

# 위치 기반 서비스를 위한 이동 객체의 시간 패턴 탐사<sup>†</sup>

백옥현, 이준욱, 류근호  
충북대학교 데이터베이스 연구실  
e-mail : {ohpaek, junux, khryu}@dmlab.chungbuk.ac.kr

## Temporal Pattern Mining of Moving Objects for Location Based Services

Ok Hyun Paek, Joon Wook Lee, Keun Ho Ryu  
Database Laboratory, Chungbuk National University

### 요 약

위치 기반 서비스는 이동중인 사용자에게 사용자의 위치와 관련된 정보를 제공하는 서비스를 통칭한다. 이 논문에서는 위치 기반 서비스를 위해서 이동 객체의 시간 패턴을 탐사함으로써 사용자에게 최적화된 서비스를 제공할 수 있는 기법을 제안한다. 제안하는 기법은 이동 객체 중 이동 점의 과거 위치 정보에 대한 시간에 따른 변화 패턴을 구하는 기법이다. 이것은 시간에 따라 위치가 자주 바뀌는 이동 객체의 특성을 고려하여 효율적으로 패턴을 갱신하여 항상 최근의 정보를 유지할 수 있도록 하였다.

### 1. 서론

최근 무선 인터넷을 사용할 수 있는 이동 전화와 PDA의 소개와 정확한 위치정보기술의 발달로, 이를 이용한 응용 서비스에 대한 관심이 고조되고 있으며, 그 중 위치 기반 서비스(Location Based Services)가 무선 인터넷의 새로운 이슈로 떠오르고 있다[8].

위치 기반 서비스는 이동 기기를 사용하는 사용자의 위치를 추적해 다양한 응용 서비스를 제공하는 서비스를 통칭한다. 이것은 개인의 정확한 위치를 파악해서 각종 마케팅이나 서비스에 활용함으로써 개인화되고 동적인 서비스를 제공한다. 위치 기반 서비스의 대상인 PDA, 이동 전화, 차량, 비행기, 선박 등은 자유롭게 이동하면서 시간에 따라 위치를 변경하는 특성을 가진다. 이와 같이 시간의 흐름에 따라 객체가 이동하면서 위치 및 모양을 연속적으로 변경하는 특성을 가지는 데이터를 이동 객체(Moving Objects)라 한다[3]. 이동 객체의 시간의 따른 변화는 특정한 경향 혹은 패턴을 가진다. 이러한 경향이나

패턴과 같은 지식은 데이터 마이닝 기법을 통해 발견할 수 있다. 찾아낼 수 있는 지식의 예는 다음과 같다.

- ◆ 전체 고객의 25%는 A→B→C→D 순으로 지역을 방문한다
- ◆ A→B→C 지역으로 이동하는 자동차의 90%는 C 지역에서 130km/h 이상의 속력을 낸다
- ◆ 저녁 6시부터 8시 사이에 고객의 30%는 B→C→D 지역을 순서적으로 방문한다.
- ◆ Shopping Zone 안에서 50%의 고객들은 A→B→C 순으로 이동한다.
- ◆ 오후 1시부터 3시 사이에 25%의 고객이 쇼핑 지역에서 식당 지역으로 이동한다.

이런 지식들은 위치 기반 서비스에서 사용자의 이동 패턴과 프로필 등의 성향을 이용함으로써 사용자에게 더욱 특화된 서비스를 제공한다.

이 논문에서는 이동 객체의 시간에 따른 위치 변화에

<sup>†</sup> 이 논문은 한국통신가입자망연구소 위탁 연구로 수행되었음

서 패턴을 탐사함으로써 위치 기반 서비스에 적용할 수 있는 기법을 제안한다. 이동 객체는 시간에 따라 빈번하게 공간적인 속성이 변하기 때문에 패턴 역시 최근의 경향을 반영하기 위해 빈번하게 탐사되어야 한다. 그러나, 기존의 방법은 정적인 환경을 전제로 했기 때문에 이동 객체의 시간 패턴을 탐사하는 것에는 부적절하다. 따라서 이 논문에서 제안하는 방법은 패턴 탐사의 효율성을 보장하기 위해 점진적인 방법을 사용한다. 점진적인 방법을 사용함으로써 마이닝 연산의 효율성과 규칙의 질을 보장할 수 있다.

## 2. 관련연구

이 장에서는 관련연구로써 위치 기반 서비스와 이동 객체, 그리고 기존의 데이터 마이닝 기법에서 활용되었던 시간 패턴에 대해 기술한다.

### 2.1 위치 기반 서비스

위치 기반 서비스 기술은 위치 측정 기술(Location Determination Technology), 위치 응용 플랫폼, 그리고 응용 및 서비스로 구성된다. 전체적인 구성은 그림 1과 같다.

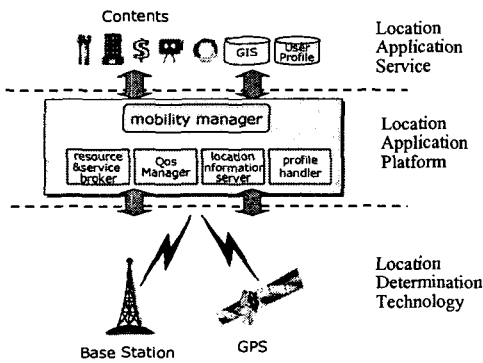


그림 1 위치 기반 서비스의 구성

위치 측정 기술은 GPS를 이용하거나, 기존 통신망의 기지국의 위치와 파일럿 신호로부터 이동중인 사용자의 위치를 알아내는 통신망 기반 방식이 있다. 위치 응용 플랫폼은 위치 기반 응용 서비스를 제공하기 위한 기반이 된다. 위치 기반 응용 서비스는 위치 측정 기술과 플랫폼을 이용하여 위치 기반의 응용 서비스를 수행할 수 있는 기술로써, 표 1과 같이 분류될 수 있다[8].

표 1 위치 기반 응용 서비스의 분류

| 분류          | 서비스  |
|-------------|--|
| 위치 기반 정보서비스 | 지도서비스, 지도 전화번호부, 인접 지역정보, 교통정보 가이드, 여행 가이드       |
| 항법/추적 서비스   | 최적경로 계산, 차량항법 및 물류관제, 영업 배치 등 기업용 인터넷, 개인 항법     |
| 위치 기반 상거래   | Mobile 전자상거래, 위치기반 비교쇼핑, 예약, 위치 기반 광고, 위치 기반 CRM |
| 공익 서비스      | 구조 요청, 범죄 신고, 자동차 사고에의 대응, 기상예보                  |
| 기타          | 게임, 위치기반의 이동통신 요금 차등화                            |

### 2.2 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]. 이동 점은 시간에 따라 객체의 위치가 변하는 것으로 이런 이동 점의 예는 택시, 배, 비행기, 사람 등이 있다. 이동 점에 대한 질의는 “화재가 난 곳과 가장 가까운 소방차는 어디에 위치하는가?”, “437번 비행기의 이동 경로를 보이시오”와 같이 예를 들 수 있다. 이동 영역은 시간에 따라 객체의 위치 뿐만 아니라 모양까지 변하는 것으로, 이러한 이동 영역의 예는 민족의 이동, 숲의 발달, 폭풍의 이동 경로 등이 있다. 이동 영역에 대한 질의는 “2001년 5월 10일 오전 10시에 폭풍의 위치와 모양을 보이시오”와 같은 예를 들 수 있다. 이 논문에서는 이동 객체의 시간 패턴 마이닝에 대해서 이동 점만을 대상으로 한다.

### 2.3 시간 패턴

시간 패턴은 시간 속성을 가진 데이터에서 찾아 낼 수 있는 시간적인 지식의 하나로써, 환자의 병력, 상품 구매 이력, 웹 로그 등과 같은 데이터에 적용되어왔다. 시간 패턴은 그 지식의 형태에 따라 순차 패턴, 주기 패턴, 시간 관계로 분류될 수 있다.

순차패턴 탐사는 항목집합으로 구성된 트랜잭션들 간에 특정 항목집합이 순차적으로 발생하는 패턴을 탐사하는 기법이다[1]. 순차패턴의 예로, 위치 기반 서비스를 사용하는 사용자들은 특정 시간 간격동안 A → B → C 지역을 순차적으로 방문하는 경향이 있다는 것을 들 수 있다. Srikant 은 [7]에서 일반화한 순차패턴을 탐사하기 위해 GSP 알고리즘을 제안하였다. 주기 패턴이란 시간 상에서 정기적으로 발생하는 현상으로, 예를 들어 매년 또는 매주, 매일과 같은 시기에 주기적으로 발생하는 패턴을 가리킨다. [6]에서는 주기(Cyclic) 연관규칙을 소개하였다. 시간 관계(Temporal relationship)는 시간 간격을 가진 여러 사건 간의 시간 관계를 포함한 지식이다. 순차패턴과는 달리, 단순히 사건이 종료된 후에 다음 사건이 발생하는 것이 아니라, 여러 사건이 특정 기간 동안 중복되어 발생한다는 특성을 가진다.

## 3. 데이터 모델

이동객체의 위치는 시간에 따라 변하므로 기존의 데이터 모델로는 이를 표현할 수 없었으며, 따라서 이를 위한 방법들이 연구되어 왔다[2, 3].

이 논문에서 고려한 데이터 모델은 시공간 이동 객체를 유한한 점들의 집합으로 표현한 이산적 모델로써, 그림 2와 같이 나타낼 수 있다. 그림 2(a)의 굵은 선은 점 객체의 이동을 나타낸 것이며, 2(b)는 시공간 지원 공간에서 몇 개의 객체 이동을 보여준다.

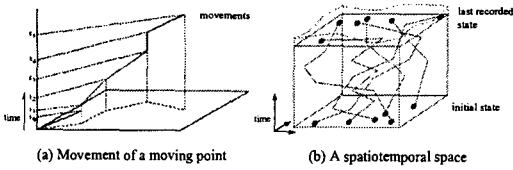


그림 2 이동과 공간

이동 점 객체는 해당 객체에 대한 버전들의 집합으로 저장된다. 이동 점 객체를 표현하기 위해 제시된 여러 가지 데이터 모델이 있지만[2,3], 이 논문에서는 일반성을 갖기 위해 기본적인 부분만을 고려한다. 이동 점 객체는 이동 점의 공간 속성의 유효시간의 집합으로 표현된다. 위치 기반 서비스에서의 공간 속성은 어떤 위치 측정 기술을 사용하느냐에 따라서  $x, y$  좌표를 가지는 좌표점으로 표현할 수도 있고, 기지국에 포함되는 영역만으로도 만족하는 경우도 있다. 이동 점에 대한 시간 패턴 탐사를 단순화하기 위해서 공간 속성은 이동 객체가 속해 있는 지역으로 제한한다. 이동 점 객체는 다음과 같이 정의된다.

[정의 3.1]  $Mpoint = oid, \{(Lid_j, VT_j), \dots, (Lid_n, VT_n)\}$

단,  $oid$ 는 객체의 식별자를 가리키며,  $Lid_j$ 는 이동 객체가 속해 있는 지역의 식별자이다.  $VT_j$ 는 유효 시간이고,  $(From, To)$ 의 두 타임스탬프로 구성되며,  $\forall i, j \in n$  일 때,  $VT_i \cap VT_j = \emptyset$  이다.

4. 문제정의

4.1 이동 객체의 시간 패턴 마이닝

데이터베이스  $D$ 를 이동 점 객체의 집합이라 하자. 하나의 이동 점 객체는 객체 식별자( $oid$ )와 이동 점이 존재했던 공간 속성인 지역과 유효시간으로 구성된다.

$L = \{l_1, l_2, \dots, l_m\}$ 은 공간 속성을 나타내는 지역의 집합이다.

시퀀스는 지역  $l$ 의 순서화된 리스트이다. 시퀀스  $s$ 는  $\langle s_1, s_2, \dots, s_n \rangle$ 으로 표기된다.  $k$ -시퀀스는  $k$ 개의 지역으로 구성된다.

시퀀스  $\langle a_1, a_2, \dots, a_n \rangle$ 은 만약  $a_1 \subseteq b_1, a_2 \subseteq b_2, \dots, a_n \subseteq b_n$ 인 정수  $i_1 < i_2 < \dots < i_n$ 이 존재한다면 다른 시퀀스  $\langle b_1, b_2, \dots, b_n \rangle$ 의 부분 시퀀스이다. 예를 들어  $\langle 2, 5, 9, 8 \rangle$ 은  $\langle 6, 2, 1, 5, 9, 3, 8 \rangle$ 의 부분 시퀀스이다.

데이터베이스에 있는 한 객체에 대해 시간의 증가에 따라서 지역을 순서적으로 정렬한 것을 이동 시퀀스(Moving Sequence)라 부른다. 시퀀스  $s$ 의 지지도는  $s$ 를 포함하는 전체 이동 시퀀스의 비율로 정의된다. 이동 시퀀스는  $s$ 가 이동 시퀀스의 부분 시퀀스이면 시퀀스  $s$ 를 포함한다. 하나의 시퀀스는 그것의 지지도가 사용자가 지정 한 최소 지지도  $min\_sup$ 을 초과하면 빈발 시퀀스이다.

이동 점 객체의 데이터베이스가 주어졌을 때, 시간 패턴 마이닝은 모든 빈발 시퀀스를 찾는 것이다.

4.2 시간 패턴의 점진적 갱신

$|D|$ 는 원래의 데이터베이스  $D$ 에 있는 이동 시퀀스의

수이고,  $min\_sup$ 이 최소 지지도를 나타낸다고 가정하자. 시간의 흐름에 따라 이동 객체의 위치가 바뀌었고, 이동 객체의 새로운 위치가 데이터베이스  $D$ 에 추가되었다. 이런 새로운 정보는 추가된 데이터베이스  $\Delta D$ 에 저장될 수 있다.  $UD$ 는  $D$ 와  $\Delta D$ 에 포함되는 갱신된 데이터베이스이다. 즉,  $UD = D \cup \Delta D$ 가 된다.  $S_k^D$ 가  $D$ 에 있는 모든 빈발  $k$ -시퀀스의 집합이고,  $S_k^{\Delta D}$ 는  $\Delta D$ 에 있는 모든 빈발  $k$ -시퀀스의 집합, 그리고  $S^D$ 와  $S^{\Delta D}$ 는 각  $D$ 와  $\Delta D$ 의 모든 시간 패턴을 나타낸다.

$D$ 의 각 시간 패턴  $s$ 에 대해서 그것의 지지도 값을  $S_{count}^D$ 라고 하자. 이전과 같은  $min\_sup$  값에 대해서 시퀀스  $s$ 는 그것의 지지도 값이  $min\_sup \times |UD|$ ( $UD$ 에 있는 전체 이동 시퀀스)보다 크면 빈발 시퀀스이다.

시간 패턴을 점진적으로 갱신하는 것은  $UD$ 에 있는 새로운 빈발 시퀀스  $S^{UD}$ 를 찾는 것이다.  $UD$ 의 빈발 시퀀스  $S^{UD}$ 를 찾기 위한 이전의 접근 방법들은 모든 시퀀스에 대해 탐색을 수행해야만 했다. 그러나 제안하는 방법은 이전에 탐사되었던 지식과 추가된 데이터베이스에 대해서만 고려함으로써 탐색 공간을 줄여서 프로세스를 효율적으로 수행할 수 있도록 하였다. 이런 접근 방법에 대한 개념적인 비교는 그림 3에서 볼 수 있다.

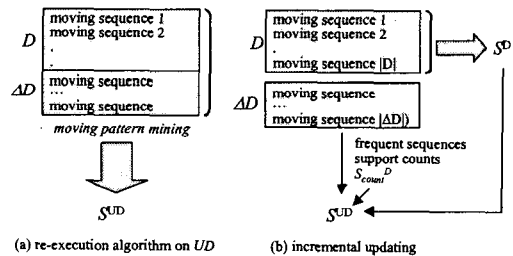


그림 3 재탐사와 점진적 갱신

예 1)  $Oid$ 로 정렬된 이동 객체의 이동 시퀀스가 표 2와 같이 있다고 가정하자. 여기서 최소 지지도는 33%이다. 이것은 최소 이동 시퀀스의 개수가 2개가 되어야 한다는 것을 의미한다. 여기서의 패턴은  $\langle 1 \rangle, \langle 2 \rangle, \langle 3 \rangle, \langle 4 \rangle, \langle 5 \rangle, \langle 8 \rangle, \langle 1, 2 \rangle, \langle 8, 2 \rangle, \langle 8, 1 \rangle, \langle 2, 1 \rangle, \langle 2, 8 \rangle, \langle 5, 1 \rangle, \langle 5, 3 \rangle, \langle 3, 1 \rangle$ 이 된다.

표 2 6개의 이동 시퀀스를 가지는 원래 데이터베이스  $D$

| oid | Moving Sequence                    |
|-----|------------------------------------|
| 1   | $\langle 8, 2, 1 \rangle$          |
| 2   | $\langle 1, 2, 8, 7 \rangle$       |
| 3   | $\langle 1, 4 \rangle$             |
| 4   | $\langle 1, 4 \rangle$             |
| 5   | $\langle 5, 3, 1 \rangle$          |
| 6   | $\langle 2, 5, 8, 3, 1, 2 \rangle$ |

시간이 흐름에 따라 이동 객체의 위치가 변하게 되었고 데이터베이스에는 새로운 데이터가 더해졌다고 가정하자.  $oid$  순으로 정렬된 추가된 데이터는 표 3과 같다. 이제 전체에서 빈발 패턴이 되려면 최소 3번 이상 나타나야 한다. 이전의 패턴이었던  $\langle 3 \rangle, \langle 8, 2 \rangle, \langle 2, 1 \rangle, \langle 2, 8 \rangle, \langle 5, 1 \rangle, \langle 5, 3 \rangle$  그리고  $\langle 3, 1 \rangle$ 은 더 이상 빈발패턴이 아니다. 대신,

<7>과 <8 7>은 새로운 패턴이 되었다.

표 3 추가된 이동 시퀀스 데이터베이스  $\Delta D$

| oid | Moving Sequence |
|-----|-----------------|
| 7   | <5 7 8 7>       |
| 8   | <8 1 7>         |
| 9   | <1 2 4>         |

5. 점진적인 기법을 사용한 이동 객체의 시간 패턴

마이닝

이 절에서는 앞에서 설명되었던 것을 기반으로 이동 객체의 시간 패턴을 마이닝 할 수 있는 점진적인 알고리즘을 제안한다. 제안하는 알고리즘은 Agrawal 에 의해 제시되었던 Apriori 알고리즘[1]을 기반으로 한다. 알고리즘은 크게 세 단계로 나누어진다.

단계 1. 증가된 데이터베이스  $\Delta D$  를 스캔해서 존재하는 1-시퀀스의 발생수를 센다.

단계 2.  $\Delta D$  의 후보 1-항목집합에서  $min\_sup$  를 만족하는 주요 1-시퀀스를 생성하고,  $\Delta D$  의 후보 1-시퀀스와  $D$  의 주요 1-시퀀스를 비교해서 전체  $min\_sup$  를 만족하는 전체 주요 1-시퀀스 생성한다.

단계 3. 각  $\Delta D$  의 주요  $(k-1)$ -항목집합과  $UD$  의 주요  $(k-1)$ -항목집합은  $\Delta D$  와  $UD$  각각의 후보  $k$ -시퀀스 만들기 위해 사용된다. 후보집합 생성은 [1]에서 제시한 apriori-gen() 함수가 사용된다.  $\Delta D$  가 스캔되면서 각  $k$ -시퀀스의 지지도가 세어지고, 각각의 후보집합에서  $min\_sup$  를 만족하는 시퀀스가 선택된다. 이 과정은 새로운 빈발한 시퀀스가 찾아지지 않을 때까지 반복된다.

다음은 이동 객체에서의 시간 패턴을 탐사하기 위한 점진적인 알고리즘의 단계별 알고리즘을 보인다.

```

1)  $C_1^{\Delta D} = \{1\text{-후보항목집합}\} / * \Delta D$  의 1-후보항목집합*/
2) foreach sequence  $S \in \Delta D$  do
3)   count support
4)  $L_1^{\Delta D} = \{c \in C_1^{\Delta D} \mid c.freq_{\Delta D} / |\Delta D| \geq min\_sup\}$ 
5) foreach  $X \in L_1^{\Delta D}$  do
6)   if  $X.freq_D + X.freq_{\Delta D} / (|D| + |\Delta D|) \geq$ 
        $min\_sup$  then
7)      $L_1^{UD} = L_1^{UD} \cup \{X\}$ ;
8) for  $(k=2; L_{k-1}^{UD} \neq \emptyset$  and  $L_{k-1}^{\Delta D} \neq \emptyset; k++)$  do begin
9)    $C_k^{\Delta D} = apriori-gen(L_{k-1}^{\Delta D})$ ;
10)   $C_k^{UD} = apriori-gen(L_{k-1}^{UD})$ ;
11)  foreach sequence  $S \in \Delta D$  do begin
12)    foreach candidate  $c \in C_k^{\Delta D}$  do
13)      count support
14)    foreach candidate  $c \in C_k^{UD}$  do
15)      count support
16)  end
17)  $L_k^{\Delta D} = \{c \in C_k^{\Delta D} \mid c.sup_{\Delta D} / |\Delta D| \geq min\_sup\}$ 
18)  $L_k^{UD} = \{c \in C_k^{UD} \mid c.sup_D + c.sup_{\Delta D} /$ 
        $|D| + |\Delta D| \geq min\_sup\}$ 
19) end
    
```

그림 4 시간 패턴 탐사를 위한 점진적 알고리즘

6. 결론 및 향후 연구

이 논문에서는 위치 기반 서비스에 적용하기 위하여 하나의 이동 기기를 사용하는 사용자를 이동 객체로 설정하고, 이동 객체의 시간에 따른 공간 속성의 변화에 대한 패턴을 마이닝하는 기법을 제안하였다. 시간에 따른 위치가 자주 변하는 이동 객체의 특성을 고려하여 효율적으로 패턴을 탐사할 수 있도록 필요한 부분에 대해서만 데이터베이스에 대한 스캔을 수행하는 점진적인 방법을 사용하였다. 점진적인 방법은 마이닝 연산을 효율적으로 수행할 수 있게 하고, 적은 비용으로 데이터의 최근 경향을 반영할 수 있게 한다는 장점이 있다.

향후 제시한 기법의 구현과 시간, 공간적인 특성을 가지는 데이터에 각각의 차원에 대해 제약조건과 개념 계층을 적용하기 위한 연구가 필요하다.

참고문헌

- [1] R. Agrawal and R. Srikant, Mining Sequential Patters, 11<sup>th</sup> International Conference on Data Engineering, 1995.
- [2] 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.
- [3] R.H.Guting, and at. Al, A Foundation for Representing and Querying Moving Objects, ACM Transactions on Database Systems, 2000.
- [4] J. Han, G. Dong, and Y. Yin, Efficient Mining of Partial Periodic Patterns in Time Series Database, 11<sup>th</sup> International Conference on Data Engineering, 1999.
- [5] P.S. Kam and A. Fu, Discovering Temporal Patterns for Interval-Based Events", the 2nd Int. Conf. on Data Warehousing and Knowledge Discovery, (DaWak), Springer Verlag, LNCS, London, UK, 4-6 Sept, 2000.
- [6] B. Ozden,, S. Ramaswamy, and A. Silberschatz, Cyclic Association Rules, Proceedings of the 14<sup>th</sup> International Conference on Data Engineering, 1998.
- [7] R. Srikant and R. Agrawal, Mining Sequential Patterns : Generalizations and Performance Improvements, International Conference on Extending Database Technology, Springer-Verlag, 1996.
- [8] 김옥, 지규인, 이장규, 위치 기반 무선 인터넷 서비스, Telecommunications Review, 제 10 권 6 호, 2001.
- [9] 류근호, 이준욱, 이용준, eCRM 을 위한 시간 데이터 마이닝 기술, 정보과학회 데이터베이스연구회지, 제 17 권 제 1 호, 2001.
- [10] 안윤애, 류근호, 이동 객체의 불확실한 위치 정보 관리, 충북대학교 컴퓨터정보통신 연구 제 9 권 제 1 호 pp.81-91, 2001년 5월.