고 영역별 위치종속 데이터를 구성할 때의 모호성으로 인해 정확히 평가하고 표현하기 어려운 점이 존재한다. 본 논문은 질의에 거리 정보를 사용함에 있어서 발생할 수 있는 문제점을 고려하고 위치종속 데이터간의 관련성 및 질의되는 이동 클라이언트의 위치 그리고 요구되는 지형물에 대한 거리간의 관계를 정량화하여 위치종속 데이터를 위한 데이터 영역인 LDD 영역을 제안한다. 모의 실험에서는 이동 클라이언트가 요청하는 위치종속 질의 및 위치의 성향들 그리고 영역의 두 가지 조밀도 설정하여 제안하는 LDD 영역에서 질의 처리가 지리적인 지역만을 고려한 데이터 영역에서보다 데이터베이스 접근 수를 줄일 수 있음을 보인다.

키워드 : 위치종속 데이터 관리(Location Dependent Data Management), 이동 컴퓨팅(Mobile Computing)

Abstract In mobile computing environments, queries based on the location of mobile clients (MCs) may cause different results. We say that the data of these results are location dependent data (LDD). Location-dependent queries to LDD need to be processed in conjunction with the geographical distance. The efficiency of query processing may also be increased by LDD relationship, etc. But there is the problem of fuzziness about how the distance used in location-dependent queries is evaluated and the data regions are organized. In this paper, we quantify the fuzziness of a location-dependent query on LDD. And we propose data regions for LDD, called LDD regions, by relationship of accessed data and the degree of distance between data objects and MCs' locations. In simulation studies we show that the number of database access for location-dependent queries, which have several settings on MCs' favor and two granularity of regions, can be smaller in proposed LDD regions than that in geographical regions.

Key words : Location Dependent Data Management, Mobile Computing

1. 서 론

최근 휴대용 단말 및 이동 통신 네트워크와 같은 하

드웨어의 발달과 기반 인프라의 구축으로 사용자들이 언제, 어디서나 자신이 필요로 하는 정보를 접근할 수 있게 되었고 현재 이동 컴퓨팅(mobile computing) 환경은 과거의 예상[1, 2, 3]을 대부분 실현하고 있을 뿐만 아니라 앞으로의 수요와 기술 발전의 급속한 성장을 예상하고 있다[4]. 가까운 미래를 보더라도 휴대용 장치들의 자원이 지속적으로 향상되고 차세대 통신 시스템의 출현으로 이동 컴퓨팅 환경에서 사용자들이 수행하는 작업과 요구하는 서비스의 영역이 더욱 다양해져 가고 있다. 이러한 이동 컴퓨팅 환경은 이동 단말의 사용

· 이 논문은 2001년도 한국학술진흥재단의 지원에 의하여 연구되었음 (KRF-2001-041-E00239).

[†] 비 회 원 : 고려대학교 컴퓨터학과
jhryu@disys.korea.ac.kr

^{**} 종신회원 : 고려대학교 컴퓨터학과
hwang@disys.korea.ac.kr

논문접수 : 2002년 7월 29일
심사완료 : 2002년 12월 18일

자들이 자유롭게 이동하면서 도처에 있는 정보를 얻고, 처리하고자 하는 자원을 원활히 접근하는 기능을 제공할 수 있어야 한다.

이동 컴퓨팅 환경에서 적극적인 비즈니스 및 작업처리를 위한 이동 클라이언트의 이동성(mobility)에 대한 요구는 지속적으로 증가하고 있다. 또한 유선 네트워크에 접속한 상태에서처럼 데이터베이스 시스템을 접근하고자 하는 요구가 점차 증가하고 있다. 하지만 그에 적합한 환경 및 서비스를 지원하고자 할 때, 이동 컴퓨팅 환경에 존재하는 본질적인 특징들로 인해 유선 네트워크와는 달리 고려해야 할 사항들이 존재한다. 그러한 사항 중에서 이동 클라이언트가 이동한 위치에 따라 요구하는 질의에 대한 결과가 다를 수 있다는 점이 있다. 이렇게 위치에 따라 질의에 대한 결과 값이 달라질 수 있는 데이터를 위치종속 데이터(LDD: Location Dependent Data)라 한다[5, 6, 7].

위치종속 데이터에 대하여 예를 들어 설명하기 위하여 지리적으로 떨어져 있는 두 개의 국립공원 P , Q 와 각 국립공원에서 콘도미니엄을 운영하는 기업 A 가 있다고 가정하자. P 와 Q 의 위치는 불변이라는 사실에 기반하여, 이동하는 클라이언트가 머물고자 하는 “가까운” 콘도미니엄을 찾고자 할 때, A 기업이 운영하는 콘도미니엄이라도 이동 클라이언트가 위치한 곳에 따라서 질의에 대한 결과 값이 A 기업의 국립공원 P 에 위치한 콘도미니엄이 될 수도 있고 국립공원 Q 에 있는 콘도미니엄이 될 수도 있다. 따라서, 이동 클라이언트가 이렇게 요구하는 위치종속 데이터를 적합하게 전달하기 위해서는 데이터가 요구하는 위치에 종속되어 마련되어 있을 필요가 있으며 또한 그들과 관련된 데이터를 함께 관리함으로써 추후의 관련된 데이터 요청에 효과적으로 대처할 수 있다. 이렇게 지역에 따라 어떻게 위치종속 데이터를 해당 영역으로 구성하느냐의 문제 해결은 보다 신속하고 효과적인 데이터 전달을 유도하여 협소한 무선 접속의 비용을 줄이고 가용성을 높이는 데 중요하게 작용할 것이다.

위치종속 데이터들의 영역은 그림 1에서처럼 논리적인 영역으로 분할될 수 있다. 그림에서 각 지형물의 첫 번째 침자는 지형물이 실제 존재하는 영역을 가리키고 두 번째 침자는 그 침자의 데이터영역에 포함되어 있는 데이터임을 나타낸다. 침자가 AB 인 경우 A 셀의 데이터를 B 영역이 가지고 있음을 나타낸다. 따라서, 기존의 무선망에서 하나의 BS(Base Station)가 하나의 셀(cell)을 서비스한다고 할 때, 논리적인 영역은 하나 이상의 셀에 있는 데이터를 관리할 수도 있다. 이 경우 실제로 두 개

셀의 지역을 갖는 것이 아니라 해당하는 데이터만을 관리함을 의미한다. 위 예에서 뿐만 아니라 응급환자 이송의 경우와 같은 긴급상황에서 이러한 위치종속 데이터는 매우 긴박할 수 있다[8]. 즉, “가까운” 병원을 찾고자 할 경우, “가까운”이라는 질의에 대한 신속한 처리를 제공할 수 있어야 할 것이다. 그림 1에서 A 영역과 B 영역에는 실제 병원이 존재하지 않으며 각 영역에서 가까운 병원은 C 영역의 병원임을 A 영역 B 영역에서도 제공할 수 있게 데이터 영역을 구성함으로써 신속히 대응할 수 있게 할 수 있다. 그러나 여기서 “가까운”이라는 말에는 정확한 값으로 표현하기 어려운 불명확함(uncertainty)의 문제가 있다.

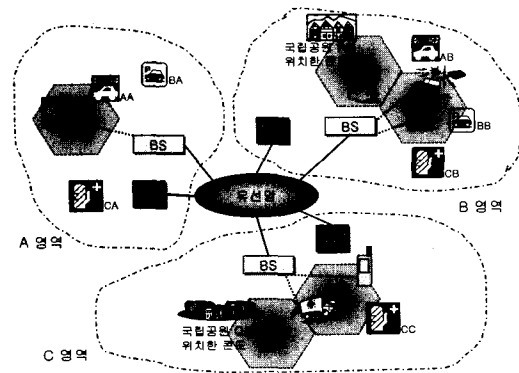


그림 1 위치종속 데이터를 위한 데이터 영역의 예

따라서, 위치종속 데이터에 대한 질의에 대하여 거리 정도의 모호성을 정량화하고, 이동 컴퓨팅 환경에서 위치종속 데이터를 위한 영역을 구성할 필요가 있다. 본 논문에서는 위치종속 데이터들의 관련도와 거리 정도에 의한 퍼지 그래프를 통하여 위치종속 데이터 영역을 구성하여 위치종속 데이터에 대한 질의의 모호성을 줄이고 신속히 답변할 수 있는 기법을 제시할 것이다. 이와 같은 이동 컴퓨팅 환경에서 위치종속 데이터의 관리에 여러 고려사항들을 낳는 분야이며 [5, 9], 최근에 들어서는 그러한 문제들에 대한 연구가 점차 활발해지고 있다[5, 7, 10].

본 논문의 구성은 다음과 같다. 2장에서는 관련연구를 살펴보고 3장에서 위치종속 데이터에 대한 질의의 모호성을 정량화하고 데이터 영역구성이 위치종속 데이터에 대한 질의 처리에 있어서 필요함을 살펴본다. 4장에서는 위치종속 데이터에 대한 관련도 및 거리에 따른 데이터 구성과 영역 구성 방법을 제시한다. 5장에서는 본 논문

에서 구성하는 위치중속 데이터에 대한 질의 처리가 데이터베이스 접근 수를 줄임으로써 효과적인 질의를 처리할 수 있음을 모의실험을 통하여 보인다. 마지막으로 6장에서 결론을 기술한다.

2. 관련연구

이동 컴퓨팅 환경에서 데이터 관리에 대한 기존의 연구들은 잦은 접속 단절, 빈약한 무선 대역폭, 이동 단말의 자원 부족, 배터리 전원의 한계와 같은 제약사항들을 극복하며 작업환경을 개선하려는 여러 가지 접근이 있었다. 그런데 대부분의 기존 연구들은 한 지역에서의 데이터 접근에 관련된 환경을 가정하고 있으므로 이동 클라이언트의 위치에 따라 요구하는 데이터의 결과가 달라짐을 고려하지는 않았다[11, 12, 13]. 그러나 최근에는 이동성과 관련한 위치중속 데이터에 대한 필요성이 점차 증가하고 있다.

[1]에서는 사용자의 위치에 대한 불명확성을 제한된 범위에서 가정하며, 사용자의 이동성 패턴에 대한 빈도에 따라서 영역 분할을 정의하여 전체적인 비용을 줄이고자 하였다. 그리고 의사 SQL로 작성된 질의에 대한 해석 전략을 기술하였다.

[14]에서는 동적 속성(dynamic attributes)이라는 것을 추가한 MOST 라는 새로운 모델을 제시한다. 시간적이거나 지속적인 위치의 질의를 고려하였으며 이동하는 대상의 위치를 시간 함수로 표현하였다. 지속적으로 갱신되는 동적 속성에 대하여 미래의 값을 위한 질의를 가능하게 하였으며 또한 이동하는 대상의 위치에 대한 인덱싱을 연구하였다.

[5]에서는 위치중속 데이터를 기존의 데이터와 구별해서 관리하는 접근을 시도하였다. 이 연구는 시간적 복사본(temporal replica)과 공간적 복사본(spatial replica)을 정의하고, 이들이 의미가 서로 다르다는 점에 착안하여 공간적 복사본의 일관성, 질의, 트랜잭션 등을 재고하였다. 물리적 지형의 위치가 변하지 않는다는 진리에 따라 위치중속 데이터가 하나의 옴은 값을 갖는 논리적 영역을 정의하였다. 질의에 대해서는 데이터와 질의 두 가지 모두를 수정하는 방안을 제시하였으며 분할된 트랜잭션으로 새로운 트랜잭션의 성질을 제시하였다. 그리고 위치중속 질의 처리를 위해 데이터베이스를 접근하게 되는 과정을 설명하는 LDD 구현 구조를 제시하였다. [22]에서는 위치중속 데이터를 효과적으로 접근하기 위하여 위치 속성이 식별되어야함을 지적하고 질의에서 위치와의 관련을 정형화하였다. 그리고 그러한 위치의 판별이 필요함을 설명하여 위치 인식(location

awareness)과 위치 중속의 개념을 차별화 하였다.

[10]에서는 에너지 및 접속 요소를 고려하여 질의 최적화를 위한 비용 모델로 제시하였다. 이 모델에서는 위치 정보에 대한 불확실성을 제한하기 위하여 위치 서버의 분할을 사용한다. 즉, 이동 클라이언트가 하나의 셀을 넘어 서지만 분할된 영역에 있으면 HLR(Home Location Registers)을 갱신하지 않게 하였다. 여기서는 이동 클라이언트의 위치를 파악하기 위한 질의에 초점을 두었으나 본 논문에서는 질의의 주체인 이동 클라이언트의 위치보다는 질의되는 위치중속 데이터에 초점을 둔다.

[7]은 위치 데이터에 대한 계층적 개념을 도입하였다. 이 계층들은 위치의 조밀도(granularity)에 대한 맵핑을 정의한다. 예를 들어, 미국의 한 주(state)에 속하는 도시에 대한 정보를 생성한다. 또한 특정 위치에서 위치중속 속성의 값을 갖는 관계나 데이터베이스 검색을 지원하는 분산 디렉토리로써 사용된다. 그리고 위치중속 속성들에 대한 공간적 인덱스를 포함하는 계층 개념으로 확장하며 위치중속 질의에 대한 지원을 고려하였다.

[15]는 클라이언트가 여러 셀의 무선 환경에서 셀간 이동시에 요구되는 위치중속 질의의 스케줄링을 위한 세 가지 기법을 제시하였다. 우선, 하나의 사이트에 대하여 주변 해당 지역에 있는 어느 지점들은 다른 어떤 사이트보다 더 가까운 영역을 갖는 사이트가 되는 성질이 있는 보르노이 다이어그램(Voronoi Diagram)을 사용하여 영역을 구분하였다. 그리고 그러한 영역들간에 이동시 질의에 대한 지원을 위하여 우선순위(priority), 지능적(intelligent), 혼합(hybrid) 기법을 제안하였다.

또한 유사한 연구 영역으로 기존의 분산데이터베이스에서는 데이터베이스 또는 테이블 등을 대상으로 접근 비용과 효율성을 위해 수평, 수직, 혼합 분할 방법 등의 단편화(fragmentation)방법이 있었다[16, 17, 18, 19]. 이 연구영역의 수직분할 연구에서 트랜잭션이 처리될 때 함께 사용된 속성들에 대하여 친밀도(Affinity)라 표현하고 그 정도에 따라 분할하는데 반영하였으며 [17]의 경우에는 그래프를 이용하였고 [18]에서는 분산데이터베이스 분할에서 일반적으로 분할되는 경계가 명확하지 않은 점을 해결하기 위하여 퍼지개념을 도입하였다. 그러나 이러한 분산 데이터베이스에서의 분할 방법들을 이동 컴퓨팅 환경에 적용하기 위해서는 이동 컴퓨팅 환경에 고유한 특성이 고려되어야 한다.

이와 같은 기존 연구들의 동향에 대하여 질의되는 대상이 움직이는 객체를 관리하는 경우인 [14]와 달리 본 논문에서는 [5]에서처럼 질의되는 대상은 움직이지 않은 고정된 지형물에 대한 데이터를 고려할 것이며 기존

의 분산데이터베이스의 분할 방법에 거리의 정도와 위치중속 질의의 특성을 분석함으로써 위치중속 데이터의 구성 및 질의 처리를 위한 관련도를 유도할 것이다.

3. 위치중속 질의의 평가

이동 컴퓨팅 환경에서 위치중속 데이터를 접근하는 질의를 처리하기 위한 지원은 다음과 같은 세 가지의 형태로 분류될 수 있다[5]. 첫 번째 형태는 질의에 대한 수정 없이 올바른 값의 위치중속 데이터를 접근할 수 있도록 데이터들을 새롭게 구성하여 저장하는 것이다. 두 번째 형태는 거리 정보를 갖는 질의를 지원하는 것으로, 데이터에 추가되는 거리 정보가 요구될 수는 있지만 마치 질의가 위치에 대한 투명성을 갖듯이 데이터의 저장위치와 접근이 처리될 것이다. 따라서 질의만이 수정된 형태이다. 세 번째 형태는 첫 번째와 두 번째 접근을 모두 고려하는 것이다. 즉, 데이터를 어떻게 저장하며 질의를 어떻게 해석하고 실행할 것인가를 함께 고려하는 방식이다. 본 논문은 퍼지(fuzzy) 개념을 이용하여 위치중속 질의에 거리 요소를 지원할 때의 모호성을 정량화하고, 데이터간의 관련정도 및 거리 요소를 동시에 고려하여 데이터들을 구성하는 방법을 유도한다. 따라서, 세 번째 접근방법을 고려하며 본 절에서는 위치중속 질의의 거리정도를 평가하기 위한 방법을 제시한다.

이동 컴퓨팅 환경에서 위치중속 데이터를 요구할 때, 사용하는 일반적인 거리정도는 명확하지 않은 경우가 있다. 예를 들면 “가까운(Near)”과 같은 실생활의 거리 개념은 어떻게 표현하고 평가하느냐에 따라 결과 값의 집합은 달라질 수 있다. 이것은 제안하는 환경에서 뿐만 아니라 모든 현상의 세계에서 기존의 방법으로만은 표현하고 해석할 수 없는 수많은 것들 중에 하나이다. 이러한 불명확한 성질의 현상을 표현할 수 있는 방법으로 퍼지 이론을 고려할 수 있다[20]. 퍼지 이론은 기존의 수학적 집합 및 논리와 관계 표현 등의 한계를 극복할 수 있는 확장된 접근이라 할 수 있다. 예를 들어, R 라는 집합(crisp set)을 X 라는 전체집합에서 정의될 수 있다고 하자. $x \in X$ 일 때, 집합 R 은 다음 함수(characteristic function)와 같이 정의될 수 있다.

$$X_R(x) = \begin{cases} 1 & \text{if } x \in R \\ 0 & \text{if } x \notin R \end{cases} \quad (1)$$

그러나 이러한 함수에 의해 정의되는 집합은 그 소속 여부가 명확한 경우에 해당된다. 퍼지집합의 개념을 이용하면 애매한 원소의 집합 소속여부를 소속정도로 고려할 수 있다. 즉, 전체집합 X 의 원소 x 가 집합 R 에 대한 소속 여부가 애매할 경우 $[0, 1]$ 사이의 소속정도(member-

ship degree)로 표현할 수 있다. $\mu_R : X \rightarrow [0, 1]$ 에 해당한다. 값이 더 클수록 집합의 소속정도가 더 높음을 나타낸다. 이러한 함수를 소속함수(membership function)라 한다.

본 논문에서는 이러한 퍼지 개념을 위치중속 데이터를 서비스하는 영역에서 거리의 애매모호한 정도를 정량화하는데 사용한다. 예를 들어보면, “가까운”이라는 퍼지 집합을 N 이라 할 때, N 을 위한 소속함수 μ_N 으로 할당할 수 있다. 이동 컴퓨팅 환경에서 실제 적용하는 위치의 척도는 셀, 등록영역(registration area), 데이터 영역 등 적용하는 상태에 따라 그 규모가 달라질 수 있다. 여기서는 거리의 일반적인 척도인 미터계(km)에 기반하여 현재 이동 클라이언트가 존재하는 하나의 데이터 영역 내에 있는 범위를 “가까운”으로 가정하고 일반화하기 위해 다음과 같은 Sigmoid 소속함수를 사용한다.

$$\mu_N(x) = \begin{cases} 1 & 0 \leq x \leq a \\ \frac{1}{1 + \exp \frac{1}{b(x-a)}} & a < x \end{cases} \quad (2)$$

위치중속 데이터의 모호성을 평가하기 위하여 이 소속함수의 각 변수들을 다음과 같이 사용할 것이다. a 는 하나의 데이터 영역에 속하는 기지국으로부터의 영향이 미치는 범위까지의 반경이라 하자. 이 때, a 값은 위치중속 데이터를 이용하는 애플리케이션을 서비스하는 환경에 따라 “가까운”이라는 의미를 협소하게 또는 광의로 설정할 수 있는 변수이다. 그리고 기울기에 해당하는 b 의 값은 점차 현재의 위치로부터 멀어지는 다른 영역의 데이터들을 어느 정도로 정량화할 것인지를 결정하는 기울기이다. 그리고 임의의 상수 c 의 값으로 0과 1 사이에 값을 어느 정도 폭의 거리만큼 부여할 것인지를 결정할 수 있다. a 의 값을 2, b 의 값을 0.6, c 의 값을 8로 설정하였을 때, 퍼지 집합 N 의 소속함수 μ_N 의 값을 그래프로 표현하면 다음과 같다.

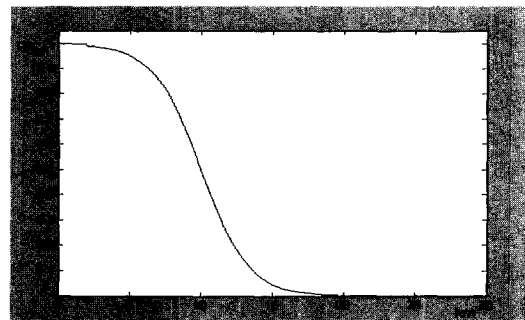


그림 2 퍼지 소속함수 μ_N 에 의한 “가까운”의 예

위 그림과 같이 가정한 소속함수에 의하면 이동 클라이언트로부터 0미터에서 a 미터의 위치에 있는 대상은 의도하고자 했던 “가까운”에 해당하는 것으로 평가할 수 있다. 위 예에서 x 의 값이 1 km 거리에 해당하는 위치중속 데이터라면 원하는 결과일 수 있으며, 3km의 위치중속 데이터라면 약 0.99정도(degree)로 “가까운”에 소속된다. 그리고 10km 이상으로 멀어지는 경우부터 0.5이하로 값이 작아지게 된다. 따라서, 현재의 위치를 관찰하는 영역과 이웃하는 영역의 데이터임을 효과적으로 수치화하고 이동 클라이언트로부터 요구되는 위치중속 데이터에 대한 거리 정도를 정량화하여 모호성을 줄일 수 있다.

위 예에서 실세계의 거리표현 정도를 일반적인 거리의 척도로 추상화(abstraction)한 것은 셀, 등록영역 등이 지역에 따라서 커버하는 범위가 다른 척도이므로 소속함수의 값을 얻을 경우, 상대적인 값으로 인해 일반적이지 못한 특정한 환경에서만 옳은 값이 될 수 있기 때문이다. 따라서 추상화한 후에 원하는 환경에 맞는 위치조밀도에 따라 적용하는 단계가 바람직한 순서이다. 이렇게 얻은 소속 정도는 다시 실제 시스템의 환경에 적용하여 단일 값을 갖도록 할 수 있을 것이다.

질의의 퍼지 표현에 의한 정도(관계의 강도)에 따라 가까운 영역을 정의할 수 있다. 즉, 위 예에서 요구하는 위치중속 데이터가 $\mu_N(x)=0.5$ 정도로 가까운 데이터라면 10 km범위의 데이터를 결과 값으로 얻게 될 것이다. 이에 대하여 적용될 위치간 거리의 단위가 셀 단위로 관리되는 영역이라 하고 셀의 반경이 2km일 경우, 소속함수에 의해 0.5정도로 “가까운”의 의미는 세 개 셀에 걸치는 범위에서 위치중속 데이터를 얻을 수 있게 된다.

이와 같이 위치중속 데이터에 대한 질의에서 거리의 정도를 정량화하여 위치중속 데이터를 적용하는 환경에 맞게 위치간 거리의 단위를 바인딩(binding)함으로써 본질적인 모호성을 명확히 평가하고 요구하는 데이터를 위한 영역을 구성하는데 사용한다. 이를 기반으로한 데이터 영역 구성은 다음절에 기술한다.

4. 위치중속 데이터 구성

데이터 영역이란 애매하고 명확하지 않은 정의라는 점에서 재고할 필요가 있다. 왜냐하면 그러한 영역의 구성은 앞서 언급한 위치중속 데이터에 대한 요구에 어느 정도의 거리가 부여되었느냐에 따라서 여러 가지 다른 결과를 낼 수 있기 때문이다. 그러므로, 본 절에서는 데이터간의 관련정도와 거리정도를 함께 고려하며 데이터 영역을 구성하기 위한 기반을 유도할 것이다. 구성하

는 방법으로는 퍼지 그래프를 사용하여 적합한 분할을 얻음으로써 위치중속 데이터들이 어느 영역에 소속되는지의 불명확성을 감소시키기 위한 방법이다. 이러한 분할을 위해서 다음을 고려한다.

- 위치중속 데이터간의 관련 정도
 - 해당 지역에서 수행되는 각 질의의 빈도 및 같은 질의에서 사용된 위치중속 데이터인가를 반영
- 위 관련 정도에 의한 위치중속 데이터와 질의된 이동 클라이언트간의 거리를 고려
 - 거리에 대한 퍼지 소속정도 적용

위와 같은 기준으로 구성된 영역을 LDD 영역이라 하자. 그리고, 실세계의 물리적인 지형간의 거리는 이미 알려져 있는 사실로써 변함이 없다고 가정한다.

4.1 위치중속 데이터 관련도: LddBond

이동 클라이언트가 요구하는 위치중속 데이터가 서비스되는 전체 영역을 $F(Field)$ 라 하고, F 에 소속되며 거리 상으로 서로 다른 영역들의 집합을 $R=\{r_1, r_2, \dots, r_n\}$ 라 하고 하자. 그리고 F 에서 관리되는 위치중속 데이터의 전체 집합을 $D=\{d_1, d_2, \dots, d_n\}$ 이라 하고 질의의 집합을 $Q=\{q_1, q_2, \dots, q_i\}$ 라 하자. 먼저, 위치중속 데이터들이 소속될 영역 구성을 위한 위치중속 데이터들간의 관련도를 구하기 위해 고려할 요소들은 표 1과 같다.

표 1 관련도에서 고려하는 요소

고려사항	조건	값
$ref(q_i, d_i)$	q_i 가 d_i 를 참조한 경우	1
	그렇지 않은 경우	0
$put(q_i, r_u)$	q_i 가 r_u 에서 발생한 경우	1
	그렇지 않은 경우	0
$sameq_acc_{qt}(d_i, d_j)$	$ref(q_i, d_i)=1$ 이고 $ref(q_i, d_j)=1$ 인 경우	1
	그렇지 않은 경우	0
$justq_acc_{qt}(d_i, d_j)$	$ref(q_i, d_i)=1$ 또는 $ref(q_i, d_j)=1$ 인 경우	1
	그렇지 않은 경우	0

위 데이터 관계도를 위한 고려사항과 함께 이동 컴퓨팅 환경에서 이동 클라이언트의 질의 및 거리 관계에 해당하는 사항을 다음과 같이 정의한다.

- 정의 1. $freq(q_i, r_u) : q_i$ 가 r_u 에서 발생하는 빈도
 - 정의 2. $loc_{q_i} = \{p_i \mid put(q_i, r_u)=1 \text{인 경우 } q_i \text{의 위치}\}$
 이며, F 에 포함되는 위치
 - 정의 3. $pos_{d_i} = \{p_i \mid ref(q_i, d_i)=1 \text{인 경우 } d_i \text{의 위치}\}$
 이며, F 에 포함되는 위치
 - 정의 4. $dist(i, j) = \mu_N(|i - j|)$
- 정의 1은 질의가 발생하는 영역의 빈도이며 정의 2는

질의가 발생한 지역 R 의 한 원소에 포함되는 한 지점이 며 정의 3은 질의가 참조한 데이터의 물리적인 위치에 해당한다. 정의 4에서 $|i-j|$ 는 두 요소 i, j 간의 거리를 의미하며 3절의 수식 (2)의 퍼지 소속함수에 의한 소속 정도이다. 표 1에서의 고려사항과 정의 1, 2, 3, 4를 이용하여 F 에서 서비스되는 임의의 두 가지 위치중속 데이터 d_i 와 d_j 간의 관련도 $LddBond(d_i, d_j)$ 를 구하는 식은 다음과 같다.

$$LddBond(d_i, d_j) = \begin{cases} \frac{\sum_{r_u | sum_{acc}(d_i, d_j)} \sum_{r_u | pos(q_i, r_u)} freq(q_i, r_u) * DQD}{\sum_{r_u | sum_{acc}(d_i, d_j)} \sum_{r_u | pos(q_i, r_u)} freq(q_i, r_u)} & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases} \quad (3)$$

, where $DQD = \min[\text{dist}(pos_{d_i}, pos_{d_j}), \text{MAX}[\text{dist}(loc_{q_i}, pos_{d_i}), \text{dist}(loc_{q_j}, pos_{d_j})]]$

수식 (3)에서 질의된 영역에서 접근된 데이터에 대한 빈도를 고려하여 데이터간의 관련도를 평가한다. 여기에 위치중속 데이터의 특성상 접근된 데이터 d_i 와 d_j 그리고 질의되는 위치의 거리를 반영하는데 이 세 요소간의 거리의 정도를 반영한 수식이 DQD (the Degree of distance between Query and Data)이다. DQD 는 접근된 두 데이터의 거리가 멀 경우 해당하는 영역 하나에 구성될 가능성을 작게 하고 서로 가까운 거리일 경우는 같은 영역으로 유지될 수 있는 가능성을 높도록 조정된다. 이때, 질의한 위치의 성향을 반영하기 위하여 질의된 지형물과 질의한 위치간의 거리를 고려하여 이들간의 거리가 멀고 가까움에 따라서도 $LddBond(d_i, d_j)$ 의 값을 크고 작게 하는데 영향을 준다. 이것은 위치중속 데이터의 특성을 반영하여 데이터 영역을 구성하기 위한 기반을 $LddBond(d_i, d_j)$ 가 제공하도록 하기 위함이다. DQD 에서 고려하는 바와 같이 데이터간 거리와 질

의한 이동 클라이언트의 위치 사이의 구도를 고려해야 할 하나의 예를 살펴보면 그림 3과 같다.

그림 3에서 보면 수식 (3)에서 DQD 의 세 요소를 거리 a_b 와 a_{MC} , MC_b 로 표현하였다. 여기서 MC 는 질의를 제출한 이동 클라이언트의 위치가 된다. 예를 들어 MC 의 반경 m 거리 정도의 범위에 있는 호텔을 원했을 경우, 그림에서처럼 호텔이 ③와 ④에 있고 일정한 지리적인 범위의 셀을 서비스하는 BS 를 가정할 때, MC 는 일차적으로 ③를 받게 되지만 사실은 ④가 더 가까운 호텔이다. 물론 이 경우에 이웃하는 BS 지역의 데이터를 관리하는 데이터베이스를 다시 접근하여 그 정보를 가져올 수 있다. 하지만 본 논문에서는 이러한 경우가 많은 경우, $LddBond$ 에 의하여 하나의 데이터 영역으로 될 수 있는 가능성을 높임으로써 실제로 가까운 거리의 데이터를 얻을 수 있도록 하여 데이터베이스의 접근 수와 통신비용을 감소시킨다. 위 그림의 경우 DQD 에 의해서 이러한 삼각구도의 거리에 대한 퍼지 소속함수 값이 반영되어 두 데이터간 거리에 의한 퍼지 소속함수 값을 관련도에 반영한다. 그리고 정의 4에 의해서 Sigmoid함수의 소속정도를 따르면서 거리 정도를 반영하기 때문에 이와 같이 바로 이웃하는 경계간의 소속정도는 1보다는 작지만 매우 높게 나타난다. 그러므로 이러한 경우에 두 데이터가 하나의 영역에 구성될 가능성을 높일 수 있다. 따라서, $LddBond$ 에서 거리의 정도를 반영한 이유는 데이터간의 빈도뿐만 아니라 접근된 데이터와 질의된 곳의 위치를 고려하여 거리가 먼 경우, 최종의 $LddBond$ 값을 좀 더 작게 하고, 거리가 가까운 경우에는 두 데이터간의 최종 $LddBond$ 를 증가시킴으로써 위치중속 데이터의 특성을 반영한다. 이와 같이 $LddBond$ 는 함께 접근되는 위치중속 데이터의 질의 빈도와 거리정도의 상호관계에 의하여 최종 데이터 영역을 구성하기 위한 관련도를 제공한다. 다음절에서는 이 관련도에 기반하여 영역을 구성하는 방법을 제시한다.

4.2 위치중속 데이터 영역 구성

위치중속 데이터의 집합 D 의 각 원소에 대하여 $LddBond(d_i, d_j)$ 를 구하면 대칭형의 관련도 행렬을 얻을 수 있다. 이를 기반으로 데이터 영역을 구성하는 기법은 기존의 분산 데이터베이스와 퍼지 그래프에서 분할하는 기법을 응용할 것이다[16, 17, 18, 19, 20]. 이 퍼지 그래프에서 노드 집합은 위치중속 데이터의 집합이며, 간선의 집합은 앞서 고려한 $LddBond(d_i, d_j)$ 에 의한 관련도에 해당한다. 본 논문은 먼저 퍼지 그래프의 α -cut을 이용하여 분할을 할 때, 관련성이 높고 근

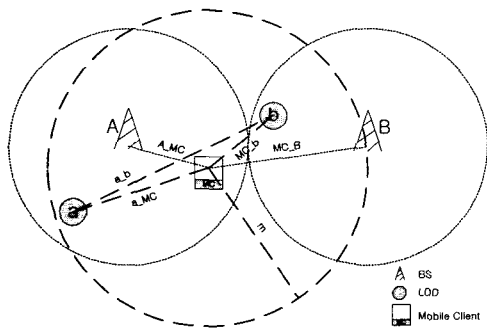


그림 3 이동 클라이언트 및 위치중속 데이터의 위치 관계

거리의 위치중속 데이터들을 동일한 LDD 영역으로 만들기 위한 최적의 α 값을 구한다. 퍼지 그래프에서 α -cut은 α 값보다 소속정도가 작거나 같은 모든 간선을 제거하여 생성되는 그래프 분할을 적용한다.¹⁾ 여기서, 위치중속 데이터가 포함되는 영역을 결정하기에 적합한 α 값을 선정하기 위하여 다음과 같은 기준을 고려한다.

- 분할된 각 서브그래프를 연결하는 노드간에 관련도(결합도)를 최소화한다. 이것은 분할된 영역간 데이터들간의 소속정도가 낮아야함을 의미한다.
- 분할된 각 서브그래프내의 노드간 관련도(응집도)를 최대화한다. 이것은 분할된 한 영역에 속하는 데이터들 간에는 높은 소속정도를 가져야함을 의미한다.

위와 같은 기준은 앞서 언급한 위치중속 질의 처리의 비용을 최소화하기 위함이다. 그러한 α 값을 구하기 위해서는 각 α 값으로 분할을 하였을 때, 위 두 가지 사항에 가장 적합한 최적의 α 값을 선택하여야 한다. 먼저, 두 가지 기준은 각각 상반되는 관계를 가지고 있으므로 올바른 평가를 위해서 한 가지 기준으로 비용을 통일하여 계산이 이루어져야 한다. 분할된 영역간의 응집도를 기준으로 하려면 결합도에 해당하는 성분을 “1-(분할에 사용된 α 값에 의해 제거된 간선의 값)”으로 계산해야 전체적인 퍼지 그래프의 분할 비용을 통일하여 산출할 수 있다. 이때, 관계정도의 차이를 계산하기 위하여 본 논문에서는 유클리드 거리(Euclidean distance)를 사용한다.

먼저, 분할된 하나의 서브그래프 A에 대하여 m 개의 간선값이 있을 때, 순서에 상관없이 두 개를 선택하여 만든 쌍 n 개에 대하여 유클리드 거리(이 계산 값을 $Con(\alpha)$ 라 하자)를 다음 식으로 계산한다.

$$e(x_i, y_j) = \sqrt{\sum_{x_i, y_j \in A} (x_i - y_j)^2}, \quad \text{단, } 1 \leq i \neq j \leq m \quad (5)$$

그리고, 분할에 의해 생성된 모든 서브 그래프에 대하여 유클리드 거리를 구하여 합산한다. 다음, 그래프 분할에 사용한 α 값에 의해 제거된 간선들의 결합도에 해당하는 응집도는 1-(제거된 각 간선들의 값)이고, 이 값들의 개수 m' 개의 원소에 대하여 순서에 상관없이 두 개를 선택하여 만든 쌍들의 유클리드 거리를 구한다(이 계산 값을 $Discon(\alpha)$ 라 하자). 그러면 임의의 α 값에 의하여 분할된 퍼지 그래프의 개수가 g 개일 때, 최종적으로 분할을 평가하는 평가함수 $Eval(\alpha)$ 는 다음과 같다.

$$Eval(\alpha) = \sum_{i=1}^g Con_i(\alpha) + Discon(\alpha) \quad (6)$$

$Eval(\alpha)$ 의 결과 중 최소값일 때 α 는 전체 퍼지 그래프에 대하여 최선의 분할을 유도하며 이를 데이터 영역을 구성하는데 적용한다. 만일 질의된 데이터가 이동 클라이언트의 현재 지역으로부터 멀리 떨어져 있다면 그 질의들은 원하는 데이터를 얻기 위해 라우트되어야 하는데 이러한 상황이 빈번할 경우, 각 영역에 존재하는 데이터베이스 접근 수를 증가시킨다. 또한, 질의가 발생한 지역에서 바로 이어지는 이동 클라이언트의 다른 관련 질의들이 발생할 경우 유사한 데이터베이스 접근의 비용을 초래할 수 있다. 따라서, 앞서 설명한 $LddBond(d_i, d_j)$ 는 데이터들과 이동 클라이언트의 질의가 발생한 위치간의 거리 정도뿐만 아니라 접근되는 지역성을 고려하기 때문에 질의가 다른 영역으로 라우트되어 데이터베이스를 접근하는 비용을 효과적으로 줄일 수 있다. 제안하는 새로운 데이터 영역인 LDD 영역을 구성하는 단계를 살펴보면 표 2와 같다.

위 단계를 완료하면 지리적인 데이터 영역들을 제안하고자 하는 LDD 영역들로 재구성할 수 있다. 이와 같은 절차로 구성된 LDD 영역은 이동 클라이언트가 동일한 위치에서 접근하는 데이터들간의 관련도 및 어느 정도 근거리의 데이터들인가에 대하여 동일한 데이터 영역에 포함될 수 있는 가능성을 부여한다. 따라서, 이동 클라이언트가 이어서 요구하는 관련 질의에 대한 위치중속 데이터의 접근을 신속히 처리할 수 있는 기반을 제공할 수 있다. 또한, 위치중속 데이터가 관리되는 LDD 영역을 구성하는데 있어서, 거리의 척도에 따라 얼마나 크고 작게 분할될 것인가는 적용하는 이동 컴퓨팅 환경의 설정에 따라 다를 수 있다. 그러므로, 본 논문에서는 우선적으로 위치중속 데이터에 대한 질의에 있어서 거리의 모호성을 정량화하는 절차를 통하여 적용되는 환경의 척도에 따라 거리의 정도를 표현할 수 있게 함으로써 거리 척도의 상이한 조밀도에 유연성을 제공한다. 여기서 자주 변경되지 않는 지형물을 대상으로 LDD 영역을 구성하므로 지형물 변화에 대한 영역의 갱신은 고려하지 않기로 한다.

5. 모의 실험

제안한 LDD 영역을 평가하기 위하여 위치중속 데이터에 대한 질의 처리에 대하여 모의실험을 한다. 실험되는 위치중속 데이터 및 이동 클라이언트의 위치 설정은 [21]에서 사용한 확률분포를 사용하였다.

1) 여기서 α 값은 포함하지 않고 제거하는 strong α -cut과 구별한다[20].

표 2 LDD 영역 구성 절차

단계 1: LDD간의 관계도 계산	<ul style="list-style-type: none"> • $LddBond(d_i, d_j)$ - 질의되는 데이터와 이동 클라이언트가 질의한 위치간의 거리 및 접근 빈도에 기반하여 d_i와 d_j간의 관련도를 계산한다.
단계 2: 전체 퍼지그래프를 분할하기 위한 평가함수 계산	<ul style="list-style-type: none"> • $\sum_{i=1}^n Con_i(a)$ - 적용한 α 값에 의해 생성된 서브그래프 집합 $G = \{A1, A2, \dots, Ag\}$의 각 원소마다 존재하는 간선 값에 대하여 순서에 상관없이 두 개를 선택하여 만든 쌍의 유클리드 거리로써 차이정도를 계산하며, 이들을 모두 합산하여 응집도를 계산한다. <hr/> <ul style="list-style-type: none"> • $Discon(a)$ - 적용한 α 값에 의해 제거된 간선들의 결합도를 응집도로 변환하여 표현하면 $1-(\text{각 간선들의 값})$ 이고, 이 값들의 각각에 대하여 순서에 상관없이 두 개를 선택하여 만든 쌍들의 유클리드 거리로써 차이정도를 계산하여 더한다. <hr/> <ul style="list-style-type: none"> • $Eva(a) = \sum_{i=1}^n Con_i(a) + Discon(a)$ - 퍼지 그래프를 분할하기 위해 가장 작은 $Eva(a)$값의 α 값을 선택한다.
단계 3: LDD 영역 구성	<ul style="list-style-type: none"> • 지리적 데이터 영역에 있는 각 데이터베이스의 데이터 항목을 재구성한다. 여기서, 데이터 항목은 단계 2에서 최종으로 분할된 서브그래프의 노드들에 해당한다. - 만일 d_i-d_j가 서브그래프 중에 하나이고 기존의 어떤 지리적 데이터 영역이 d_j는 포함하지 않고 d_i만을 포함하고 있다면 현재 지리적 데이터 영역으로부터 가장 가까운 데이터 d_j를 복사한다. 나머지 서브그래프들에 대해서도 같은 방식으로 적용한다.

5.1 모델 설정

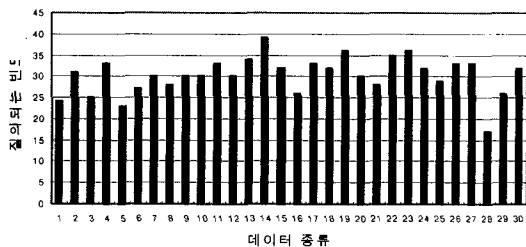
본 실험 모델에서는 제안하는 데이터 영역인 LDD 영역의 평가를 위해 [5]에서 서술한 현재 위치의 위치중속 데이터들을 관리하는 데이터베이스에 원하는 데이터가 없을 경우 요구하는 데이터가 있는 데이터베이스로 질의를 라우트하는 방식과 비교할 것이며, 이를 지리적인 데이터 영역이라 부르기로 한다. 주어진 서비스 지역의 지리적인 데이터 영역은 자신의 영역에 존재하는 위치중속 데이터들에 해당하는 데이터 집합을 갖는다. 실험 모델에서는 144개의 지리적인 데이터 영역을 서비스 지역으로 설정하였으며 각 지리적인 데이터 영역이 갖는 데이터 집합을 관리하는 데이터베이스를 갖는다. 그리고 본 실험에서는 제안하는 LDD 영역을 구성함으로써 얻을 수 있는 성능향상을 비교하기 위하여 적용되는 서비스 지역과 영역의 수를 지리적인 데이터 영역의 수와 동일하게 적용하여 실험한다. 서비스되는 데이터에 해당하는 물리적인 위치와 데이터의 종류(예, 호텔, 공항 등)의 설정은 일양(Uniform)분포로 설정하였다. 그리고 질의를 요청하는 이동 클라이언트가 전체 영역들에 걸쳐서 거의 일정하게 분포하는 경우와 소수의 영역에 매우 많은 이동 클라이언트가 분포하는 설정으로 나누어 실험을 한다. 또한 협의의 '가까운'과 광의의 '가까운'을 영역 크기에

설정하여 실험하기 위하여 거리는 km로 가정하고 하나의 영역의 크기를 2km와 4km로 각각 설정하였다. 환경 설정에 대한 요약은 표 3과 같다.

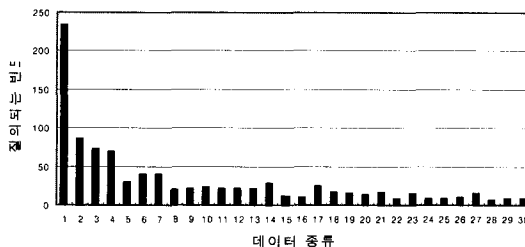
표 3 매개변수 설정

파라미터	의미	설정
NoGeoR	지리적인 데이터 영역의 수	144개
NoLDDR	LDD 영역의 수	144개
ServDT	서비스되는 데이터의 수	400개
KndOfDT	서비스되는 데이터의 종류	30
NoMC	위치중속 질의의 수	60~ 600개
QPttrn	KndOfDT에 대한 이동 클라이언트의 질의 성향	다양성: 17~39, 소수 집중성: 6~234
LocMC	질의하는 이동 클라이언트의 위치 분포	일양분포, 지프(Zipf)분포
RgSz	영역 하나의 크기	2km, 4km

각 이동 클라이언트는 하나의 질의를 하며 그 질의에 의해 접근하는 데이터 집합을 갖는다. 각 실험에서 LDD 영역 구성을 위해 사용되는 질의들은 질의하는 이



(a) 전체 LDD 종류에 골고루 질의되는 성향



(b) 소수의 LDD 종류에 집중되는 질의 성향

그림 4 이동 클라이언트의 LDD 질의 성향 설정

동 클라이언트가 접근하는 데이터에 대한 설정과 같은 일양분포와 지프분포 두 가지를 사용하였다. 즉, 극소수의 데이터 종류에 질의가 집중되는 이동 클라이언트의 성향과 모든 데이터 종류에 대하여 골고루 요구하는 이동 클라이언트의 성향으로 나누어 LDD 영역 구성을 하였으며 평가를 위한 질의도 각각 같은 분포를 따르게 하였다. 그림 4는 실험에서 사용한 각 데이터 종류 당 질의되는 빈도로써 질의 성향을 나타낸 예이다.

실험에서 평가하기 위한 LDD 영역을 구성하기 위해 얻은 값들은 표 4와 같다. 표 4는 4장에서 서술한 기법에 의하여 계산된 값들로 최종 LDD 영역 구성을 위해 유도된 데이터 쌍을 이용하여 본 실험에서 설정한 각 환경에서의 LDD 영역을 구성한다.

두 가지 이동 클라이언트의 질의 성향을 먼저, 각각 이동 클라이언트가 많이 집중되는 지역 설정 및 전체적인 지역에 널리 분포하는 영역을 설정하여 실험을 한다. 두 번째로는 그러한 이동 클라이언트의 특성 설정과 함께 '가까운'의 의미를 광의 및 협의로 설정하여 실험한다.

표 4 LDD 영역구성을 위해 유도된 정보

RgSz	QPttm	선정된 α	Eval(α)	LDD 영역 구성을 위해 유도된 데이터 쌍
광의의 가까운	다양성	0.28	0.799750	3-28, 12-24, 15-18,
	소수 집중성	1.00	0.558838	-
협의의 가까운	다양성	0.49	5.014381	1-4-9-18-25, 6-15-13-23, 20-24
	소수 집중성	0.42	3.658763	1-4-3-23, 6-15-12-9-10-19, 13-28, 18-21

5.2 실험 및 결과

먼저, 그림 4와 같은 두 가지 성향의 질의를 지리적인

데이터 영역과 제한한 LDD 영역에서의 실험을 비교 평가한다. 하나의 영역 크기를 4km로 설정하였으며 실험 모델에서 만일 질의된 데이터가 이동 클라이언트의 현재 영역에 존재하지 않을 경우, 질의들은 가까운 영역의 데이터베이스부터 질의된 데이터를 찾기 위해 접근한다. 다음은 이동 클라이언트의 위치 설정에 따른 두 가지 영역에서의 실험 결과이다

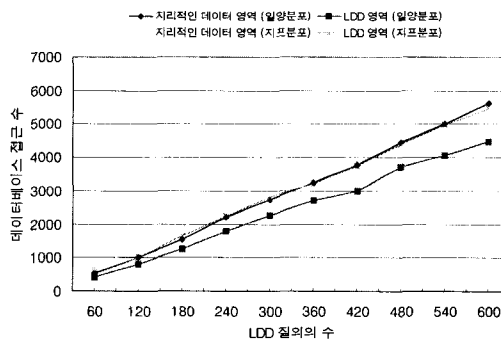


그림 5 다양한 LDD를 요구하는 성향의 이동 클라이언트 질의 처리(광의의 '가까운')

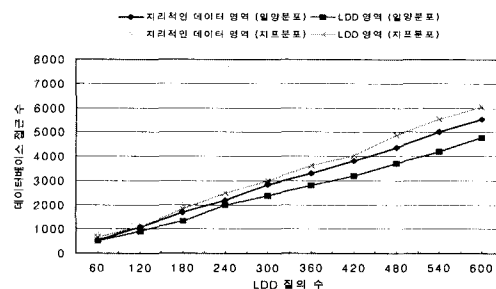


그림 6 소수의 LDD에 집중되는 성향의 이동 클라이언트 질의 처리(협의의 '가까운')

그림 5는 실험 모델에서 설정한 이동 클라이언트들의 위치가 전체 영역에 걸쳐서 거의 일정하게 분포한 경우(일양분포 사용)와 소수의 지역에 집중적으로 분포되어 있는 경우(지프분포 사용)로 나누어 지리적인 데이터 영역과 LDD 영역에서 질의 요청을 실험한 결과이다. 이때 이동 클라이언트의 질의 성향으로는 여러 가지 LDD의 종류에 대하여 골고루 접근하는 경우를 실험하였다. 그러한 LDD 질의를 처리하기 위해 데이터베이스를 접근하는 횟수를 측정한 결과 두 분포에서 모두 제안하는 LDD 영역의 데이터베이스 접근 수가 더 작음을 알 수 있다. 또한 몇 개의 지역에 집중적으로 이동 클라이언트가 분포되는 경우, 질의의 성향이 전체적으로 다양한 데이터를 요구하는 이동 클라이언트들이므로 그 지역에서 바로 찾을 수 없는 데이터가 많을 수 있다. 이것은 그만큼 다른 영역의 데이터베이스를 접근하게 되는 빈도를 높게 하여 일양분포에서 보다 지프분포에서 모두 데이터베이스 접근수가 더 많음을 알 수 있다. 또한 지리적인 데이터 영역(일양분포)과 LDD 영역(지프분포)의 결과를 보면 제안하는 LDD 영역에서는 이동 클라이언트가 집중적으로 모이는 경우라도 지리적인 거리를 고려한 데이터 영역에 이동 클라이언트가 고르게 분포한 경우의 질의 처리를 위한 데이터베이스 접근수와 거의 유사한 정도만 데이터베이스 접근이 발생하고 있다. 이것은 이동 클라이언트가 더 많이 집중될수록 그 성능의 차이가 더욱 커질 수 있음을 보이는 결과로 제안하는 LDD 영역 구성의 중요성을 나타낸다.

그림 6은 그림 5에서와 같이 실험 모델에서 설정한 두 가지의 이동 클라이언트 분포를 갖는 각 영역설정에서 질의의 성향이 소수의 위치중속 데이터에 집중되는 경우를 실험하였다. 즉, 대부분의 이동 클라이언트가 공통적으로 갖는 큰 관심사가 있거나 꾸준한 인기를 갖는 위치중속 데이터가 존재하는 상황을 설정하였다. 이 상황에서는 그림 5의 결과와 비교하여 전체적으로 모두 데이터베이스 접근 수가 더 증가하였음을 알 수 있다. 이것은 하나의 영역에서 집중적으로 요청되는 데이터가 없을 경우 그만큼 더 많은 다른 데이터베이스를 접근해서 원하는 데이터를 찾기 때문이다. 그러나 이 작업부하에서도 제안하는 LDD 영역의 성능이 모두 우세함을 알 수 있다.

다음 실험은 하나의 영역 크기를 더 작게하여 '가까운'의 의미를 협소하게 설정한 결과이다. 즉, 3절의 수식 2에 의해서 1이 될 수 있는 범위를 작게 설정하여 세밀한 조밀도를 부여한 환경에 해당한다. 그림 5, 6에서와 같은 이동 클라이언트의 설정에 기반하여 영역 구성의

크기를 2km 이내로 협소하게 설정한 실험의 결과는 그림 7, 8과 같다.

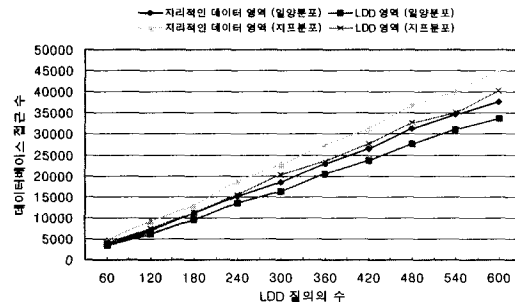


그림 7 다양한 LDD를 요구하는 성향의 이동 클라이언트 질의 처리(협의의 '가까운')

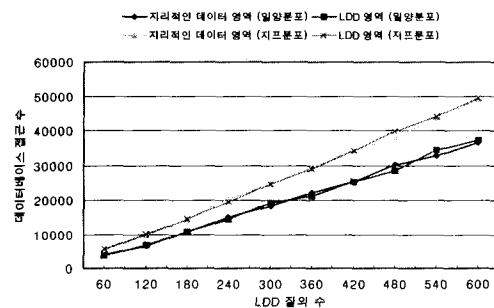


그림 8 소수의 LDD에 집중되는 성향의 이동 클라이언트 질의 처리(협의의 '가까운')

그림 7과 8의 결과를 보면 영역의 크기를 좀 더 작게 설정한 경우 데이터베이스 접근 수가 그림 5와 6에 비하여 많아졌음을 알 수 있다. 서비스되는 전체 영역이 작아짐으로써 '가까운'의 정도도 보다 더 가까워져야 하고, 데이터 구성도 좀더 세밀하게 구성해야 합당한 정확도를 높일 수 있다. 이것은 하나의 영역에 해당하는 범위가 작아지므로 조밀도를 그만큼 세분화시킨 상황이다. 그림 7의 결과에서 보면 '가까운'의 의미를 광의로 설정한 경우와 유사한 결과를 얻을 수 있음을 알 수 있다. 소수의 LDD에 집중되는 질의 성향과 작은 영역의 범위를 설정한 그림 8의 결과에서는 지리적인 데이터 영역과 제안하는 LDD 영역에서의 결과가 거의 유사하였다. 또한 이 경우에서 질의하는 이동 클라이언트의 위치가 소수의 영역에 집중되더라도 성능에 크게 영향을 주지 않음을 알 수 있다. 그러나, 협소하게 영역을 설정한 두

결과 모두 데이터베이스 접근 수는 모두 증가하였지만 역시 지리적인 영역에서보다 LDD 영역에서 동일하거나 보다 작은 데이터베이스 접근으로 질의를 처리할 수 있었다.

6. 결론

활성화되고 있는 위치중속 데이터에 대한 연구는 이동 컴퓨팅 환경에서 데이터를 접근하는 요구 및 서비스의 증가를 잘 반영한다. 위치중속 데이터의 지원은 이동 클라이언트에게 새로이 이동한 위치의 알맞은 정보를 얻을 수 있는 여러 가지 위치기반 서비스를 지원할 수 있다. 본 논문에서는 그러한 위치중속 데이터의 접근을 효율적으로 지원하기 위한 데이터 영역을 구성하는데 초점을 두었다. 위치중속 데이터를 요구하는 질의에서 거리의 모호성은 정확하고 적합한 데이터 값을 제공하기 위하여 고려해야할 사항이므로 본 논문에서는 위치중속 데이터간의 지역성, 관련성 그리고 거리 정도를 정량화하여 그러한 모호성을 해결하고자 하였다. 그 관련도에 기반하여 제안한 LDD 영역은 지역별 위치중속 데이터만을 관리하는 지리적인 데이터 영역에 비하여 이동 클라이언트의 위치중속 데이터 접근 성향 및 군집성 그리고 광의 및 협의의 '가까운' 정도에 따라 모두 우세한 결과를 얻을 수 있었다.

참고 문헌

- [1] T. Imielinski and B. R. Badrinath, "Querying in Highly Mobile and Distributed Environments," *Proceedings of International Conference on Very Large DataBase*, pp.41-52, August 1992.
- [2] T. Imielinski and B. R. Badrinath, "Data Management for Mobile Computing," *SIGMOD Record*, 22(1), pp.34-39, March 1993.
- [3] G. H. Forman and J. Zahorjan, "The Challenges of Mobile Computing," *CSE Technical Report 93-11-03, Univ. of Washington* (and in *IEEE Computer*, pp.39-47, 1994), 1994.
- [4] A. Rakotonirainy, "Trends and Future of Mobile Computing," *10th International Workshop on Database and Expert Systems Applications*, pp.136-140, September 1999.
- [5] M. H. Dunham and V. Kumar, "Location Dependent Data and its Management in Mobile Databases," *9th International Workshop on Database and Expert Systems Applications*, pp.414-419 1998.
- [6] J. B. Lim, A. R. Hurson and K. M. Kavi, "Concurrent Data Access in Mobile Heterogeneous Systems," *Proceedings of the 32nd Hawaii International Conference on System Sciences*, January 1999.
- [7] S. K. Madria, B. Bhargava, E. Pitoura and V. Kumar, "Data Organization Issues for Location-Dependent Queries in Mobile Computing," *ADBIS-DASFAA 2000, LNCS 1884*, pp.142-156, 2000.
- [8] K. Ku and Y. Kim, "Moflex Transaction Model for Mobile Heterogeneous Multidatabase Systems," *10th International Workshop on Research Issues on Data Engineering*, pp.39-46, 2000.
- [9] M. H. Dunham and A. S. Helal, "Mobile Computing and Databases: Anything New?," *SIGMOD Record*, Vol. 24, No. 4, Dec. 1995.
- [10] H. Kottkamp and O. Zukunft, "Location-Aware Query Processing in Mobile Database Systems," *Proceedings of the ACM Symposium on Applied Computing*, pp.416-423, 1998.
- [11] K. Tan and B. Ooi, *Data Dissemination in Wireless Computing Environments*, Kluwer Academic Publishers, 2000.
- [12] D. Barbara and T. Imielinski, "Sleepers and Workholics: Caching Strategies in Mobile Environments," *Proceedings of the ACM SIGMOD Conference on Management of Data*, pp.1-12, 1994.
- [13] E. Pitoura and G. Samaras, *Data Management for Mobile Computing*, Kluwer Academic Publishers, 1997.
- [14] A. P. Sistla, O. Wolfson, S. Chamberlina, and S. Dao, "Modeling and Querying Moving Objects," *Proceedings of the IEEE International Conference on Data Engineering*, pp.422-432, 1997.
- [15] B. Zheng and D. L. Lee, "Processing Location-Dependent Queries in a Multi-cell Wireless Environment," *Proceedings of the 2nd ACM International Workshop on Data Engineering for Wireless and Mobile Access (MobiDE01)*, pp.54-65, May 2001.
- [16] S. Navathe, S. Ceri, G. Wiederhold and J. Dou, "Vertical Partitioning Algorithms for Database Design," *ACM Transaction on Database Systems*, Vol. 9, No. 4, pp.680-710, December 1984.
- [17] S. B. Navathe and M. Ra, "Vertical Partitioning for

- Database Design: A Graphical Algorithm," *ACM SIGMOD Record*, vol.18, no.2, pp.440-450, June 1989.
- [18] J. Son, K. Choi and M. Kim, "A Vertical Partitioning Algorithm based on Fuzzy Graph" *Journal of KISS: Database*, Vol. 28, No. 3, September, 2001.
- [19] M. Tamer Ozsü and, P. Valduriez, *Principles of Distributed Database Systems*, Second Edition, Prentice Hall, 1999.
- [20] G. J. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, 1995
- [21] M. Nascimento and Y. Theodoridis. Benchmarking spatio-temporal databases: The gstd software. <http://www.cti.gr/RD3/GSTD>.
- [22] A. Seydim, M. Dunham and V. Kumar, "Location Dependent Query Processing," *Proceedings of the 2nd ACM International Workshop on Data Engineering for Wireless and Mobile Access (MobiDE01)*, pp.47-53, May 2001.



유 재 혁

1997년 고려대학교 전산학과 이학사
 1999년 고려대학교 일반대학원 컴퓨터학과 이학석사. 1999년~현재 고려대학교 일반대학원 컴퓨터학과 박사과정. 관심분야는 동시성 제어, 분산 데이터베이스 이
 동 컴퓨팅 시스템

황 중 선

정보과학회논문지 : 정보통신
 제 30 권 제 1 호 참조