

이동 환경에서의 위치 데이터에 기반한 마이닝 시스템[†]

백옥현⁰ 이준욱 이금우 김영균 류근호
충북대학교 데이터베이스연구실
(ohpaek, junux, gmoo, ykkim khryu)@dblab.chungbuk.ac.kr

A Mining System Based on Location Data in Mobile Environments

Ok Hyun Paek⁰ Jun Wook Lee Keum Woo Lee Young Kyun Kim Keun Ho Ryu
Dept. of Computer Science, Chungbuk National University

요 약

이동 컴퓨팅 환경에서의 응용은 사용자의 위치 정보를 이용해서 개인의 위치에 따른 서비스를 제공할 것이 요구된다. 기존의 데이터 마이닝 분야에서는 이동 환경에서 발생하는 데이터는 고려되지 않았다. 이 논문에서는 이동 컴퓨팅 환경에서 발생하는 위치 데이터를 기반으로 한 마이닝 기술과 요구사항을 논의하고, 이를 기반으로 위치 데이터에서 유용한 지식을 추출할 수 있는 데이터 마이닝 시스템을 제안한다. 이 시스템은 위치 데이터베이스와 GIS, 사용자 데이터베이스, 콘텐츠 데이터베이스와 통합, 연계되어 유기적으로 데이터 마이닝을 수행한다. 또한 사용자와의 상호작용이 가능한 인터페이스와 가시화 도구를 제공한다. 시스템을 구성하는 각각의 구성요소에 대해 기술하며, 마이닝된 결과를 응용할 수 있는 부분을 보인다.

1. 서 론

이동 컴퓨팅 환경이라는 새로운 패러다임에 따라 이것을 이용한 m-Business, m-CRM 등 새로운 개념의 응용 부분들이 등장하고 있다. 이동 전화, PDA 등의 핸드 헬드 기기를 포함한 무선 기기가 보편화 되고 있으며, 이러한 무선 기기의 확산은 소형화, 넓은 대역폭의 지원, 그리고 향상된 디스플레이 기술의 지속적인 발전에 의해 가능하게 되었다. 또한 GPS(Global Positioning System) 등의 위치 추적 기술에 의해 어느 사용자가 어느 위치에 있는지를 알 수 있게 되었다. 이러한 인프라가 형성됨에 따라 위치 데이터를 이용한 응용들이 등장하였으며, 데이터 마이닝 분야에도 새로운 도전을 제공하였다. 이동 환경에서 발생하는 위치 데이터에서의 마이닝에 대한 연구에 대한 전례는 거의 없는 실정이다. 위치 데이터에서 추출된 지식으로 다음과 같은 예를 들 수 있다.

- 지식 1 : A지역을 방문하는 고객의 70%는 B지역을 방문한다.
- 지식 2 : 전체 고객의 25%는 A → B → C → D 순으로 지역을 방문한다.
- 지식 3 : 20대 여성 고객의 60%는 B → D → E 의 이동 패턴을 가진다.
- 지식 4 : A → B → C 지역으로 이동하는 자동차의 90%는 C 지역에서 130km/h 이상의 속력을 낸다.
- 지식 5 : 저녁 6시부터 8시 사이에 B지역을 방문하는 고객의 50%는 20대 여성이다.

지식 1의 예는 지역 이동 간의 연관성을 탐사한 예이며, 지식 2는 순차 패턴을 적용한 예이다. 지식 3은 분류된 고객의 순차패턴을 구한 예이고, 지식 4는 규칙 탐사에 제약조건을 적용한 예이다.

지식 5는 시간 지식과 분류된 지식의 결합을 보이고 있다.

이런 지식들은 향후 핵심적인 서비스로 보고되고 있는 위치 기반 서비스(Location Based Service)에 적용될 수 있다. 위치 기반 서비스란 이동중인 사용자에게 유선 및 무선 통신을 통하여 사용자의 위치와 관련된 정보를 제공하는 서비스를 통칭한다[7]. 위치 기반 응용 서비스로는 인접 지역 정보 제공의 무선 GIS, 최적 경로 계산, 교통 혼잡 정보등과 같은 교통 정보, 차량 항법 및 물류 관계, 위치 기반 CRM(Customer Relationship Management) 그리고 사용자 지역을 기반으로 한 비교 쇼핑 등의 인터넷 부가 서비스 등이 다양하다. 이러한 서비스들은 단순히 위치만을 기반으로 하는 것이 아니라 위에서 예시한 것과 같은 사용자의 이동 패턴과 프로필 등의 성향을 이용함으로써 더욱 유용한 정보를 제공할 수 있다.

이 논문에서는 위와 같은 부분에 응용하기 위해서 바탕이 되는 기술들을 보이고, 이를 기반으로 하는 시스템의 모델과 설계 원칙을 기술한다. 이런 요구사항을 반영하는 위치 데이터 기반의 마이닝 시스템을 제안하고, 마지막으로 이런 시스템이 적용될 수 있는 응용분야를 기술한다.

2. 위치 데이터에서의 데이터 마이닝을 위한 기술

2.1 이동 객체 데이터베이스

이동 객체[5]는 시간에 따라 객체의 공간 정보가 연속적으로 변경되는 개체로 크게 이동 점과 이동 영역으로 나눌 수 있다. 이동 점은 시간에 따라 객체의 위치가 변하는 것으로, 사람이나 동물, 자동차, 비행기 등을 예로 들 수 있다. 이동 영역은 시간에 따라 객체의 위치 뿐 아니라 모양까지 변하는 것으로서, 한 국가의 행정 영역이나 폭풍의 영향권 등을 예로 들 수

[†] 이 논문은 1999년도 KOSEF 특정기초(#1999-2-303-006-3) 연구비 지원으로 수행되었음

있다.

시간에 따라 위치가 연속적으로 변하는 이동 객체를 표현 하기 위한 데이터 모델은 연속적 모델과 이산적 모델로 구분된다. 연속적 모델은 시공간 이동 객체를 무한한 점들의 집합으로 표현하고, 이산적 모델은 이동 객체를 유한한 점들의 집합으로 표현한다. 기존의 데이터베이스로 이동 객체를 직접 표현할 경우 이동 객체의 위치가 계속적으로 변하므로 이를 기록하기 위한 빈번한 연산이 요구되고, 이동 객체의 현재 상태만을 기록하므로 과거 및 향후 이동 위치와 관련된 질의에 적절히 응답할 수 없다. 이동 객체 데이터베이스는 객체의 유효 시간과 공간 좌표를 기본적으로 가지며, 객체의 식별자, 위치 좌표, 시간 데이터를 사용하여 특정 식별자를 가지는 객체가 시간 T에 어느 위치에 있었는지를 나타내는 시공간 데이터 형태로 저장된다.

2.2 시공간 데이터 마이닝

이동 환경에서의 객체의 위치는 시간에 따라 지속적으로 변하기 때문에 시간과 공간의 개념이 모두 포함될 것이 요구된다. 시공간 데이터 마이닝은 '시공간 데이터로부터 이전에 잘 알려지지 않았지만, 묵시적이고 잠재적으로 유용한 시공간 지식을 추출하는 기술'로 정의될 수 있다. 시공간 마이닝은 첫째, 공간 시스템에 시간개념이 포함된 것과, 둘째, 시간 시스템에 공간 차원을 추가한 두 가지로 살펴볼 수 있다. 일반적으로, 타임 스탬프가 있는 공간 데이터가 사용되기 때문에 시공간 데이터마이닝에 관한 연구는 전자 쪽으로 진행되어 왔다[2, 4]. [3]에서는 기후분석을 위한 패턴을 찾기 위해 대규모의 데이터셋에 적합한 시스템이 필요하다는 것에 착안하여 분산 병렬 질의와 분석을 수행할 수 있는 CONQUEST 시스템의 설계, 구현, 응용에 관해 기술하였다. [4]는 시공간 데이터셋에서 빈발적으로 발생하는 이벤트의 시퀀스를 시공간 패턴이라 정의하고, 환경 데이터에서 이러한 시공간 패턴을 효율적으로 마이닝 할 수 있는 DFS_MINE 알고리즘을 제안하였다.

3. 위치 데이터 기반의 마이닝 시스템 설계

3.1 위치 데이터 마이닝 모델

위치 데이터 기반의 마이닝 시스템을 설계하기에 앞서 위치 데이터에서의 지식 탐사를 위한 처리 흐름을 모델링한다. 모델은 그림 1과 같이 Interface, Focus, Mining Operation, Evaluation, Update 프로세스로 구성된다. 데이터는 Interface

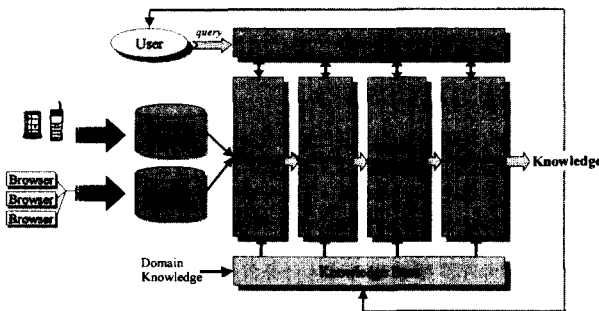


그림 1. 위치 데이터 마이닝 모델

를 통해 추출되며, Focus 프로세스를 통해 필요한 부분만을 취한다. Mining Operation 프로세스에서 탐사된 규칙은 Evaluation 프로세스에서 평가되고, 유용하다고 판단되는 결과만이 사용자에게 보여지고 지식 베이스에 저장된다.

3.2 시스템 설계 원칙

위치 데이터를 기반으로 하는 마이닝은 객체 위치의 잦은 업데이트로 인한 마이닝 연산의 비효율성, 이전에 찾았던 지식의 무효화 가능성, 그리고 실시간 지식 제공 등의 특성을 지니고 있다. 따라서 정확한 설계 원칙에 따라 개발할 필요가 있다. 위치 데이터 기반의 마이닝 시스템을 설계하는 데 있어서의 원칙은 다음과 같은 네 가지로 요약될 수 있다.

첫째, 시공간 데이터 마이닝 연산을 제공한다. 시간에 따라 공간 속성이 변하는 위치 데이터를 위해서 기존의 데이터 마이닝 시스템에서는 제공하지 않았던 시공간 마이닝 연산을 제공함으로써 시간 및 공간 의미를 갖는 지식을 탐사할 수 있도록 한다.

둘째, 이동 객체의 위치가 시간의 흐름에 따라 빈번하게 변화하므로, 그에 따른 효율적인 연산의 수행을 보장한다.

이동 객체는 시간에 따라 동적으로 그 위치가 변화하기 때문에 이에 대한 마이닝 연산 역시 빈번하게 수행되게 된다. 따라서, 연산의 수행을 효율적으로 수행할 수 있도록 새로운 부분에 대해서만 규칙을 탐사하는 점진적인 방법을 이용한다.

셋째, 발견된 지식은 최근의 경향을 반영한다.

지식 자체의 유용성 측면 외에도 탐사된 지식의 유효 기간에 대한 평가가 필요하다. 즉, 과거에 탐사된 시점은 과거 시점에서는 옳지만 현재 시점에서는 더 이상 옳지 않을 수 있다. 따라서, 규칙은 지속적으로 유지·관리되어야 한다.

넷째, 특정한 데이터베이스에 종속적이지 않아야 한다.

이동 객체의 위치 정보를 저장하는 이동 객체 데이터베이스는 데이터가 채택하는 모델에 따라 다른 방식으로 저장된다. 따라서, 특정한 이동 객체 데이터베이스에 종속되기 보다는 인터페이스를 통해 이질적인 데이터베이스를 지원하도록 해야한다. 이런 방식은 기존 데이터베이스를 변경하지 않고 저렴한 비용으로 데이터 마이닝 기능을 구현할 수 있다.

3.3 위치 데이터 기반의 마이닝 시스템 아키텍처

앞서 기술한 데이터 마이닝 모델과 설계 원칙을 고려하여 위치 데이터를 위한 마이닝 시스템은 그림 2와 같이 설계되었다. 수행해야 할 마이닝 태스크를 하나의 질의로 간주하여, 수행해야 할 태스크, 입력 데이터, 사용하는 알고리즘, 알고리즘의 임계치 등을 사용자 인터페이스에서 질의로 입력받는다. 수행될 마이닝 태스크가 질의로 명시됨으로써 다양한 마이닝 연산을 표현할 수 있는 표현성과 확장성을 제공하며, 이질적인 데이터베이스 시스템에서도 사용할 수 있다는 장점을 가진다.

수행해야 할 마이닝 연산은 시간 패턴, 공간 연관 규칙, 공간 분류, 클러스터링의 대표적인 네 가지 범주로 규정하였으며, 추출된 지식은 지식 저장소에 저장된다. 지식은 최근의 경향을 담고 있어야 하므로 점진적으로 유지·갱신된다. 데이터 마이닝 연산의 효율성을 보장하기 위해서, 마이닝 연산을 수행할 때마다 전체 데이터베이스를 탐색하는 기존의 방법보다는, 필요한 부분에 대해서만 지식 탐사를 수행함으로써 효율적인 마이닝 연산이 가능하도록 하였다. 데이터 소스는 이동 객체 데이터베이스와 GIS, 트랜잭션 데이터베이스, 그리고 사용자 프로파일 데이터베이스를 고려하였으며, 이러한 데이터베이스에서의 데이터 추출은 인터페이스에 의해 이루어진다.

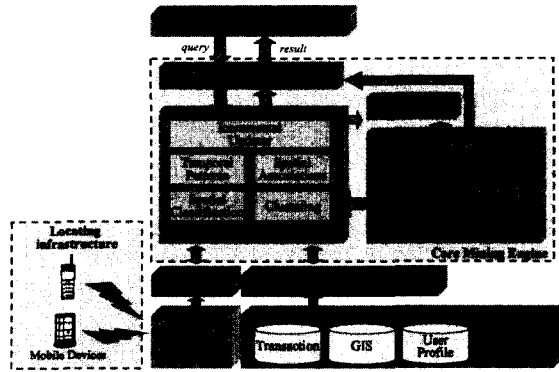


그림 2. 위치 데이터 기반의 마이닝 시스템 아키텍처

각 구성 요소의 역할은 다음과 같다. 사용자 인터페이스(User Interface)는 사용자의 입력값에 따라 반복적인 작업을 통해 유용한 지식을 추출하거나 수행할 마이닝 기법을 선택할 수 있게 해주며, 지식 탐사 결과를 2차원, 3차원 형태로 축약하여 보여주는 가시화 도구를 제공한다. 마이닝 질의 처리기(Mining Query Processor)는 사용자가 입력한 질의의 문법과 의미를 분석하여 마이닝 작업을 수행하고 결과를 전달한다. 마이닝 모듈(Mining Modules)은 위치 데이터와 다른 데이터를 연계해서 유용한 지식을 추출하기 위한 마이닝 연산을 구현한 모듈이다. 사용자가 질의에서 명시한 입력 데이터에서 지식을 탐사한 후, 결과를 지식 베이스에 저장한다. 지식 저장소(Knowledge Base)는 지식 탐사 모듈에 의해 발견되어진 규칙이 저장된다. 생성된 규칙은 규칙 적용에 있어 이전에 생성된 규칙이 항상 유용한 규칙으로 적용되어지지 않을 수 있기 때문에 갱신 작업이 이루어진다

4. 적용 및 평가

4.1 Mobile 전자상거래

Mobile 전자 상거래란 휴대용 컴퓨터, PDA 또는 이동 전화 등의 모바일 장비들을 통해서 구매를 하는 것을 말한다. Mobile 전자상거래의 트랜잭션에 대한 지식 탐사와 위치 정보를 결합하는 것은 유용한 마케팅, 판매 정보를 얻을 수 있게 해 줄 뿐 아니라, 고객 개인에게 개인화된 서비스를 제공해 줄 수 있다. 예를 들면 전자상거래 시, 고객의 평소 이동 성향을 고려해서 위치 기반의 광고를 제공한다는지, 관련된 상품을 추천 하는 등의 서비스가 가능하다.

4.2 개인화 추천 시스템

개인화 추천 시스템은 고객 정보 데이터, 제품 정보 데이터, 그리고 고객의 제품 구매 정보 등에 데이터 마이닝 기법을 적용하여 고객의 취향을 알아내고 각 고객의 취향에 맞는 추천 리스트를 제공하는 시스템이다. 이것은 사용자의 정보와 사용자의 위치 정보를 얻어 현재 위치에서 제공 가능한 서비스 중 사용자가 관심을 가질만한 정보를 제공한다. 이러한 시스템에서 중심이 되는 부분은 사용자에게 적절한 정보를 선별하기 위한 마이닝을 수행하는 모듈이다.

4.3 지능적 위치 기반 서비스

위치기반 서비스(Location-Based Service)는 이동중인 사용자에 게 무선 및 유선 통신을 통하여 사용자의 위치에 기반한 검색과

다양한 정보를 제공하는 서비스를 청한다. 위치만을 기반으로 하는 것이 아니라, 사용자가 과거에 요청한 콘텐츠의 이력을 마이닝하여 사용자 개인이 관심 있어 할 만한 내용을 보여주는 등의 지능화된 위치 기반 서비스가 가능하다. 적용 가능한 데이터 마이닝 기법은 연관규칙, 순차패턴, 클러스터링 등 응용하고자 하는 부분에 따라 다양하다.

5. 결론 및 향후연구

이 논문에서는 이동 컴퓨팅 환경에서 발생하는 위치 데이터를 기반으로 한 데이터 마이닝에 대하여 논의하였다. 데이터 마이닝 시스템을 설계하기 위한 데이터 마이닝 모델과 설계 원칙을 기술하였고, 이를 바탕으로 하여 위치 데이터와 GIS, 고객 데이터를 연계하여 유용한 지식을 추출할 수 있는 데이터 마이닝 시스템을 설계하였다. 마이닝된 지식은 Mobile 전자상거래에 이용되거나, 개인화 추천 시스템, 그리고 지능적 위치 기반 서비스에 적용될 수 있다.

향후에는 마이닝 시스템 내에서의 시간과 공간 의미를 갖는 마이닝 질의어의 설계와 제시한 시스템 내의 여러 연산들을 구현하는 작업이 필요하다.

참고문헌

- [1] C.S.Jensen, A.Friis-Christensen, T.B. Pedersen, D. Pfoser, S.Saltenis, N.Tryfona, Location-Based Services -A Database Perspective, In Proceedings of ScanGIS, 2001.
- [2] K.Koperski, J.Adhikary, J.Han, Spatial Data Mining : Progress and Challenges Survey Paper, In Proceedings of ACM-SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, 1996.
- [3] P.Stolorz, H.Nakamura, Fast Spatio-Temporal Data Mining of Large Geophysical Datasets, 1st international conf. on Knowledge Discovery and Data Mining, 1995
- [4] E.Tsoukatos and D.Gunopoulos, Efficient Mining of SpatioTemporal Patterns, 7th International Symposium on Spatial and Temporal Databases (SSTD), 2001.
- [5] O.Wolfson, B.Xu, S.Chamberlain, L.Jiang, Moving Objects Databases : Issues and Solutions, In Proceedings of the 10th International Conf. on Scientific and Statistical Database Management, 1998.
- [6] 이용준, 시간 관계 규칙 탐사를 위한 데이터마이닝 기법, 이학박사학위 논문, 충북대학교 대학원, 2001.
- [7] 이준표, LBS(위치 기반 서비스), 한국 지리 정보, 통권 53호 2001년 6월 24-15, 2001.