

이동객체 위치 일반화를 이용한 시공간 이동 패턴 탐사

이 준 옥[†] · 남 광 우^{††}

요 약

현재의 이동객체를 기반으로 하는 다양한 시공간 응용환경에서의 서비스 지원 시스템 개발을 위하여 중요한 문제 중의 하나는 방대한 이동객체의 위치 이동 데이터로부터의 의미 있는 지식인 시공간 이동 패턴을 탐사하는 것이다. 이를 위하여 시간적 위상관계, 공간적 위상관계 그리고 시공간적 위상관계에 대한 접근이 지식 탐사를 위하여 고려되어야 한다. 이 논문에서는 효율적인 시공간 이동 패턴 탐사 기법인 MPMine 알고리즘을 제안하였다. 제안한 기법은 시간 제약조건과 공간 제약조건 등을 함께 고려하며 또한 공간 위상 연산인 contain()을 이용한 공간 개념화를 수행할 수 있다. 제안한 기법은 기존의 일반적인 시간 패턴 탐사 기법과 달리 이동객체 데이터 집합으로부터 위치 요약 및 일반화를 통하여 탐색 공간을 줄일 수 있어 효율적으로 유용한 이동 패턴을 탐사할 수 있다.

Spatiotemporal Moving Pattern Discovery using Location Generalization of Moving Objects

Jun Wook Lee[†] · Kwang Woo Nam^{††}

ABSTRACT

Currently, one of the most critical issues in developing the service support system for various spatio-temporal applications is the discovering of meaningful knowledge from the large volume of moving object data. This sort of knowledge refers to the spatiotemporal moving pattern. To discover such knowledge, various relationships between moving objects such as temporal, spatial and spatiotemporal topological relationships needs to be considered in knowledge discovery. In this paper, we proposed an efficient method, MPMine, for discovering spatiotemporal moving patterns. The method not only has considered both temporal constraint and spatial constrain but also performs the spatial generalization using a spatial topological operation, contain(). Different from the previous temporal pattern methods, the proposed method is able to save the search space by using the location summarization and generalization of the moving object data. Therefore, Efficient discovering of the useful moving patterns is possible.

키워드 : 시공간 이동 패턴(Spatiotemporal Moving Pattern), 패턴 탐사(Pattern Discovery), 시공간 지식 탐사(Spatiotemporal Knowledge Discovery), 위치 일반화(Location Generalization)

1. 서 론

최근 들어 휴대폰, PDA, 자동차 등과 같이 시간에 따라 공간 특성이 빈번하게 변화하는 방대한 시공간 이동 데이터를 관리하기 위한 데이터베이스 기술과 위치 측위 기술이 발전함에 따라 위치기반 서비스[21], 위치기반 추천 시스템[23], 지능적 물류 관제 서비스[22], 텔레매틱스 등의 새로운 시공간 응용 서비스개발에 대한 관심과 연구가 증대되고 있다. 이러한 실세계의 시공간 응용 분야에서 이동데이터의 이력을 기반으로 하는 지능적 서비스를 위해서는 시공간 이동 데이터 집합으로부터 의미 있는 시공간 지식 탐사를 필요로 한다.

시공간 지식 탐사(spatiotemporal knowledge discovery) [1-4]는 시공간 데이터 집합으로부터 변화 이력에 대한 시간(temporal) 규칙이나 공간(spatial) 규칙뿐만 아니라 시공간(spatiotemporal) 규칙 등과 같이 유용한 지식 탐사를 위한 연구 분야이다. 시공간 데이터 집합으로부터의 시간 규칙이나 공간 규칙의 탐사는 기존의 시간 데이터 마이닝[5, 6] 기법이나 공간 데이터 마이닝 기법[7, 8] 등을 통하여 탐사될 수 있다. 그러나 시공간 규칙은 객체의 시간 요소와 공간 요소를 함께 고려함으로써 생성될 수 있다[1, 9]. 시공간 지식 탐사의 주요 태스크는 시공간 객체의 특성을 찾아내는 것이다. 따라서 시공간 관계는 시공간 객체와 관련된 사건들 간의 인과관계(casual relationship)를 탐사하는데 매우 중요한 의미를 갖는다. 이를 위하여 시간적 위상관계, 공간적 위상관계 그리고 시공간적 위상관계에 대한 접근이 지식 탐사를 위하여 고려되어야 한다.

[†] 정 회 원 : 한국전자통신연구원 LBS 연구팀 PostDoc. 연구원

^{††} 정 회 원 : 한국전자통신연구원 LBS 연구팀 선임연구원
논문접수 : 2003년 7월 24일, 심사완료 : 2003년 9월 8일

시간과 공간 의미를 함께 고려함으로써 획득할 수 있는 시공간 규칙 유형 중 시공간 이동 패턴의 예로는 “매 주중 저녁에 20~35세까지의 미혼 여성들이 지역을 A-B-C의 순으로 빈번하게 방문하는 경향이 있다”와 같은 일반화된 시공간 이동 패턴, “9월에서 11월 동안에 A 지역에서 B 지역으로의 주변 500M 범위를 지나는 차량의 수가 증대되었다.”와 “매 주말 저녁에 지역 A-B-C를 방문하는 경향이 높은 여성들의 소비가 지난 3개월 사이에 감소하였다.”와 같은 시공간 메타규칙 등이 있다.

시공간 이동 패턴은 이동하는 객체의 위치 패턴으로 객체의 위치 특성에 따른 개인화되고 알맞은 콘텐츠나 서비스 제공을 가능하게 하는 시공간 규칙이다[24]. 지금까지는 사건(event) 시퀀스에서 빈발하게 발생하는 패턴 규칙을 탐사하는 시간 패턴 탐사 연구[10-13]가 진행되었다. 그러나 이러한 패턴 탐사 기법은 공간 데이터의 특성 및 공간 연산을 패턴 탐사에 이용하지 않기 때문에 공간 일반화 연산[14, 15]을 고려해야 하는 시공간 이동 패턴 탐사 문제에 적용하기 어렵다.

따라서 이 논문에서는 이러한 문제점을 해결하기 위하여 패턴 탐사를 위한 시공간 요소들을 식별하고 그 관계를 표현하며 다양한 시간, 공간연산 및 위상관계를 이용하는 새로운 지식 탐사 기법인 시공간 이동 패턴 탐사 기법, MP Mine을 제안한다.

이 논문의 구성은 다음과 같다. 제 2장에서는 시공간 이동 패턴 탐사와 관련된 기존의 연구를 정리하고 그 문제점을 제시한다. 제 3장에서는 시공간 이동 패턴 탐사를 위한 문제에 대한 정의를 제시한다. 제 4장에서는 시공간 이동 패턴 탐사 기법을 설계하고 알고리즘을 제시한다. 제 5장에서는 제안한 시공간 이동 패턴 탐사 기법의 성능 특성을 다양한 평가 기준에 의해 분석한다. 마지막으로 제 6장은 결론으로 제안한 시공간 이동 패턴 탐사 기법에 대한 문제점과 향후 연구 방향을 제시한다.

2. 관련 연구

기존의 패턴 탐사기법인 Apriori 기반의 패턴 탐사[10], SPIRIT[17] 및 GSP[13] 등은 주로 사건 이벤트 시퀀스 내에서의 빈발 패턴이나 에피소드 탐사를 주요 탐사 대상으로 하고 있다. 이러한 탐사 기법은 이동객체와 같이 시간 및 공간 차원을 포함하는 객체의 다양한 이동 패턴을 탐사하는데 있어 복잡한 과정을 통하여 비효율적으로 패턴 탐사를 제공할 수 있다. 이동객체 데이터의 급격한 증가와 탐사 빈도를 고려할 때 이동 패턴 탐사의 효율성은 응용환경에 있어서 매우 중요한 문제이다. 특히 시공간 위상 관계성을 패턴 탐사에 접목시키기 위해서는 패턴 탐사 기법의 데이터 모델과 위상 관계 처리 등에 관한 부분 등이 함께 고

려되어야 한다. 아직까지 Ewrig[18] 등이 제시한 시공간 위상관계 연산을 이동 패턴 탐사에 고려하지 않고 있기 때문에 시공간 데이터 집합에 잠재하고 있는 객체들 간의 의미 있는 이동 패턴은 탐사되지 못하고 있다.

앞에서 고찰한 연구와 같이 최근까지의 시공간 패턴 등은 서론에서 제시한 것과 같은 유형의 시공간 이동 패턴 지식을 제공하지 못한다. 시간 패턴 탐사[10-13]나 몇몇 시공간 패턴[3, 4, 16]과 달리 시간과 공간 특성을 함께 고려하는 시공간 이동 패턴 탐사를 위하여 해결해야 할 문제들은 다음과 같다.

시공간 이동객체 데이터에 대한 형식적인 정의가 필요하다. 기존의 패턴 탐사 기법들은 이동 패턴 문제를 위한 시공간 이동객체를 정의하고 있지 않고 비공간 속성만을 갖는 데이터 집합에 대한 정의나 이미지 데이터 집합에 대한 정의만을 제시하고 있다.

효율적인 시공간 이동 패턴 탐사를 위한 이동 시퀀스 생성이 가능하여야 한다. 시공간 이동 데이터 집합은 빈번한 이동 데이터를 포함하고 있어 방대한 이동 시퀀스 집합을 생성한다. 각 객체 단위의 긴 이동 시퀀스는 중요하지 않은 이동 정보를 포함할 수 있기 때문에 효율적 패턴 탐사를 위해서는 사용자가 원하는 수준의 요약이 가능해야 한다.

공간 개념 지식의 적용을 통한 일반화된 개념 수준으로의 이동 시퀀스를 생성해야 한다. 이동 시퀀스는 이동객체의 시간에 따른 연속적인 이동 변화로써 위치수준의 이동 변화로부터는 의미 있는 패턴 탐사가 불가능하다. 상위의 공간 개념으로의 일반화는 의미 있는 패턴 탐사를 가능하게 하는 이동 시퀀스를 생성할 수 있다.

시공간 이동 패턴의 유용성 측정의 방법을 제시하여야 한다. 일반적으로 패턴 탐사의 유용성 측정 지표로 많이 사용되는 지지도(support threshold)는 이동 시퀀스간의 포함 관계에 있어서 시간 제약조건을 고려가 필요하다. GSP[13] 등은 시퀀스간의 포함 관계 정의에 있어 일반 항목간의 시간 제약조건을 제시하고 있다. 시공간 이동 시퀀스 내에서의 이동간 시간 제약 조건에 대한 명확한 정의를 통해 지지도를 통한 유용한 패턴 탐사가 가능해야 한다.

3. 시공간 이동 패턴 탐사 문제정의

3.1 시공간 이동객체에 대한 표현

시공간 이동객체를 표현하기 위한 모델은 매우 다양하다. 이 논문에서는 시공간 이동 패턴 탐사의 대상 데이터 집합을 정의하기 위하여 시공간 객체와 객체 영역 그리고 범위를 정의한다.

[정의 3.1] 시공간 객체 O 는 시간(time), 공간(space), 속성(attribute)의 3가지 차원(dimension)을 가지며 유

일한 객체 식별을 위한 식별자를 포함하여 $O = \langle OID, S, T, A \rangle$ 로 표현된다. 여기서 OID 는 식별자, S 는 공간, T 는 시간, A 는 속성 차원이다.

객체를 구성하는 3가지 차원은 값의 영역인 도메인(D)을 갖는다. 도메인 공간은 연속적이거나 이산적일 수 있다. 일반적인 지리공간 영역은 3차원을 말한다. 시간 차원 도메인에 있어 Snodgras[19]는 실세계에 객체가 존재하는 시간을 의미하는 유효시간과 시스템에 기록된 시간을 의미하는 거래시간의 두 개념을 규정하였다.

[정의 3.2] 시공간 객체 영역 U 는 시공간 객체가 존재하는 도메인으로 3차원 도메인의 카테이션 곱($U = S \times T \times A$)으로 표현된다.

각 시공간 객체는 시공간 객체 영역 속에서 식별자(OI ; object identifier)를 통해 서로를 유일하게 식별할 수 있다. 시공간 객체 식별자 OID 와 시공간 객체의 시간차원 T 에서의 한 시간값에 ($T_p \in T$)에 의해 시공간 객체 영역에서의 모든 객체는 유일하게 구별된다.

각 시공간 객체 인스턴스는 각 차원의 도메인 내에서의 범위(extent)을 갖는다. 따라서 각 도메인 내에서 한계를 다음과 같은 기호로 표시한다.

- 공간 범위 $E_S \in S$: 공간상에서의 한 시공간 객체의 위치와 모양 등 공간 객체의 외양(appearance)를 말한다. 예를 들면 2차원 상에서의 위치, 병원이나 학교 등의 위치, 국가 경계나 산림의 범위 등이 있다.
- 시간 범위 $E_T \in T$: 시간라인 상에서의 한 시공간 객체의 위치(point), 간격(interval) 또는 기간(period)를 말한다. 예를 들면 차량의 이동 시점이나 태풍의 이동 간격 등이 있다.
- 속성 범위 $E_A \in A$: 한 시공간 객체의 모든 속성의 값을 말한다. 예를 들면 차량의 이동 속도나 방향 등이 있다.

[정의 3.3] 시공간 객체 인스턴스 O_i 는 시공간 객체 영역 U 에서의 한 범위로 3차원에서의 범위 쌍으로 구성된다.

$$O_i = \langle OID_i, E_S, E_T, E_A \rangle$$

[20, 25]에서는 이동객체 시간 패턴 탐사를 위하여 이객체에 대한 표현 모델로써 이동객체 데이터는 유일하게 식별되는 객체 식별자(OID)와 함께 이산적으로 샘플링된 연속적인 이동 위치데이터로 구성된다. 이동 위치데이터는 2차원 공간상의 좌표 값과 측정시간이라는 유효시간으로 구성되는데 이러한 이동객체의 개념적 표현으로 $Mpoint = \langle OID, \{(VT_1, L_1), \{(VT_2, L_2), \dots, \{(VT_n, L_n)\} \rangle$ 로 정의하였다. 여기서 (VT_i, L_i) 는 이동객체의 i 번째 위치데이터로써 VT_i

는 유효시간 L_i 는 샘플링된 위치 (x_i, y_i) 로 2차원 상의 좌표이다. 이 논문에서는 2차원 공간상의 시공간 이동 데이터 집합을 다음과 같이 정의한다.

[정의 3.4] 2차원 공간상에서 어떤 시점 T_P 에서의 이동객체에 MO 의 인스턴스 MO_i 는 시공간 객체 영역 U 상 있으며 다음과 같이 정의한다. 이때 2차원 공간범위 $E_S = (x_i, y_i)$ 이고, $E_T = T_{Pi}$ 로 T_{Pi} 는 시공간 객체 영역상 시간 라인상의 한 인스턴스이다.

$$MO_i = \langle OID_i, (x_i, y_i), T_{Pi}, E_A \rangle$$

[정의 3.5] 어떤 객체 식별자 OID_i 에 대한 이동객체 인스턴스들은 이동객체 버전 집합 $MV(OID_i)$ 을 구성한다.

$$MV(OID_i) = \{MO_i | MO_i \notin U, MO_i[OID_i] = OID_i\} \subset U$$

[정의 3.6] 이동객체에 대한 이력 객체 MH_i 는 어떤 이동객체 버전 집합으로부터 객체 식별자를 통해 유도된 객체이다.

$$MH_i = \langle OID_i, \{(x_{iu}, y_{iu}), T_{Piu}, E_{Ai}\}, \dots, \langle (x_{in}, y_{in}), T_{Pin}, E_{An} \rangle \rangle$$

이러한 이력객체 집합에 대한 정리는 앞서 정의되었던 M Point의 표현에 하지만 시공간 이동객체와 이동객체 집합을 좀더 명확하게 표현하였다. 마지막으로 앞서 [정의 3.4], [정의 3.5], [정의 3.6]을 통해 이력객체들의 집합으로써 이동객체 데이터 집합을 정의하면 다음과 같다.

[정의 3.7] 이동객체 데이터 집합 MD 는 이동객체의 이력객체들의 집합이다.

$$MD = \{MH_i | MH_i[OID] \neq MH_j[OID], \text{ for } \forall i, j \ i \neq j\}$$

임의의 두 이력 객체 MH_i, MH_j 는 시공간 데이터 객체 영역상의 구분되는 식별자를 갖는 두 이동객체의 이력객체이다. 따라서 $MH_i[OID] \neq MH_j[OID]$ 이다. 이러한 시공간 데이터 영역상의 이동객체에 대한 이력 객체들은 이동객체 데이터 집합 MD 을 표현한다.

3.2 시공간 이동 패턴 탐사 문제 정의

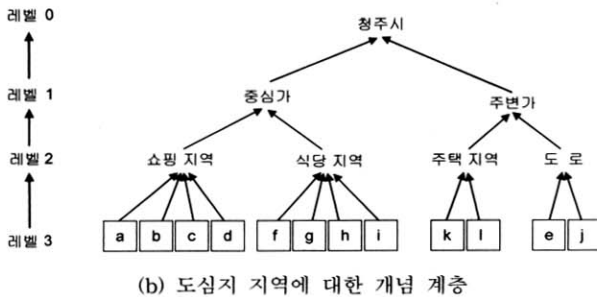
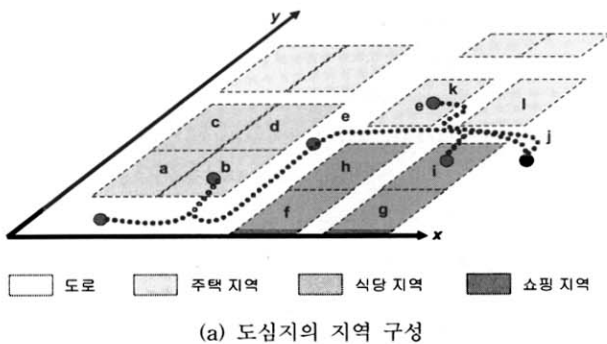
3.2.1 이력객체의 위치 일반화

위치 일반화는 어떤 위치를 공간상의 상위 개념으로 추상화 하는 것을 말한다. 개인의 상세한 수준의 위치 정보에서는 의미 있는 패턴을 찾기가 힘들다. 따라서 대상 데이터 집합으로부터의 시공간 이동 패턴 탐사를 위한 데이터는 전처리(preprocessing)를 통하여 이동에 일반화된 데이터 집합을 형성함으로써 의미있는 수준의 지식을 획득할 수 있다.

[정의 3.8] 이동 가능한 지역(region)들의 집합 R 은 $R = \{R_i\}$ 로 각 지역 $R_i = \langle R_{id}, E_S, E_A \rangle$ 로 표현한다. R_{id} 는 구분되는 지역에 대한 식별자이다.

이동 위치는 공간상의 특정 지역 수준의 의미로 일반화되기 위하여 공간 지역에 대한 공간 개념 계층이 요구된다. 공간 개념 계층에 대하여 공간 연산을 이용하여 각 이동 위치는 [정의 3.8]에서 정의되고 있는 개념 수준의 지역으로써 일반화 된다.

(그림 3.1)에서는 지역에 대한 4 레벨의 공간 개념 계층 예를 보여준다. 단말 수준에서는 건물이나 도로 등을 표현하고 있고 이들은 공간 범위로서 영역 타입에 해당한다. 또한 상위로 포함관계를 통하여 지역들로 일반화된다. 레벨 3의 단말 수준의 지역이나 도로 등은 레벨 2의 쇼핑지역이나 식당지역, 주택지역, 거리 등으로 일반화되고 있음을 알 수 있다.



(그림 3.1) 지역에 대한 공간 개념 계층

위치의 공간 일반화를 위한 일반화 연산은 다음과 같이 정의한다.

[정의 3.9] 이동객체의 이력객체 MH_i 와 공간 개념 계층 SC 그리고 일반화 수준 I 이 주어졌을 때 이동객체의 이력객체 일반화는 이력객체의 각 위치 값을 공간 개념 SC 상에서의 I 수준의 값으로 변환하는 것을 말한다. 일반화 연산 $SCGeneralize()$ 은 다음과 같이 정의한다.
 $SCGeneralize(MH_i, SC, I) : \exists P_i \rightarrow R_i, \text{for } \forall P_i$

$$\in MH_i[ES], R_i \in SC(I)$$

$SC(I)$ 는 공간 개념계층 상의 I 수준의 모든 지역 객체들의 집합이다. 이동객체의 이력객체 MH_i 에 대한 위치 일반화는 공간 개념 계층을 이용하여 이동 위치를 개념계층 상의 값으로 변환하는 것이다. 공간 위치의 일반화를 하는데 있어 공간 위치와 지역에 대한 공간 개념 계층 간에는 일반화 관계가 명시적으로 존재하지 않는다. 일반적으로 공간상의 두 객체간의 위상관계에 관한 공간 서술자를 통해 점객체와 면객체간의 포함관계를 조사할 수 있다. 이러한 공간 서술자는 $contain()$ 이라는 용어로 정의되어 사용된다. 이 논문에서는 이러한 공간 서술자인 $contain()$ 연산을 지원한다고 가정하였다.

3.2.2 이력객체의 위치 요약

이동객체의 빈번한 위치 이동에 따라 이력객체를 구성하는 각 위치 변화의 카디날리티가 상당히 방대하다. 이는 패턴 탐사를 위한 탐색영역을 증가시키는 문제가 있다. 특히 각 이동을 지역으로 일반화하여 패턴 탐사를 함에 따라 지역이 중복적으로 시퀀스를 구성하게 된다. 빈번한 이동에 대한 지역의 중복적 표현을 시퀀스로부터 제거함으로써 시퀀스의 크기를 줄이고 탐사의 효율성을 증대시킬 수 있다.

한 이동객체의 이동 변화는 연속적이지만 빈번한 이동 패턴 탐사를 위하여 이동객체는 다수의 이동 변화를 가질 수 있다. 이동 변화는 특정 시점의 특정 위치로부터 시작하여 다음 위치 이동의 시작이 있을 때까지의 연속적인 위치 변화로써 하나의 시퀀스를 형성한다.

이동객체의 이력객체에 대하여 위치 요약을 위하여 거리 제약조건을 정의한다.

[정의 3.10] 이동객체의 이력객체 MH_i 에 대하여 거리 제약 조건(distance constraint) min_dist 는 이력객체 내의 인접한 두 이동 P_i, P_{i+1} 간에 만족해야 하는 최소 이동 거리이다.

이동객체의 위치 요약은 인접한 두 이동은 최소 거리 임계치 min_dist 를 적용하여 다음의 식 (3.1)을 만족하도록 하는 것을 말한다. 여기에서 $dist()$ 함수는 두 이동간의 공간상의 Euclidian 거리를 구하는 함수이다.

$$dist(P_i, P_{i+1}) > min_dist, \text{for } 1 \leq i < n \quad (1)$$

위치 요약을 위하여 min_dist 임계치를 적용함에 있어서 발생할 수 있는 문제점은 이동 위치를 이동 지역 수준으로 일반화 할 경우에 위치 요약으로 인하여 다음 이동 지역에 해당하는 이동 위치가 요약된다는 것이다. 이는 패턴 탐사 시 중요한 의미를 가질 수 있는 데이터의 손실을 유발하기 때문에 위치 요약을 위한 min_dist 적용시 같은 지역수준에서의 min_dist 적용이 필요하다.

3.2.3 이동 시퀀스

이동객체 데이터 집합 MD 는 객체 식별자 OID 와 객체의 위치 이동의 연속적인 나열로 구성되어 있다. 이동 패턴 탐사를 위한 이동 시퀀스 S 는 위치 이동의 연속적인 모임으로써 $S = \{LC_1, LC_2, \dots, LC_n\}$ 으로 정의한다. 여기서 $LC_i = \langle (x_i, y_i), VT_i \rangle$ 로써 2차원 공간 평면상의 좌표 x_i, y_i 와 이동 시간 VT 로 구성된다. 예를 들어 $S = \{\langle (100, 95), 10 \rangle, \langle (125, 93), 22 \rangle, \langle (160, 117), 27 \rangle, \langle (180, 120), 32 \rangle\}$ 와 같이 표현된다.

한 이동객체의 이동 시퀀스 S 는 구분되는 이동객체들의 연속적인 위치 이동으로 이동객체가 존재하는 한 변경 위치 추가만 존재하는 매우 긴 시퀀스이다. 이동 패턴 탐사 시 탐사 영역이 방대해 지고 또한 모든 이동간의 시간 간격을 고려할 때 모든 이동은 연속적인 한 이동이 아닌 단위 이동으로 볼 수는 있다. 특정 이동간의 이동 지연시간은 이동에 대하여 의미적으로 이동의 끝과 새로운 이동의 시작으로 규정할 수 있는 것이다. 따라서 효과적이고 빈발한 이동 패턴 탐사를 위하여 각 이동간의 시간 제약조건을 고려함으로써 한 이동객체의 이동 시퀀스 S 는 단위 이동 시퀀스로 나누어 진다. 이를 위하여 시간 제약조건을 정의한다.

[정의 3.11] 최대 시간 간격 max_gap 은 연속하고 있는 두 이동 LC_{i-1}, LC_i 간에 만족해야 하는 시간 제약 조건이다.

[정의 3.11]의 시간 제약조건 max_gap 는 이동 시퀀스 내의 두 위치 이동이 다음 식을 만족하지 못하면 서로 다른 이동 시퀀스로 분리하기 위하여 사용할 수 있다.

$$LC_i[VT] - LC_{i-1}[VT] \leq max_gap \quad (3.2)$$

[정의 3.12] 단위 이동 시퀀스 US 는 이동 시퀀스 S 를 구성하는 시간 순서에 의해 중첩되지 않는 인접한 두 이동간 최대 시간 간격을 만족하는 시퀀스이다.

정의에 의해 한 시퀀스 S 상의 한 이동 LC_i 는 유일하게 한 단위 이동 시퀀스 US_i 에만 포함된다. 만일 시퀀스 S 상에 LC_i 에 인접한 이동 LC_{i+1} 이 있어 두 이동이 식 (3.2)를 만족하지 못하면 LC_{i+1} US_i 이다. 따라서 이동 시퀀스 S 는 중첩되지 않는 연속적 단위 시퀀스의 집합으로 구성될 수 있다. 모든 단위 이동 시퀀스 간에 다음의 식 (3.3)을 만족하여야 한다.

$$min(US_i[VT]) - max(US_{i-1}[VT]) > max_gap, 2 \leq i \leq m \quad (3.3)$$

이동 시퀀스 S 에 대하여 패턴 탐사를 바로 적용할 수 없다. 패턴 탐사를 위해서는 각 패턴에 나타날 수 있는 항목(item)의 집합적 규정이 필요한데 이동객체의 경우 각 이동

위치의 구분 가능한 항목집합 구성이 힘들고 패턴에서도 의미가 없다. 따라서 이 논문에서는 일반화된 이동 시퀀스를 정의한다.

[정의 3.13] 일반화된 이동 시퀀스는 위치 이동에 따른 연속적인 이동 지역들의 모임으로써 $GS = \{AC_1, AC_2, \dots, AC_k\}$ 로 정의한다.

여기서 $AC = \langle a_i, VT_i \rangle$ 로 $a_i \in A$ 는 이동 지역이고 VT_i 는 이동시간이다. 이동 지역 $A = \{A_1, A_2, \dots, A_m\}$ 는 공집합이 아닌 m 개의 이동 가능한 지역(area)들의 집합이다. 이동객체 데이터 집합 MD 에 존재하는 이동시퀀스 S 를 이동 패턴 탐사를 위하여 일반화된 이동시퀀스 GS 로의 변환을 통한 이동 시퀀스 데이터베이스 GSD 를 형성할 필요가 있다. 이동 시퀀스 데이터베이스 생성을 위해서는 각 이동 시퀀스를 시간 제약 조건 max_gap 을 적용함으로써 단위 이동 시퀀스로 변환한다. 그리고 최소 이동 거리 min_dist 적용을 통한 이동 위치의 요약과 $SCGeneralize()$ 연산을 통한 위치 일반화 과정이 수행된다.

3.2.4 이동 패턴 탐사 문제

두 이동 시퀀스 $a = \langle a_1, a_2, \dots, a_n \rangle$ 와 $b = \langle b_1, b_2, \dots, b_m \rangle$ 가 있을 때 $i_1 < i_2 < \dots < i_n$ 와 같은 정수에 대하여 $a_1 = b_{i_1}, a_2 = b_{i_2}, \dots, a_n = b_{i_n}$ 이면 이동 시퀀스 a 는 b 의 부분 시퀀스(subsequence)이다. 이때 시퀀스 b 는 시퀀스 a 를 포함한다고 말한다. 예를 들어 시퀀스 $\langle a_1, a_4, a_5 \rangle$ 는 $\langle a_2, a_1, a_3, a_4, a_5 \rangle$ 의 부분 시퀀스이다. 이러한 포함관계에 대한 정의는 기본적으로 Agrawal[10]과 동일하다. 하지만 이동 시퀀스 간의 포함관계는 시간 범위에 대한 제약조건을 추가적으로 필요로 한다.

두 이동 시퀀스는 포함관계를 만족하기 위해서는 최대 이동 시퀀스 범위 max_span 내에서의 이동에 대하여 고려해야 하며 또한 각 이동 역시 최대 시간 간격 max_gap 내에 존재해야 한다. 이를 형식적으로 표현하면 두 이동 시퀀스 $a = \langle a_1, a_2, \dots, a_n \rangle$ 와 $b = \langle b_1, b_2, \dots, b_m \rangle$ 가 있을 때 $i_1 < i_2 < \dots < i_n$ 와 같은 정수에 대하여 다음을 만족하면 이동 시퀀스 b 는 a 를 포함한다.

- $a_1 = b_{i_1}, a_2 = b_{i_2}, \dots, a_n = b_{i_n}$ 이고,
- $b_{i_n}[VT] - b_{i_1}[VT] \leq max_span$ 이고
- $b_{i_j}[VT] - b_{i_{j-1}}[VT] \leq max_gap, 2 \leq j \leq i_n$

어떤 이동 시퀀스에 대하여 시퀀스의 지지도가 사용자가 지정한 임계치 이상일 경우 이 시퀀스는 빈발하다(frequent)라고 말한다. 어떤 시퀀스가 빈발한지를 조사하기 위해 사용자가 지정하는 최소한의 임계치를 최소 지지도(minimum support)라 하고 min_sup 로 표기한다. 따라서 빈발 시퀀스(frequent sequence)는 사용자가 지정한 최소 지지도 $min_$

*sup*를 만족하는 이동 시퀀스를 말한다.

시공간 이동 패턴 탐사는 객체의 이동 데이터로부터 공간상에서의 시간에 따른 이동 패턴을 탐사하는 문제로서 다음과 같이 정의한다.

[정의 3.14] 시공간 이동 패턴 탐사는 이동 데이터 집합 *MD*로부터, 시간 제약조건 *max_gap*, *max_span* 과 공간 제약조건 *min_dist* 그리고 최소지지도 *min_sup* 을 만족하는 모든 빈발한 이동 시퀀스를 탐사하는 문제이다.

4. 시공간 이동 패턴 탐사 기법

4.1 시공간 이동 패턴 탐사 단계

탐사 기법은 크게 이동 시퀀스 추출 모듈과 이동 패턴 탐사 모듈로 구분된다. 이동 시퀀스 추출 모듈은 패턴 탐사를 위한 전처리 과정으로써 시공간 이동 패턴 탐사를 위하여 이동객체 데이터 집합으로부터 일반화된 이동 시퀀스를 생성하는 과정이다. 데이터베이스상의 이동 이력 객체를 정렬하고 이동객체의 이력 객체는 위치 요약과 위치 일반화 과정을 통하여 일반화된 이동 시퀀스가 생성된다. 이러한 과정에 있어서 시간 제약조건 *max_gap* 과 공간 제약조건 *min_dist* 가 함께 적용된다. 이동 패턴 탐사 모듈은 일반화된 이동 시퀀스들로부터 *max_span* 과 *max_gap* 의 시간 제약조건을 만족하며 사용자가 지정한 최소지지도 *min_sup* 을 만족하는 빈발 시퀀스를 생성한다.

4.1.1 이동 시퀀스 추출

이동 시퀀스 추출 모듈은 정렬단계과 이동 시퀀스 생성 단계들로 구성된다 이동 시퀀스 생성 단계에서는 위치 요약 및 일반화 과정을 이용한다. 이동 시퀀스 추출 모듈을 각 단계별로 자세히 설명한다.

① 정렬 단계

정렬 단계는 이동객체 데이터 집합 *MD*를 이동객체 식별자(*MOID*)와 이동측정 시간(*VT*)에 의하여 각 이동객체 별 시간에 따른 순서로 정렬한다.

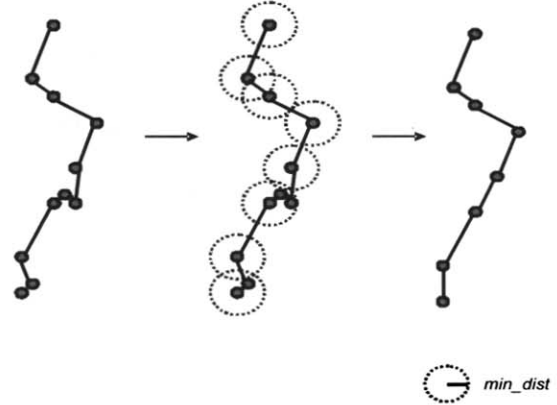
② 이동 시퀀스 생성 단계

정렬된 이동객체 데이터 집합으로부터 이동 시퀀스를 생성하는 단계는 최소 거리 임계치를 이용한 위치 요약 과정, 최대 시간 간격 임계치를 이용한 단위 이동시퀀스 분리 과정 그리고 일반화 연산 *SCGeneralize()*를 이용한 위치 일반화 과정의 세 과정으로 구성된다.

● 위치 요약 과정

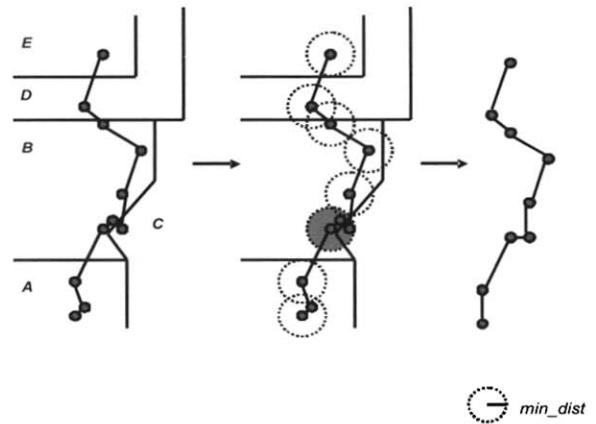
위치 요약 과정은 (그림 4.1)과 같이 최소 거리 임계치 *min_dist*를 적용하여 연속적 위치 이동을 요약하게 된다.

이러한 *min_dist* 임계치 값의 설정은 응용 영역에서 다양하게 규정할 수 있으나 한가지 방법으로는 이동객체의 평균 이동속도 *V*와 이동 시간 *T*의 곱으로써 설정할 수 있다.



(그림 4.1) 이력 객체의 위치 요약

지역 경계를 고려한 이력객체의 위치 요약은 같은 지역수준 내에서의 이동에 대하여 이동간의 최소거리인 *min_dist*를 적용함을 의미한다. (그림 4.2)의 경우는 (그림 4.1)과 다른 이력객체의 위치 요약을 보여준다. 이동에 대한 제약조건 적용 시 지역 경계를 고려함으로써 B 지역에서 C 지역으로의 이동에 대한 위치 이동 정보를 잃어버리지 않게 한다.



(그림 4.2) 지역 경계를 고려한 이력 객체의 위치 요약

이전의 위치와 현재의 위치간의 유클리디안 거리가 최소 거리 임계치 보다 작으면 단위시퀀스에 포함되지 않고 배제된다. 이러한 요약과정에 있어서 지역 경계를 고려 함으로써 최소 임계치 내에 존재해도 지역 경계를 벗어난 경우에는 요약되지 않도록 한다.

● 단위 이동 시퀀스 분리 과정

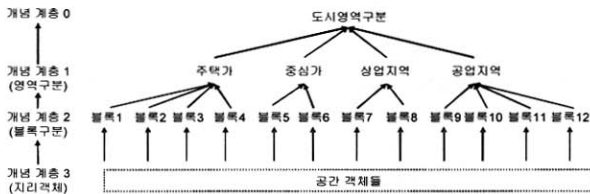
단위 이동 시퀀스 분리 과정은 최대 시간 간격 임계치

*max_gap*을 적용하여 하나의 이동객체의 연속적 이동을 여러 개의 단위 이동 시퀀스로 분리하는 과정이다. 각 단위 이동 시퀀스는 유일한 단위 이동 시퀀스 식별자를 포함하는데 한 이동객체의 이동 시퀀스에서 두 이동간에 시간 간격이 임계치를 벗어나면 새로운 식별자를 할당하여 새로운 단위 이동 시퀀스를 형성한다.

● 일반화 과정

일반화 과정은 분리된 단위 이동 시퀀스들에 대하여 공간 개념 계층을 적용하여 요구되는 수준의 공간 개념으로 일반화하는 과정이다. 일반화 연산인 *SCGeneralize()*를 적용함으로써 일반화된 이동 시퀀스를 생성한다. 공간 일반화를 수행하기 위해 지식 컴포넌트는 배경지식을 구성하고 있어야 한다. 이동객체의 각 이동의 위치를 개념계층 *SC*상에서의 *l* 수준으로 일반화하는 *SCGeneralize()* 연산은 다음 세 단계로써 이루어진다.

- 단계 1: 각 위치에 대하여 공간 서술자 *contain()* 연산을 이용하여 공간 개념 계층상의 단말 수준 지역으로 일반화한다.
- 단계 2: 사용자의 일반화 수준 레벨에 따라 개념 계층 *SC*를 따라 상위 개념으로 일반화 한다.
- 단계 3: 개념 계층을 따라 상위 노드로 수준 *l*에 도달할 때까지 단계 2과정을 반복한다.



(그림 4.3) 공간 개념 계층의 예

(그림 4.3)는 공간 개념 계층의 예로 일반화된 이동 시퀀스 추출에 적용함으로써 지리객체 수준으로 일반화된 이동 시퀀스를 얻는다. 이동 데이터 집합에 대하여 개념수준 2인

MOID	USID	X_COORD	V_COORD	VT	PRID
100	1	100	610	2002-10-10 10:10:15	블록 1
100	1	130	630	2002-10-10 10:25:05	블록 2
100	2	110	645	2002-10-10 10:42:01	블록 1
100	2	150	650	2002-10-10 10:51:30	블록 4
100	2	155	620	2002-10-10 10:05:49	블록 5
100	3	256	300	2002-10-10 10:01:31	블록 2
100	3	211	300	2002-10-10 10:11:31	블록 4
100	3	186	298	2002-10-10 10:25:31	블록 5
100	3	150	299	2002-10-10 10:40:11	블록 9
100	4	550	425	2002-10-10 10:04:51	블록 7
100	4	510	430	2002-10-10 10:19:00	블록 4
100	4	515	458	2002-10-10 10:47:00	블록 5
100	4	517	477	2002-10-10 10:56:26	블록 6

MOID	MovingSequences
100	<블록1, 블록2>, <블록1, 블록4, 블록5>
102	<블록2, 블록4, 블록5, 블록9>
103	<블록7, 블록4, 블록5, 블록9>

(그림 4.4) 일반화된 이동 시퀀스의 예

블록수준의 개념으로 일반화하여 얻게 되는 이동 시퀀스의 예를 (그림 4.4)에서 보여준다.

4.1.2 빈발 이동 패턴 탐사

빈발 이동 패턴 탐사 모듈은 크게 다음의 세 가지 단계를 거친다.

- 빈발 1-시퀀스(frequent 1-sequence) 단계 : 최소 지지도 *min_sup*를 만족하는 모든 길이가 1인 빈발 시퀀스를 형성하는 단계이다.
- 후보 시퀀스 생성단계(candidate sequence generation) : 이 단계는 길이 *k*>2인 후보 시퀀스를 생성하는 단계로 길이가 *k*인 후보 시퀀스 *C_k*는 이전의 빈발 시퀀스인 *F_{k-1}*의 조인을 통해 생성한다.
- 후보 시퀀스 가지치기 단계(candidate sequence pruning) : 이 단계는 후보 시퀀스 *C_k*에 대하여 이동 시퀀스 데이터 집합에 대하여 최대 이동 시퀀스 범위 *max_span*과 최대 시간 간격 *max_gap* 임계치를 만족하는 이동객체에 대하여 최소 지지도 임계치를 만족하는지 여부를 조사함으로써 후보 시퀀스로부터 빈발 시퀀스를 생성하는 단계이다.

빈발 이동 패턴 탐사는 빈발 1-시퀀스로를 이용하여 빈발한 *k* 단계의 시퀀스가 더 이상 존재하지 않을 때까지 후보 시퀀스 생성단계 외 후보 시퀀스 가지치기 단계를 반복적으로 수행한다.

4.2 시공간 이동 패턴 탐사 알고리즘 MPMine

이 절에서는 이러한 각 단계에 해당하는 알고리즘을 가상코드로 표현하여 기술한다. (알고리즘 4.1)은 시공간 이동 패턴 탐사 알고리즘인 MPMine을 가상코드로 기술한 것이다. MPMine알고리즘은 입력으로 이동객체 데이터 집합, 시간 제약조건, 공간 개념계층, 공간 개념 일반화 수준 임계치, 배경 지도 데이터 등과 임계치로 최소 거리 임계치, 최대 이동 시퀀스 범위 임계치, 최대 시간 간격 임계치 및 최소 지지도 임계치 등을 취한다. 결과로써는 시공간 이동 패턴 집합을 반환한다. 알고리즘에서 후보 이동 시퀀스 생성 및 가지치기의 효율적인 수행을 위해 해쉬트리를 이용함으로써 효과적으로 패턴을 탐사할 수 있다.

(알고리즘 4.2)는 정렬된 이동객체 데이터 집합에 대하여 공간 개념 계층과 임계치 등을 적용하여 이동객체에 대한 단위 이동 시퀀스를 추출해 내는 알고리즘 *MovingSequence Extract()*의 일련의 과정을 가상코드로 기술한 것이다. 입력인 이동객체 데이터 집합 *RS*의 각 이동 시퀀스에 대하여 시간과 거리 임계치인 *min_dist*, *max_gap*을 만족하도록 단위 이동 시퀀스를 형성하며 이때 최소 거리 임계치 적용 시 경계를 고려한 단위 이동 시퀀스를 형성한다. 마치

막 단계에서는 *SCGeneralize()* 연산을 통해 공간 개념 계층인 *SCH*에 대하여 요구되는 일반화 임계치인 *gen_threshold* 수준의 개념 값으로 각 단위 이동 시퀀스의 이동변화를 일반화 한다.

```

Algorithm : MPMine()
Input : MODB : 이동객체 데이터 집합
        TP : 시간 제약 조건(temporal predicate)
        MAP : 배경 지도 데이터 집합
        SCH : 공간 개념 계층
        gen_threshold : 공간 개념 일반화 수준 임계치
        min_dist : 최소 거리 임계치
        max_span : 최대 이동 시퀀스 범위 임계치
        max_gap : 최대 시간 간격 임계치
        min_sup : 최소 지지도 임계치
Output : MPRuleSet : 시공간 이동 패턴 집합

Method :
// 시간 제약조건을 만족하는 이동객체 데이터 집합 형성
RS ← GetMiningRelevantData(MODB, TP);

// 객체 식별자와 시간을 기준으로 정렬
RS ← Sortby(RS, moid, vt);

// 임계치와 공간 개념계층의 적용을 통한 이동 시퀀스 추출
MS ← MovingSequenceExtract(RS, MAP, SCH, gen_
threshold, min_dist, max_gap);

// 빈발 이동 패턴 탐사
FMP ← FrequentMovingPatternMine(MS, max_span, max_
gap, min_sup);

// 이동 패턴 규칙 생성
MPRuleSet ← GenerateRules(FMP, TP);
return MPRuleSet;
    
```

(알고리즘 4.1) 이동 패턴 탐사 알고리즘

```

Algorithm : MovingSequenceExtract();
Input :
        RS : 이동객체 데이터 집합
        MAP : 배경 지도 데이터 집합
        SCH : 공간 개념 계층
        gen_threshold : 공간 개념 일반화 수준 임계치
        min_dist : 최소 거리 임계치
        max_gap : 최대 시간 간격 임계치
Output :
        MS : 일반화된 이동 시퀀스 집합

Method :
        MS = empty;

// 각 이동객체의 이동 시퀀스에 대하여
FOREACH (s IN RS)
{
        usid 1; // 단위 이동 시퀀스 식별자 초기화
        PreLC 1; // 이전의 이동 인덱스 초기화

// 새로운 단위 시퀀스 생성
        sid = getSid(s);
    
```

```

        US = new US(sid, usid);
        US = s[1];
// 시퀀스 분리 단계
        FOR (i = 2; i < s.length; i++) {
// 최대 시간 간격 임계치 적용
                IF (s[i-1], s[i] > max_gap) {
                        MS = MS + US;
// 시퀀스 분리 및 새로운 단위 이동 시퀀스 생성
                        usid = getNextUsid();
                        US = new US(sid, usid);
                        US = s[i];
                        PreLC = i;
                }
                ELSE {
// 최소 거리 임계치 적용
                        IF (dist(s[PreLC], s[i]) > min_dist) {
                                US = US + s[i];
                                PreLC = i;
                        }
                        ELSE {
// 지역 경계를 고려한 시퀀스 요약
                                PreArea = GeneralizeIntoRegion
                                        (S[PreLC], MAP);
                                CurArea = GeneralizeIntoRegion(S[i],
                                        MAP);
                                IF (PreArea <> CurArea) {
                                        US = US + s[i];
                                        PreLC = i;
                                }
                        }
                }
                MS = MS + US;
        }
// 공간 개념 계층을 적용한 일반화 수행
        MS SCGeneralize(MS, SCH, gen_threshold);
return MS;
    
```

(알고리즘 4.2) 이동 시퀀스 추출 알고리즘

```

Algorithm : FrequentMovingPatternMine();
Input :
        MS : 이동 시퀀스
        max_span : 최대 이동 시퀀스 범위 임계치
        max_gap : 최대 시간 간격 임계치
        min_sup : 최소 지지도 임계치
Output :
        FMP : 빈발 이동 패턴

Method :
        F1 ← (frequent 1 - sequence); // 빈발 1-시퀀스 생성
        FMP ← ∅;
        for (k = 2; Fk ≠ ∅; k++)
        {
// 후보 시퀀스 생성
                Ck ← GenerateCandate(Fk-1);
// 지지도 조사 및 가지 치기 수행
    
```



```

    Fk ← CountCandidateAndPrune(Ck, max_span,
    max_gap, min_sup);
    // 빈발 이동 패턴 추가
    FMP FMP Fk with support value;
}
return FMP;

```

(알고리즘 4.3) 빈발 이동 패턴 탐사 알고리즘

(알고리즘 4.3)은 일반화된 이동 시퀀스 MS와 최대 이동 시퀀스 범위, 최대 시간간격, 최소 지지도 등의 임계치 등을 입력으로 하여 빈발한 시공간 이동 패턴을 탐사하는 알고리즘 알고리즘 FrequentMovingPatternMine()의 각 단계를 가상코드로 기술한 것이다.

5. 실험 및 평가

새롭게 제안한 시공간 이동 패턴 탐사를 이동 패턴 탐사를 문제로 하지 않는 기존의 탐사 알고리즘(순차 패턴, 일반화된 순차 패턴)과 수행 효율성에 대하여 비교 평가하는 것은 적절하지 못하다. 제안된 시공간 이동 패턴 탐사 기법의 성능 평가는 세가지 측면에서 고려될 수 있다. 첫째, 이동 시퀀스 증가에 따른 탐사 기법의 효율성, 둘째, 이동 시퀀스 길이 증가에 따른 패턴 탐사 효율성 마지막으로 임계치 변화에 따른 패턴 탐사 효율성 등이다. 이동시퀀스 증가의 경우에는 이동 객체의 수에 변화와 관련이 있으며 이에 대한 평가는 기존의 순차 패턴의 경우와 충분히 유사하다고 판단되기 때문에 이 논문에서는 특별히 언급하지 않겠다. 임계치를 통한 평가는 일반적인 순차패턴[10]의 경우과 같이 최소 지지도와 GSP[13]의 기법처럼 최대 시퀀스 범위의 임계치를 이용하여 가능하다. 하지만 이동시퀀스의 길이 증가나 이동 시퀀스 증가의 경우에는 다음 절에서 설명하는 바와 같이 이동 데이터 생성 및 이동 시퀀스 요약을 위한 다양한 파라미터 변수들을 이용함으로써 가능하다.

5.1 실험 데이터 생성

<표 5.1>에서 제시한 바와 같이 제안한 기법의 효율성 분석을 위한 실험데이터 생성을 위한 도구로써 Java로 구현된 IBM CitySimulator를 이용하여 기본 이동 데이터 집합을 생성하였다. 이동 데이터 집합은 도심지의 3차원 공간 상에서 이동하는 객체에 대하여 시간 간격이 매우 작은 방대한 양의 데이터 집합을 생성한다. 기본 이동 데이터 집합을 실험에 적합한 형태의 실험 데이터로 변환하기 위하여 다양한 파라미터를 통하여 이동 데이터 집합을 생성한다. 생성된 이동 데이터 집합은 Oracle 8i의 Loader를 통하여 DBMS상에 적재 된다. 이러한 실험 데이터 생성 및 적재를 위한 도구로써 C#을 이용한 실험 데이터 생성기를 구현하였다. 실험 데이터 생성 환경은 Pentium4 1.5 GHz CPU를

장착한 Windows 2000서버이다. 실험 데이터 생성기는 입력 파라미터로 이동객체의 수 및 이동 시간, 이동 기간 등이 입력되면 CitySimulator를 통해 기본 이동 데이터 집합이 생성된다. 이러한 이동 데이터 집합은 위치발생 시간 간격을 통해 필터링되고 DBMS에 적재되기 위한 형식으로 변환된 후 정렬된다.

성능 평가를 위한 데이터 생성 및 실험을 위해 적용되는 파라미터 변수들은 <표 5.1>과 같다. 데이터 집합 생성을 위한 파라미터 변수와 이동 패턴 탐사에 영향을 미치는 변수들로 나누어 표현되었다.

<표 5.1> 성능 평가를 위한 파라미터 변수

단 계	파라미터 변수	의 미
데이터 집합 생성	<i>N</i>	이동객체의 수
	<i>D</i>	이동 기간
	<i>M</i>	최대 이동 수
	<i>G</i>	측정 간격(샘플링 비율)
시간 이동 패턴 탐사	<i>max_gap</i>	최대 시간 간격
	<i>min_dist</i>	최소 이동 거리
	<i>min_sup</i>	최소 지지도
	<i>max_span</i>	최대 이동 시퀀스 범위
	<i>gen_threshold</i>	일반화 수준

5.2 평 가

시공간 이동 패턴 탐사 기법의 성능 평가를 위하여 데이터 생성기를 통해 생성된 각 실험 데이터 집합에 대한 특성은 <표 5.2>과 같다.

<표 5.2> 이동 시퀀스 증가에 따른 실험 평가를 위한 실험 데이터

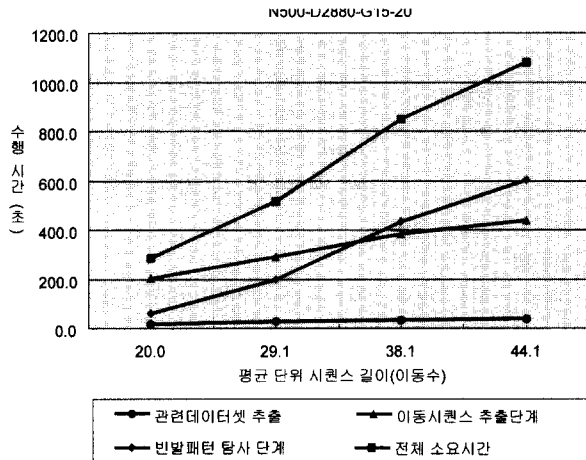
데이터 집합 이름	이동 객체수 (N)	측정 기간 (D)	원시 데이터 크기	측정 간격 (G)	데이터 크기
N500-D2880-G15-20	500	2일	18.1GB	15분~20분	4MB
N500-D2880-G15-30	500	2일	18.1GB	15분~30분	2.88MB
N1000-D2880-G15-20	1000	2일	38GB	15분~20분	8.75MB
N1000-D2880-G15-20	1000	2일	38GB	15분~30분	6.82MB

이동 시퀀스 길이 증가에 따른 패턴 탐사의 효율성 분석은 <표 5.3>와 같이 실험 파라미터를 설정하여 수행하였다. 이동 시퀀스 길이 증가에 따른 패턴 탐사의 효율성은 최대 시간 간격 임계치인 *max_gap*을 감소 시키거나 시간 한정 범위를 증가 시킴에 따라 이동 시퀀스 길이의 증가와 성능 간의 관계를 분석할 수 있다. 이 실험에서는 실험 데이터 특성상 각 이동간 간격은 어느 정도 일정 범위에 존재하므로 *max_gap* 변화에 따른 이동 시퀀스 길이 증가는 의미가 없다. 따라서 시간 한정 크기에 변화를 주어 단위 이동 시퀀스의 길이를 증대시켜 실험하였다.

〈표 5.3〉 이동 시퀀스 길이에 따른 성능 평가를 위한 실험데이터

데이터 집합 이름	N500-D2880-G15-20
공간 개념 계층 및 수준	지역구분(레벨 2: 블록수준)
최소 이동 거리 (min_dist)	25M
최대 시간 간격 (max_gap)	60분
최대 이동 시퀀스 범위 (max_span)	시간 한정 값과 동일
최소 지지도 변화 (min_sup)	3%
시간 한정	7 시간, 10 시간, 13 시간, 16 시간

실험 결과 (그림 5.5)과 같이 시퀀스의 길이 증가에 따라 전체 소요시간에서 빈발 이동 패턴 탐사 단계에 소요되는 시간이 급격히 증가하고 있는 반면 관련 데이터 추출 및 이동 시퀀스 추출 단계는 일정하게 증가하고 있음을 알 수 있다. <표 5.4>과 <표 5.5>은 각 단계에 소요되는 시간과 수행 결과를 보여준다.



(그림 5.5) 이동 시퀀스 길이 증가에 따른 수행 성능 비교

〈표 5.4〉 이동 시퀀스 길이 증가에 따른 단계별 소요시간

시간 한정 크기(시간)	소요시간(초)				
	관련 데이터 집합 추출단계	이동시퀀스 추출단계	빈발 패턴 탐사 단계	DB 소요시간	총 소요시간
7(8시~15시)	17.9	202.7	1275.1	275.8	283.0
10(8시~18시)	25.9	290.9	195.8	502.9	512.6
13(8시~21시)	33.6	382.6	432.3	836.4	848.5
16(8시~24시)	37.1	440.8	600.3	1064.1	1078.2

시간 한정에 의하여 기본적으로 각 객체는 요약되기 이전에 동일한 크기의 시퀀스 길이를 갖는다. 이러한 시퀀스의 길이는 최소 거리 임계치와 최대 시간 간격 임계치에 따라 변화하게 된다. 이 실험의 이동객체는 규칙적 이동을 갖지만 이와 달리 불규칙한 이동을 갖는 이동객체의 경우 제안한 두 임계치를 고려한 단위시퀀스 생성은 탐사 효율 더 많은 영향을 미칠수 있다. 따라서 응용환경에 따라 이러한 최대 간격 임계치와 최소 거리 임계치의 정의는 수행 성

능에 매우 중요한 역할을 함을 알 수 있다.

〈표 5.5〉 이동 시퀀스 길이 증가에 따른 수행 결과

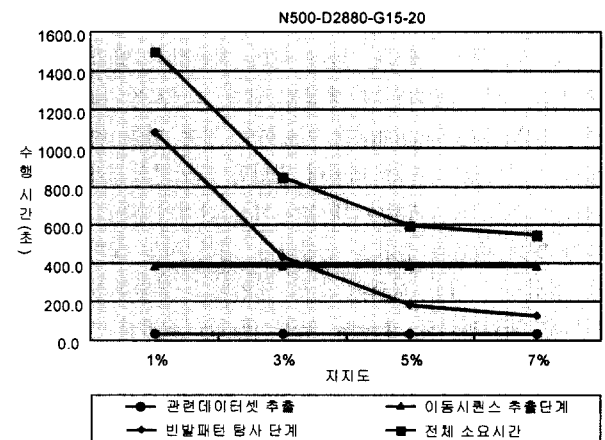
시간 한정 크기(시간)	2-Sequence	3-Sequence	총 패턴수	평균 이동수	평균 시퀀스 길이
7 (8시~15시)	25	0	25	1.8	20.0
10 (8시~18시)	51	0	51	2.1	29.1
13 (8시~21시)	79	0	79	2.5	38.1
16 (8시~24시)	94	5	99	2.7	44.1

마지막으로 임계치 변화에 따른 패턴 탐사 효율성은 최소 지지도나 최대 시퀀스 범위와 같은 임계치 변화와 수행에 미치는 영향 관계를 분석하였다. 임계치 변화에 따른 패턴 탐사 효율성 분석 실험에서는 표 이동 데이터 특성상 각 최대 시퀀스 범위는 시간 한정 값과 동일하게 가정하고 최소 지지도를 2%씩 증가 시킴에 따른 수행 성능을 실험 분석하였다.

〈표 5.6〉 지지도 변화에 따른 성능 평가를 위한 실험데이터

데이터 집합 이름	N500-D2880-G15-20
시간 한정	8시~21시(13시간)
공간 개념 계층 및 수준	지역구분(레벨 2: 블록수준)
최소 이동 거리(min_dist)	25M
최대 시간 간격(max_gap)	60분
최대 이동 시퀀스 범위(max_span)	12 시간
최소 지지도 변화(min_sup)	1%, 3%, 5%, 7%

실험 결과 (그림 5.6)과 같이 지지도를 2%씩 증가 시킴에 따라 빈발 이동 패턴 탐사 단계에 소요되는 시간이 급격히 감소하고 있는 반면 다른 두 단계는 일정한 수행 시간을 보이고 있다. 또한 지지도 증가는 탐사 패턴 수에 많은 영향을 미치고 있음을 알 수 있다.



(그림 5.6) 지지도 변화에 따른 수행 성능 비교

지지도와 최대 이동 시퀀스 범위 임계치는 탐사될 패턴의 수에 영향을 미치는 중요한 요소이다. 이러한 임계치는 반복적인 이동 패턴 탐사를 통하여 응용환경에 적합한 규모의 패턴을 탐사할 수 있다. 하지만 최대 이동 시퀀스 범위 임계치는 응용환경에 대한 사전 분석을 통하여 설정될 수 있다.

이상의 이동 패턴 탐사 기법에 대한 실험 평가 결과 및 문제점을 요약하면 다음과 같다.

첫째, (그림 5.5)에서 보는 바와 같이 시간한정과 최소거리 임계치 변화를 통해 생성된 이동 시퀀스의 길이 변화는 이동 패턴 탐사 수행 시간 및 탐사되는 패턴의 수에 많은 영향을 미친다. 특히 기법에서 제안하고 있는 최소 거리 임계치와 최대 시간 간격을 이용한 이동 시퀀스 추출 알고리즘은 성능 향상에 중요한 역할을 수행할 수 있음을 알 수 있다. 따라서 응용환경에 따라 탐사 비용을 줄이기 위해 적절한 시간 한정과 최소 거리 임계치의 설정은 매우 중요하다.

둘째, 위치 일반화를 위한 연산 SCGeneralize()는 데이터 량에 비례하여 일정한 수행 성능을 보임을 알 수 있다. 이 논문에서는 이동 시퀀스 추출 시 위치 일반화 과정을 DBMS의 저장 함수 개념을 이용하여 수행되었다. 따라서 수행 성능은 위치 일반화를 위한 개념 계층의 단계 수에 따라 DBMS내부의 조인 연산 수의 증가와 관계가 있음을 알 수 있다.

셋째, 지지도 변화는 패턴 탐사 수행 시간 및 탐사되는 규칙의 수에 밀접한 영향을 미치고 있음을 확인하였다. 이동 패턴 탐사 수행에 있어 빈발 패턴 탐사 단계가 차지하는 비율이 지지도 변화에 매우 민감하게 반응하고 있음을 알 수 있다.

그러나 이 연구에서는 위치 일반화 및 빈발 이동 패턴 탐사 단계 등에 있어 다음과 같은 문제점이 확인 되었다.

첫째, 효율적인 이동 시퀀스 추출을 위해서는 데이터량과 개념 계층 구조 고려한 위치 일반화가 필요하다. DBMS 기반의 일반화 연산을 이동 시퀀스 추출 기법 내부에 포함하기 위해서는 개념 계층 구조 및 일반화 기법에 대한 개선이 필요하다.

둘째, 빈발 패턴 탐사의 수행성능이 지지도 변화에 민감하게 반응하고 있기 때문에 효율적으로 탐사 비용을 줄이기 위해서는 최대 이동 시퀀스 범위나 데이터 량을 고려하여 이동 시퀀스 탐사를 위한 효율적인 검색 구조로 개선하는 연구가 필요하다.

셋째, 전체적으로 실험에 사용된 이동 데이터는 균일한 이동 분포를 갖기 때문에 실 데이터를 통한 이동 패턴 탐사 기법의 효율성 분석은 부족하였다. 또한 반밀결합 구조를 통한 이동 패턴 탐사 시스템의 구현 구조에서 수행시간의 많은 부분이 DBMS 연산이기 때문에 DBMS 성능에 의존적인 문제점을 보였다. 따라서 향후에는 효율적인 패턴 탐사를 위해 시스템 내부의 캐싱 구조 및 인덱스 구조 등에 대한 연구가 필요하다.

6. 결 론

현재의 이동객체를 기반으로 하는 다양한 시공간 응용환경에서의 서비스 지원 시스템 개발을 위하여 중요한 문제 중의 하나는 방대한 이동객체의 위치 이동 데이터로부터의 의미 있는 지식인 시공간 이동 패턴을 탐사하는 것이다. 이러한 지식 유형을 탐사하기 위해서는 시간과 공간 특성, 다양한 시공간 관계 및 연사자 등을 활용해야 한다.

이 논문에서는 시간 제약조건인 최대 시간 간격, 최대 이동 시퀀스 범위 등과 공간 제약조건으로써 최소 이동 거리 등을 고려하였으며 공간 위상 연산인 contain()을 이용한 공간 개념화 연산 SCGeneralize()를 이용하여 효율적인 시공간 이동 패턴 탐사 기법인 MPMine 알고리즘을 제안하였다. 시공간 이동 패턴 탐사 문제 정의에 기반하여 의미 있는 패턴 탐사를 위해 요구되는 제약조건을 고려하여 시공간 이동객체 집합에 대한 정의와 이동 시퀀스 생성을 위한 다양한 시간 및 공간 제약 사항들에 대한 명확한 정의를 제시하였다. 또한 이동 패턴 탐사 문제에서 정의한 이동 시퀀스 생성을 위한 위치 요약 및 일반화를 위한 기법을 설계하였고 이를 기반으로 시공간 이동 패턴 탐사 기법 MPMine을 설계하였다.

제안한 기법은 기존의 일반적인 시간 패턴 탐사 기법과 달리 이동객체 데이터 집합으로부터 위치 요약 및 일반화를 통하여 탐색 공간을 줄일 수 있어 효율적으로 유용한 이동 패턴을 탐사하는 응용에 활용될 수 있다. 또한 제안한 시공간 이동객체 패턴 탐사 기법은 기존의 순차패턴 등에서 찾지 못했던 시간과 공간 의미가 함께 고려된 유용한 시공간 패턴 규칙을 탐사할 수 있다. 하지만 효율적 빈발 패턴 탐사를 위한 성능 개선을 통하여 향후에는 좀더 정확한 기법의 평가를 위해 다양한 이동 데이터 집합을 이용하여 다각적인 평가 수행이 필요하다.

향후 연구의 방향은 시공간 복합 위상 서술에 기반한 지식 탐사에 대한 연구이다. 시공간 복합 위상 서술은 이동 영역 객체 및 행위, 동작 등과 같은 복합적이고 복잡한 현상에 대한 분석을 가능하게 한다.

참 고 문 헌

- [1] T. Abraham, Knowledge Discovery in Spatio-Temporal Databases, School of Computer and Information Science, University of South Australia, Ph. D dissertation, 1999.
- [2] T. Bittner, "Rough sets in spatio-temporal data mining," In Proc. of the 1st International Workshop on Temporal, Spatial and Spatio-Temporal Data Mining(TSDM2000), 2000.
- [3] E. Mesrobian, R. R. Muntz, J. R. Santos, E. C. Shek, C. R. Mechoso, J. D. Farrara and P. Stolorz, "Extracting Spatio-Temporal Patterns from Geoscience Datasets," IEEE Work-

- shop on Visualization and Machine Vision, Seattle, WA, June, 1994.
- [4] I. Tsoukatos and D. Gunopulos, "Efficient Mining of Spatiotemporal Patterns," In Proc. of the 7th Int. Symp. on Spatial and Temporal Databases (SSTD), 2001.
- [5] X. Chen and I. Petrounias, "A framework for temporal data mining," In Proc. of the 9th International Conference on Database and Expert Systems Applications, 1998.
- [6] J. F. Roddick and M. Spiliopoulou, "Temporal data mining : survey and issues," Research Report ACRC-99-007, University of South Australia, 1999.
- [7] K. Koperski and J. Han, "Discovering of Spatial Association Rules in Geographic Information Databases," In Proc. of the 4th International Symposium on Large Spatial Databases, 1995.
- [8] K. Koperski, J. Han and J. Adhikary, "Mining knowledge in geographical data," to appear in Communications of the ACM, 1998.
- [9] J. F. Roddick and B. G. Lees, "Paradigms for Spatial and Spatio-Temporal Data Mining," Geographic Data Mining and Knowledge Discovery. Taylor and Francis. Research Monographs in Geographic Information Systems. Miller, H. and Han, J., Eds, 2001.
- [10] R. Agrawal and R. Srikant, "Mining sequential patterns," In Proc. of the 11th International Conference on Data Engineering, 1995.
- [11] G. Berger and A. Tuzhilin, "Discovering unexpected patterns in temporal data using temporal logic," Temporal Databases Research and Practice, Springer-Verlag, 1998.
- [12] H. Mannila and H. Toivonen, "Discovering generalized episodes using minimal occurrences," In Proc. of the Int'l Conference on Knowledge Discovery and Data Mining (KDD-96), pp.146-151, 1996.
- [13] R. Srikant and R. Agrawal, "Mining sequential patterns : generalizations and performance improvements," In Proc. of International Conference on Extending Database Technology, Avignon, France, Springer-Verlag, 1996.
- [14] Y. Cai, N. Cercone and J. Han, Attribute-Oriented Induction in Relational Databases, in G. Piatetsky-Shapiro and W. J. Frawley (eds.), Knowledge Discovery in Databases, AAAI/MIT Press, pp.213-228, 1991.
- [15] X. Lu, J. Han and B. C. Ooi, "Discovery of General Knowledge in Large Spatial Databases," In Proc. of Far East Workshop on Geographic Information Systems(FEGIS'93), pp.275-289, 1993.
- [16] R. E. Valdes-Perez, "Systematic Detection of Subtle Spatio-Temporal Patterns in Time-Lapse Imaging. I. Mitosis," Bioimaging. Vol.4 , No.4, pp.232-242, 1998.
- [17] M. N. Garofalakis, R. Rastogi, Shim K. S., "SPIRIT : Sequential Pattern Mining with Regular Expression Constraints," In Proc. of the international conference on VLDB, 1999.
- [18] M. Erwig and M. Schneider, "Spatio-Temporal Predicates," Technical Report 262, Fern University, 1999.
- [19] R. Snodgrass, "The Temporal Query Language TQuel," ACM TODS, Vol.12, No.2, June, 1987.
- [20] Jeong J. D., Paek O. H., Lee J. W. and Ryu K.H., "Temporal Pattern Mining of Moving Object for Location-Based Service," In Proc. of International Conference on Database and Expert Systems Applications (Dexa2002), (LNCS2453), 2002.
- [21] 안병익, "LBS기술동향과 전망 LBS 구조 및 구성", 한국지리정보, 10월호, pp.52-56, 2001.
- [22] 안윤애, 김동호, 류근호, "차량위치 추적을 위한 이동객체 관리 시스템의 설계, 정보처리학회논문지D, 제9-D권 제5호, 2002.
- [23] 이금우, 위치기반 서비스를 위한 개인화된 추천 시스템, 이학석사학위논문, 충북대학교, 2002.
- [24] 이준욱, 이용준, 류근호, "시간데이터마이닝 프레임워크", 정보처리학회논문지D, 제9-D권, 제3호, 2002.
- [25] 이준욱, 백옥현, 류근호, "위치기반 서비스를 위한 이동객체 시간 패턴 탐사 기법", 정보과학회논문지, 제29권 제5호, 2002.

이 준 욱

e-mail : ljw63822@etri.re.kr

1997년 충북대학교 컴퓨터과학과

1999년 충북대학교 대학원 전자계산학과
(이학석사)

2003년 충북대학교 대학원 전자계산학과
(이학박사)



2003년~현재 한국전자통신연구원 LBS 연구팀 PostDoc. 연구원
관심분야 : 시간 데이터베이스, 시공간 데이터베이스, 시간 데이터 마이닝, 시공간 데이터 마이닝, 위치기반 서비스, 텔레매틱스

남 광 우

e-mail : kwnam@etri.re.kr

1995년 충북대학교 전자계산학과

1997년 충북대학교 대학원 전자계산학과
(이학석사)

2001년 충북대학교 대학원 전자계산학과
(이학박사)



2001~현재 한국전자통신연구원 LBS 연구팀 선임연구원
관심분야 : 시간 데이터베이스, 위치기반서비스, 시공간 데이터베이스, Temporal GIS, 텔레매틱스, 이동통신 측위 시스템