

모호성을 고려한 이동 객체의 시간 패턴 탐사

이양우⁰ 이준욱 김룡 류근호
충북대학교 데이터베이스 연구실
(dooji, junux, kimlyong, khryu)@dbl lab.chungbuk.ac.kr

Temporal Pattern Mining of Moving Objects considering Ambiguity

Yang Woo Lee⁰ Jun Wook Lee Lyong Kim Keun Ho Ryu
Database Laboratory, Chungbuk National University

요 약

위치 기반 서비스가 무선 인터넷의 새로운 이슈로 떠오르고 있다. 이동 객체의 패턴 마이닝은 이동 객체의 시간 패턴을 탐사함으로써 이동 객체에 위치에 기반한 유용한 서비스를 제공할 수 있게 해준다. 이동 객체는 시간에 따라 빈번하게 이동하기 때문에 패턴도 최근의 경향을 반영하기 위해 빈번하게 탐사되어야 한다. 따라서 점진적으로 시간 패턴을 탐사하는 접근법이 요구된다.

이 논문에서는 이동 객체의 시간 패턴을 탐사하는데 있어서 측정된 위치 데이터가 가질 수 있는 모호성을 제시했다. 또한 모호성을 고려한 시간 패턴 마이닝을 위해 패턴 탐사 단계에서의 모호성의 처리를 위해 모호성을 원인에 따라 세 가지 임계치를 정의하였다. 그리고 이러한 임계치를 고려한 시간 패턴 마이닝 프로시저 구조를 제시하였다.

1. 서 론

이동 컴퓨팅(mobile computing)이란 사용자가 이동 중일 때도 휴대용 컴퓨터와 무선 통신을 통해 네트워크에 접속할 수 있도록 해주는 컴퓨터 패러다임을 말한다[1]. 이런 환경에서 위치 기반 서비스(location based services)가 무선 인터넷의 새로운 이슈로 떠오르고 있다[2].

위치 기반 서비스의 대상이 되는 이동 객체(moving object)는 시간의 흐름에 따라 객체가 이동하면서 위치 및 모양이 연속적으로 변화하는 특성을 갖는 데이터를 말한다[3]. 우리는 이런 특성을 갖는 데이터에 마이닝 기법을 적용하여 일정한 패턴이나 경향을 발견할 수 있다. 발견된 패턴의 예는 다음과 같다.

- 고객의 20%가 A→B→C순으로 지역을 방문한다.
- 저녁 6시부터 10시 사이에 고객의 30%는 B→C→D 지역을 방문한다.

이런 지식들은 위치 기반 서비스에서 사용자의 이동 패턴과 프로필 등의 성향을 이용함으로써 사용자에게 더욱 특화된 개인화된 서비스를 제공할 수 있다.

이동 객체는 시간에 따라 빈번하게 이동하기 때문에 패턴 역시 최근의 경향을 반영하기 위해 빈번하게 탐사되어야 한다. 이런 이동 객체의 시간 패턴을 탐사하기 위해 기존의 연구에서 제안된 점진적으로 시간 패턴을 탐사하는 기법을 이용한다. 이 논문에서는 기존의 점진적인 시간 패턴 마이닝에 이동 객체의 이동 경로에 포함될 수 있는 모호성을 고려하여 좀 더 유용한 패턴을 탐사하고자 한다.

모호성을 갖는 패턴의 예로 위의 첫 번째 예에서 이동 객체가 각 지역에서 머문 시간이 A(120 분)→B(10 분)→C(80 분)일 때, B 지역은 A 지역에서 C 지역으로 가기 위한 통로의 역할일

뿐, 이동 객체가 어떤 목적을 가지고 이동 했다고 볼 수 없기 때문에 방문의 측면에서 해석할 수 없다는 의미적 분석을 할 수 있다. 이처럼 기존의 기법이 이동 객체의 이동한 경로만을 이용하여 패턴을 찾기 때문에 이 기법으로 찾은 빈발 패턴에는 모호한 패턴이 포함될 가능성이 존재하게 된다.

이 논문에서는 2 장에서 위치 기반 서비스와 이동 객체와 패턴에 대해 알아보고 3 장에서 패턴 탐사를 위한 이동 객체의 데이터 모델을 정의한다. 4 장에서 기존의 시간 패턴 마이닝과 모호성에 대해 알아보고 5 장에서 이를 해결하기 위한 세 가지 임계치를 제안하고 이를 적용한 시간 패턴 마이닝 프로시저 구조를 제안한다. 마지막으로 6 장에서 결론을 맺는다.

2. 관련 연구

위치 기반 서비스는 이동 객체의 위치 정보를 측정하고 이를 이용해 개인화된 서비스를 하는 것이다[2]. 이를 위한 관련 연구로써는 이동 객체와 시간 패턴 마이닝 등이 있다.

2.1 이동 객체

이동 객체는 시간이 흐름에 따라 객체의 위치나 영역과 같은 공간 정보가 연속적으로 변화하는 객체를 말하며 크게 이동 점(moving point)과 이동 영역(moving region)으로 나눌 수 있다 [3, 4]. 이동 점은 객체의 위치가 변하는 것으로 택시, 배, 비행기, 사람 등이 그 예이다. 이동 영역은 객체의 위치뿐만 아니라 모양까지 변하는 것으로, 이러한 이동 영역의 예는 민족의 이동, 숲의 발달, 폭풍의 이동 경로 등이 있다. 이 논문에서는 이동 점에 대한 시간 패턴 마이닝에 대해서만 다룬다.

2.2 시간 패턴

시간 패턴은 데이터의 시간 속성과 관련해서 찾을 수 있는

시간적인 지식으로써 순차 패턴, 주기 패턴, 시간 관계로 분류할 수 있다.

먼저, 순차 패턴은 항목 집합으로 구성된 트랜잭션들 간에 순차적으로 발생하는 특정 항목 집합의 패턴을 말한다[5]. 주기 패턴은 특정 패턴이 정기적으로 발생하는 현상으로, 매년이나 매주와 같은 시기에 주기적으로 발생하는 패턴을 말한다[6]. 시간 관계(temporal relationship)는 시간 간격을 가진 여러 사건 간의 시간 관계를 포함한 지식이다. 순차 패턴처럼 사건의 연속이 아니라, 여러 사건이 특정 기간 동안 중복되어 발생한다는 특성을 가진다.

3. 이동 객체를 위한 데이터 모델

이동 객체는 시간에 따라 위치가 변하는 동적 데이터를 갖는다. 그러나 정적 데이터만을 다루던 기존의 데이터 모델로는 이런 이동 객체를 정확히 표현하기가 어렵다. 이동 객체의 표현을 위한 여러 가지 데이터 모델이 제안되었다[7, 3].

이동 객체는 객체의 식별자와 공간 상에서의 위치 변화의 집합으로 표현된다. 이러한 위치 변화는 공간상에서의 위치와 유효 시간으로 표현된다. 공간상에서의 위치는 좌표로 표현하거나 기지국에 포함되는 영역만으로 표현될 수 있다. 여기서는 이동 객체의 공간 속성을 2차원 공간상의 x, y 좌표로서 표현하는 방법을 택한다. 이동 객체 Mpoint는 다음과 같이 정의된다[8].

[정의 3.1] $Mpoint = oid, \{(VT_1, L_1), (VT_2, L_2), \dots, (VT_n, L_n)\}$

단, oid는 객체의 식별자를 가리킨다. VT_i는 유효 시간이고, L_i는 VT_i에 측정된 객체의 위치를 나타내며 (x_i, y_i)로 표현한다.

이동 객체 데이터베이스는 이러한 Mpoint의 집합으로 정의한다.

4. 문제 정의

4.1 이동 객체의 시간 패턴 마이닝

[8]에서 제시한 시간 패턴 마이닝에서 시간의 흐름에 따라 데이터베이스에 있는 임의의 객체의 방문 지역을 정렬한 것을 이동 시퀀스(moving sequence)라 정의했고, 한 시퀀스의 지지도는 그 시퀀스를 포함하는 전체 이동 시퀀스의 비율로 논의되었다. 여기서 한 시퀀스의 지지도가 최소 지지도(min_sup)를 만족하면 빈발 시퀀스라고 할 때, 시간 패턴 마이닝은 이런 이동 객체의 모든 빈발 시퀀스를 찾는 것이다.

이런 이동 객체의 데이터는 시간이 지남에 따라 계속 추가되기 때문에, [11]에서는 데이터의 점진적 갱신의 기법을 사용한 시간 패턴 마이닝을 제안하였다.

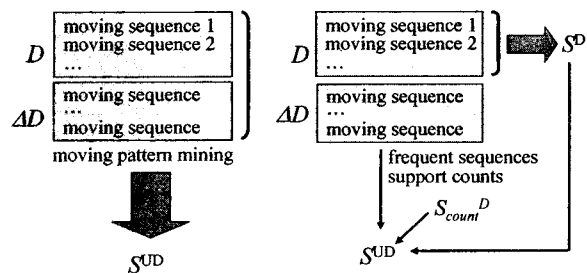
기존의 데이터베이스를 통해 얻은 빈발 시퀀스에 새로 추가된 데이터베이스에서 얻어진 시퀀스를 가지고 새로운 빈발 시퀀스를 얻는 이 기법은 전체 데이터베이스를 스캔하지 않고서 빈발 시퀀스를 얻음으로써 속도의 향상과 비용의 절감이라는 효율성을 입증하였다. 그림 1은 이 방법에 대한 개념적인 비교이다[11].

4.2 시간 패턴 마이닝에서 이동 객체 패턴의 모호성

임의의 이동 객체에 대한 시간 패턴 마이닝을 하는데 있어서 이동 시퀀스의 정확성은 매우 중요하고, 이를 통해 정확한 시간 패턴 마이닝을 할 수 있다. 이 논문에서 논의하고자 하는 모호성은 위치 추적 시스템에서 논의되는 측정 에러나 샘플링 에러[9, 10]를 말하는 것이 아니라, 이동 객체의 패턴에 잠재되어 나타날 수 있는 모호성을 의미한다.

패턴이 모호하게 되는 것은 이동 객체의 위치 데이터의 값이

나 위치 데이터들 사이의 관계에서의 모호성에 기인한다. 한 Mpoint내의 위치 데이터 중 측정하지 못해 추정된 데이터가 많은 경우로 패턴이 모호하게 될 수 있고, 위치 데이터 사이의 값이 연속적이지 못하고 시간상으로 너무 큰 갭을 갖거나 특정 지역에서 연속적으로 측정된 위치 데이터의 수가 너무 적은 경우에도 패턴이 모호하게 되는 원인이 된다. 예를 들어, Mpoint내의 많은 위치 데이터 중에서 추정 데이터의 비율이 30%를 넘게 되면 이 Mpoint의 값을 패턴 마이닝에 고려하지 않도록 조건을 설정하여 패턴의 모호성을 줄일 수 있다. 이상에서 알아본 것과 같은 원인으로 패턴에 나타날 수 있는 문제들을 패턴의 모호성이라 정의한다. 다음 장에서 이러한 패턴의 모호성을 분류하고 분류된 모호성을 다루기 위한 임계치를 각각 정의한다.



(a) Re-execution algorithm on UD (b) Incremental updating

그림 1 재 탐사와 점진적 갱신

5. 이동 객체 패턴의 모호성을 고려한 시간 패턴 마이닝

모호성을 가지는 이동 객체의 패턴에서 모호성을 제거하기 위한 여러 가지 임계치가 고려될 필요가 있다. 위치 데이터는 앞의 정의에 의해 표현되며 그 예는 다음의 표와 같다.

표 1 이동 객체 데이터베이스의 예

oid	VT	X	Y
001	2002/08/19/14/30	742410	876310
...
001	2002/08/19/15/40	742520	876410
001	2002/08/19/15/50	751330	863200
001	2002/08/19/16/00	774530	838600
...
001	2002/08/19/17/20	772541	838191
001	2002/09/21/20/20	776520	831480

5.1 공백 데이터의 추정 대한 처리

이동 객체의 위치 추적이 중간에 끊김이 없이 연속적인 것이 이상적인 것이긴 하지만, 주변 지형이나 건물의 영향, 기기 자체의 결함, 기기의 전원이 꺼졌다가 다시 들어온 경우 등의 이유로, 위치 추적은 중간에 공백 데이터를 포함하게 된다. 이런 경우 위치 추적에서 불확실성의 처리 측면에서 그 공백(missing) 데이터를 추정하게 된다.

예를 들어, 표1에서 일곱 번째 행은 공백 데이터에 대한 추정 데이터라는 것을 표현하기 위해 마지막 자리를 1로 처리했다. 그러나 같은 지역에서 연속적인 위치 데이터에서 추정 데이터가 일정 수준 이상 포함되면 신뢰도가 떨어지고 이는 패턴 대한 신뢰도를 떨어뜨리게 된다. 여기서는 이런 추정 데이터의

수준을 조절하기 위해 새로운 임계치를 정의한다. 그 정의는 다음과 같다.

[정의 5.1] 공백 데이터의 추정에 대한 임계치로 최대 추정(estimate) 수를 *max_est*로 정의한다. 이를 통해 같은 지역에서 연속적인 위치 데이터 중에 추정 데이터의 수가 *max_est*를 넘게 되면 이 위치 데이터를 시퀀스에서 제외할 수 있다.

5.2 지속 시간에 대한 임계치

위의 가정의 경우에 위치 검출 간격에 따라 동일 지역에서 얻어진 연속적인 위치 데이터의 수가 다를 것이다. 연속적인 위치 데이터의 수가 많을수록 이동 객체가 그 지역에서 머무른 시간이 길다는 것을 의미하고 위치 데이터의 수가 적은 경우에는 그 반대를 의미한다.

예를 들어, 표1에서 첫 번째 행에서 일곱 번째 행은 A→B→C지역을 차례로 이동했다고 할 수 있다. 그러나 네 번째 행과 같이 중간 지역 B에서 얻어진 위치 데이터의 수가 다른 지역에서의 데이터의 수보다 상대적으로 작기 때문에 지역을 방문했다기 보다 지나쳤다고 해석될 수 있다. 그러므로 연속적인 위치 데이터가 일정한 수 이상 존재하는 경우에만 그 지역의 방문에 대한 의미를 줄 수 있다.

[정의 5.2] 지속 시간에 대한 임계치로 최소 측정(determination)수를 *min_det*으로 정의한다. 이는 특정 지역에서 이동 객체가 *min_det*의 시간만큼 머물러야 그 지역이 방문한 지역에 포함될 수 있다는 조건으로 바꾸어 말할 수 있다.

5.3 위치 데이터의 시간 상의 간격에 대한 임계치

특정 지역에 대한 패턴 마이닝에서 대상이 되는 두 위치 데이터 사이의 간격이 너무 크면 그 두 데이터 사이에서 연관성을 찾는 것이 어렵다. 예를 들어 표 1의 일곱 번째 행과 여덟 번째 행에서 보면 두 데이터는 같은 지역에서 측정된 데이터이지만 시간 간격은 한달 이상이 된다. 그러므로 이 두 데이터를 같이 고려할 수가 없을 뿐만 아니라 패턴으로서도 별개의 것으로 고려될 필요가 있다.

[정의 5.3] 위치 데이터의 시간 상의 간격에 대한 임계치로 최대 허용 시간 간격(gap)을 *max_gap*으로 정의한다. 이를 통해 허용 시간 이상의 간격이 존재하는 경우 각각 고려되어야 한다.

5.4 시간 패턴 마이닝의 프로시저 구조

[8]에서 제안된 시간 패턴 마이닝에 임계치를 적용하여 시간 패턴 마이닝을 개선하였다. 시간 패턴 마이닝(MP:moving pattern mining)은 이동 객체의 위치 데이터를 가지고 이동 패턴을 발견하는 것으로 4단계로 구성된다[8]. 시간 패턴 마이닝은 이동 객체 데이터 베이스를 정렬하고 위치 정보를 일반화시키고 이를 일반화된 데이터베이스에 저장한다. 다음 단계인 이동 시퀀스 추출 단계에서 이 논문에서 제시한 세가지 임계치를 이용한다. 얻어진 이동 시퀀스를 이용해 빈발 패턴을 탐사하여 이동 패턴을 찾게 된다. 그림 2는 시간 패턴 마이닝의 프로시저 구조를 보여준다.

6. 결론 및 향후 연구

이 논문에서는 위치 기반 서비스에 적용 가능한 시간 패턴 마이닝을 언급했다. 이전의 연구에서는 이동 객체의 모호성을

고려하지 않았다. 이 논문에서는 시간 패턴 마이닝을 하는데 있어서 발생할 수 있는 모호성을 해결하기 위해 모호성을 언급하고 이를 해결하기 위해 임계치를 정의하고 이를 고려한 시간 패턴 마이닝 프로시저 구조를 제시하였다.

향후 연구로 여기서 제시한 처리 방법과 임계치를 고려하여 시간 패턴 마이닝을 확장하여 구현하는 것과 실험 평가하는 것이 요구된다.

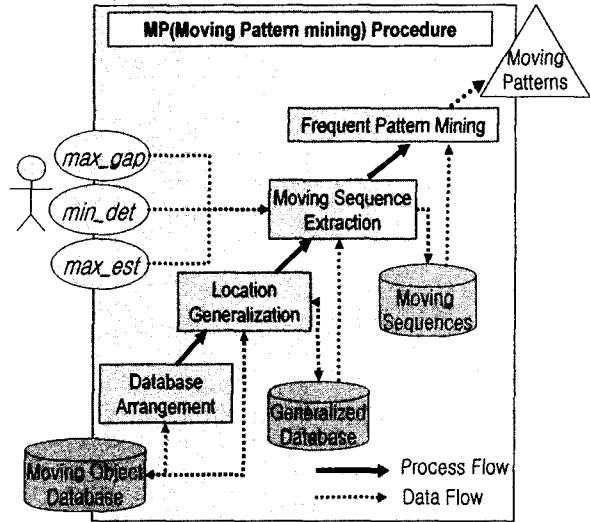


그림 2 시간 패턴 마이닝의 프로시저 구조

참고 문헌

- [1] 최미선, 김영국, 이동(Mobile) 데이터베이스 개요 및 연구 현황, 데이터베이스 연구, 17권 3호 2001.
- [2] 김옥, 지규인, 이장규, 위치 기반 무선 인터넷 서비스, Telecommunications Review, 제 10권 6호, 2001.
- [3] R.H.Guting, and at. AI, A Foundation for Representing and Querying Moving Objects, ACM Transactions on Database Systems, Vol. 25, No. 1, pp. 1-42, 2000.
- [4] M. Erwig, R.H. Guting, M. Schneider, and M. Vazirgiannis, Spatio-Temporal Data Types: An Approach to Modeling and Querying Moving Objects in Databases, GeoInformatica Vol.3, No. 3, pp.269-296, 1999
- [5] R. Agrawal and R. Srikant, Mining Sequential Patterns, 11th International Conference on Data Engineering, 1995.
- [6] B. Ozden,, S. Ramaswamy, and A. Silberschatz, Cyclic Association Rules, Proceedings of the 14th International Conference on Data Engineering, 1998.
- [7] L.Forlizzi, R.H.Guting, E.Nardelli, and M.Schneider, A Data Model and Data Structures for Moving Objects Databases, In Proc. Of the ACM SIGMOD Conf., 2000
- [8] 백옥현, 위치 기반 서비스를 위한 이동 객체의 시간 패턴 탐사 기법, 충북대학교 대학원 전자계산학과 석사학위 논문, 2002
- [9] D. Pfoser, and C. S. Jensen, Capturing the Uncertainty of Moving-Object Representations, In Proc. of Advances in Spatial Databases, 6th International Symposium, SSD'99, pp. 20-23, 1999.
- [10] 안윤애, 류근호, 이동 객체의 불확실한 위치 정보 관리, 충북대학교 컴퓨터정보통신 연구 제9권 제1호 pp.81-91, 2001년 5월.
- [11] 백옥현, 이준욱, 김영균, 류근호, 시간 연관규칙의 유지를 위한 점진적인 마이닝 기법, 한국정보과학회, 20권 1호 2001.