

ISSN 1598-9798



# 데이터베이스연구

28권 제1호 2012년 4월

## 위치 기반 서비스 환경에서 이동 객체 궤적 데이터 마이닝 기법에 대한 고찰

A Survey on Moving Object Trajectory Mining Techniques  
in Location-based Services

장미영, 윤민, 장재우

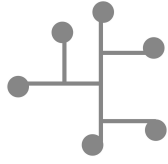
Miyoung Jang, Min Yoon, Jae-Woo Chang

데이터베이스 소사이어티  
Database Society

사단법인 한국정보과학회

The Korean Institute of Information Scientists and Engineers





# 위치 기반 서비스 환경에서 이동 객체 궤적 데이터 마이닝 기법에 대한 고찰

## A Survey on Moving Object Trajectory Mining Techniques in Location-based Services

장미영(Miyoung-Young Jang)<sup>1</sup>, 윤민(Min Yoon)<sup>2</sup>, 장재우(Jae-Woo Chang)<sup>3</sup>

### 요 약

이동 단말기의 확산과 GPS(Global Positioning System), 무선 통신 기술의 발달로 인하여 이동 위치 기반 서비스(LBS)가 더욱 중요한 사항으로 자리 잡고 있다. 따라서 대량의 시공간 데이터 베이스에서 이동 객체의 위치 정보는 이동 객체의 움직임 패턴을 분석하기 위해 널리 사용된다. 이동 객체의 패턴은 특정한 패턴을 지니고 있기 때문에, 궤적 데이터 마이닝을 통해 이동 객체의 궤적 데이터를 수집하여 분석함으로써 사람의 이동 특성(mobility)을 이해하고, 위치 데이터베이스에서 표면적으로 드러나지 않은 함축적인 지식 추출, 이동 객체 간의 패턴을 탐색할 수 있다. 아울러, 이러한 정보는 경로 예측, 영업 홍보, 도로 확장 등의 다양한 분야에서 유용하게 사용 될 수 있으며, 현재 궤적 데이터 마이닝을 통한 궤적 분석 기법들은 각각의 이용 목적에 따라 다양한 형태로 연구되고 있다. 본 논문에서는 기존 궤적 데이터 마이닝 기법을 궤적 데이터 클러스터링 기반 알고리즘, 궤적 데이터 분류 기반 알고리즘, 그리고 궤적 연관 규칙 기반 알고리즘으로 분류한다. 마지막으로, 기존의 궤적 데이터 마이닝 연구의 특징을 살펴보고, 이를 바탕으로 향후 연구 방향을 제시한다.

주제어: 이동 객체 궤적, 궤적 데이터 마이닝 알고리즘, 이동 객체 궤적 패턴

1 전북대학교 컴퓨터공학과, 박사과정

2 전북대학교 컴퓨터공학과, 박사과정

3 전북대학교 IT정보공학과, 교수

† 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2010-0023800)

+ 논문접수: 2012년 1월 4일, 심사완료: 2012년 3월 2일

## Abstract

Due to the development of geo-positioning capabilities and wireless communication technology, the Location-based Services(LBS) become an essential one. Therefore, the location information of mobile users in large spatio-temporal databases is widely used to analyze their movement patterns. This movement pattern provide us the key information to understand our daily life style and to discover useful knowledge about movement behaviour, which promotes novel applications and services such as the prediction of users movement and the navigation of the shortest path. Hence, various data mining techniques have been proposed for flexible and scalable analysis of massive spatio-temporal moving object trajectory data. In this paper, we classify the existing trajectory data mining techniques into three categories, i.e., clustering-based, classification-based and association-based mining techniques. Finally, we discuss the characteristics of state-of-art trajectory mining techniques and address future research directions based on the analysis.

Keywords: Moving object trajectory, Trajectory mining algorithm, Moving object trajectory pattern

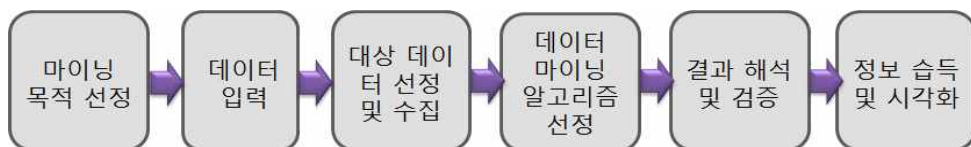
## 1. 서론

최근 이동 단말기의 확산과 GPS(Global Positioning System), 위성, WSN(Wireless Sensor Network), RFID(Radio Frequency Identification) 등 통신 기술의 발달로 인하여 이동 객체의 궤적 정보 획득 수단이 다양해지고, 실제로 방대한 양의 위치 데이터가 발생하고 있다. 일반적인 관점에서 이동 객체는 위치 측정이 가능한 Geo-센서를 장착한 야생 동물 및 스마트폰, PDA(Personal Digital Assistants) 등을 소지한 위치 기반 서비스 이용자, 그리고 GPS기반 내비게이션이 장착된 차량 등의 미하며, 실제로는 시간에 따라 움직이며 위치 데이터를 생산하는 대부분의 객체를 포함한다. 이동 객체의 움직임은 주요 시스템에서 자동으로 추적 가능하며, 이에 따라 방대한 양의 궤적 데이터를 저장하는 것이 가능하다. 이러한 이동 객체의 궤적 데이터는 이동 객체의 행동을 분석 및 예측하기 위한 정보로 사용될 수 있다. 따라서 궤적 데이터 마이닝을 통한 궤적 데이터의 분석에 대한 관심이 높아졌으며, 여러 가지 이동 객체 궤적 데이터 마이닝 기법이 연구되었다[1, 3, 5, 9, 10, 14, 18, 20, 21, 22].

일반적으로 데이터 마이닝은 데이터에 내재된 의미 있는 상관관계, 패턴, 경향 등을 탐색하는 일련의 과정으로써, 데이터베이스로부터 새로운 지식을 추출하는 전 과정을 포함하는 포괄적인 개념이다 [그림 1]. 그러나 이동 객체의 움직임, 즉 궤적 데이터는 궤적을 생성하는 이동 객체에 따라 특정한 패턴

을 지닌다. 즉, 야생 동물의 경우 먹이의 이동 경로나 기후 등의 변화에 따라 무리를 지어 움직이며, 사람은 일상생활에서 이동하는 경로의 범주가 집, 학교, 직장 등으로 제한되어 있다. 한편, 차량의 경우 도로를 따라 이동 가능한 영역을 움직이며, 다수의 차량이 통행하는 지역이 존재한다. 따라서 궤적 데이터 마이닝을 통해 이동 객체의 궤적 데이터를 수집하여 분석함으로써 사람의 이동 특성(mobility)을 이해하고, 위치 데이터베이스에서 표면적으로 드러나지 않은 지식 추출, 이동 객체 간의 패턴을 탐색할 수 있다. 이러한 정보는 이동 중인 사용자의 경로를 예측하여 최단 경로 안내 등의 서비스를 제공하는 위치 기반 서비스에 사용될 수 있다. 또한 사람들이 빈번하게 이동하는 경로 등을 분석하여 광고의 배치 선정이나 영업 홍보 등 다양한 비즈니스 분야에서도 유용하게 사용 될 수 있다.

기존 궤적 데이터 마이닝 알고리즘을 이용하여 현재 위치 기반 서비스에 필요한 지식을 추출하기 위해서는 기존 궤적 데이터 마이닝 알고리즘을 분석하고, 이를 응용에 적용하기 위한 연구가 필요하다. 특히 궤적 데이터 마이닝 기법은 추출한 지식을 적용하기 위한 응용에 따라 알고리즘의 수행 방법 및 목표가 달라지기 때문에, 이를 고려하여 기존 연구를 분석해야 한다. 따라서 본 연구에서는 기존 궤적 데이터 마이닝 알고리즘의 특징을 고려하여 크게 궤적 데이터 클러스터링 기반 알고리즘, 궤적 데이터 분류 기반 알고리즘, 그리고 궤적 연관 규칙 기반 알고리즘으로 분류하여 기술하고 각 기법의 특징을 서술



[그림 1] 데이터 마이닝 프로세스

한다. 또한, 기존의 연구를 비교 및 분석하여 향후 궤적 데이터 마이닝 연구가 나아가야 할 방향을 제시한다.

논문의 구성은 다음과 같다. 2장에서는 이동 객체 궤적 데이터의 특징과 세 가지 궤적 데이터 마이닝 알고리즘의 특징에 대해 설명하고, 이를 기반으로 궤적 데이터 마이닝이 적용 가능한 응용 분야에 대해 기술한다. 또한, 3장에서는 기존의 궤적 데이터 마이닝 기법을 클러스터링 기반 알고리즘, 궤적 데이터 분류 기반 알고리즘, 그리고 궤적 연관 규칙 기반 알고리즘으로 나누어 기술한다. 4장에서는 3장에서 서술한 연구들에 대한 비교 및 분석을 수행하고 5장에서는 궤적 데이터 마이닝 기법이 나아가야 할 향후 연구 방향을 제시한다. 마지막으로, 6장에서는 결론에 대해 서술한다.

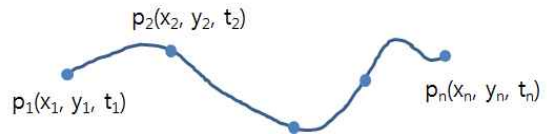
## 2. 이동 객체 궤적 데이터의 특징 및 분류

본 장에서는 이동 객체 궤적 데이터를 정의하고 궤적 데이터를 분석하는데 있어 적용되는 몇 가지 특징에 대해 서술한다. 아울러, 기존의 궤적 데이터 마이닝 기법을 세 가지 대표적인 궤적 분석 기법으로 기술하고, 마지막으로 궤적 데이터 마이닝을 통해 습득한 정보를 이용하는 대표적인 응용에 대해 살펴본다.

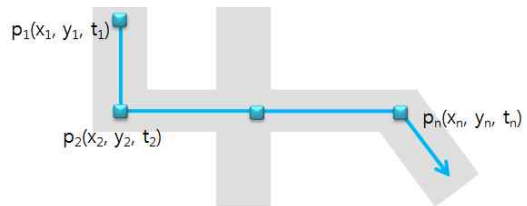
### 2.1 이동 객체 궤적 데이터

이동 객체의 궤적은 유클리디언 공간(Euclidean Space) 혹은 도로 네트워크(Road Network)로 정의되는 공간에서 이동 객체가 움직인 경로를 의미한다. 이동 객체 궤적  $Tr$  은 시간에 따라 정렬된 연속적인 점(point)들의 집합  $Tr = \{p_1(x_1, y_1, t_1),$

$p_2(x_2, y_2, t_2), \dots, p_n(x_n, y_n, t_n)\}$  으로 표현될 수 있으며,  $0 \leq i \leq n$  일 때, 각 점  $p_i$  는 GPS 등으로 측정된 데이터의 위치 좌표(예: 위도와 경도) 쌍과 시간 정보  $t_i$  를 지닌다. [그림 2]와 [그림 3]은 각각 유클리디언 공간 및 도로 네트워크를 이동하는 궤적 데이터를 나타낸다. 궤적 데이터 마이닝에서 궤적 데이터의 공간 정보, 즉 위치 정보만을 고려하여 수행하는 경우 공간 궤적 데이터 마이닝으로 정의된다. 또한 궤적 데이터의 공간 정보 및 시간을 고려하여 궤적 데이터마이닝을 수행하는 경우 시공간 궤적 데이터 마이닝으로 정의한다. 공간 궤적 데이터 마이닝에서 중요한 요소는 방대한 양의 공간 데이터를 분석하기 위한 공간 데이터 타입 정의와 공간 접근 순서 등을 파악하기 위한 공간 궤적 분석 알고리즘이다. 이러한 기법은 공간 클러스터링, 공간 특성 추출, 공간 연관 규칙 등의 기법을 적용하여 공간 데이터의 위상, 거리, 방향 관계를 바탕으로 데이터를 서술하고, 주어진 공간 객체가 이동할 때 비공간 속성의 변화를 측정하거나 공간 객체 간의 연관 관계를 모델링 하여 데이터 속성을 표현하기 위한 효율적인 알고리즘에 대한 연구를 중심으로 진행되었다. 한편, 시공간 데이터 마이닝은 초기에는 자연 과학 분야에서 지구과학 데이터를 이용한 연구를 중심으로



[그림 2] 유클리디언 공간을 이동하는 궤적 데이터



[그림 3] 도로 네트워크를 이동하는 궤적 데이터

발전하였으며, 이는 다양한 관찰 시스템에서 수집된 지구 기상 데이터로부터 데이터마이닝 기법을 이용하여 태풍, 지진, 기상변화, 화산 분출과 같은 활동을 예측하였다. 최근에는 이동 객체의 위치 데이터 혹은 궤적 데이터로부터 마이닝 기법을 적용하여 도로 교통 관제시스템, 선박, 항공 등의 이동 객체 추적 시스템 등에서 이동 객체의 궤적 분류, 궤적 예측 등을 수행하는 연구로 발전하였다.

2.2 궤적 분석 방법에 따른 마이닝 기법

가. 궤적 데이터 클러스터링(Trajectory Data Clustering)기반 알고리즘

궤적 데이터 클러스터링 기반 데이터 마이닝 알고리즘은 현재 궤적 데이터 분석에 가장 많이 적용되어 사용되는 방법으로써, 이미 알려진 클래스 레이블을 참고하지 않고, [그림 4(a)]와 같이 궤적 데이터의 각 지리적 인접성 등의 특징을 분석하여 궤적 데이터를 그룹화 하는 알고리즘이다. 따라서 각 클러스터 내 궤적 데이터의 유사성이 극대화되고, 서로 다른 클러스터 간의 유사성은 최소화 되도록 궤적이 분류된다. 이러한 기법은 궤적 데이터의 기본 지식이나 일정한 데이터 특성을 반영하지 않고도 궤적 데이터를 분류 할 수 있다는 큰 장점을 지닌다.

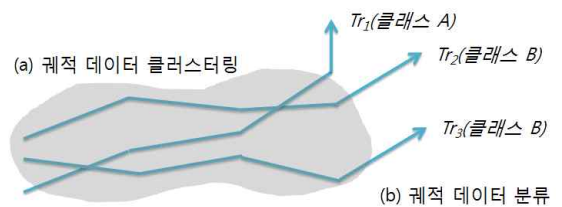
나. 궤적 데이터 분류(Trajectory Data Classification)기반 알고리즘

궤적 데이터 분류 기반 데이터 마이닝 알고리즘은 각 궤적을 분류하기 위한 분류 체계 및 클래스 레이블을 기반으로 궤적 데이터를 기술한다. 또한, 새로 추가되는 궤적 데이터에 대해 이를 분류 시스템에 적용하기 위한 구분 모형(혹은 함수)를 찾는 알고리즘이다. 따라서 이러한 기법은 궤적 데이터의 위치 정보만이 아닌 비공간 데이터인 속성 값을 기반으로

하는 하나의 클래스를 부여하고, 궤적 데이터를 이를 기준으로 분류하는 기법이다[그림 4(b)]. 예를 들어 전체 궤적을 요일 및 시간을 기준으로 분류하고 이에 대한 특징을 추출하는 알고리즘은 시간(요일)이라는 궤적의 클래스를 정의하고 이를 기반으로 궤적 데이터를 분류 및 분석하는 기법이다.

다. 궤적 연관 규칙 (Trajectory Association Rule)기반 알고리즘

궤적 연관 규칙 기반 데이터 마이닝 알고리즘은 수치적으로 관련성(동시 발생 빈도)이 높은 지점(혹은 영역) 정보를 연관 (Association) 관계로 정의하고, 이에 대한 빈도 및 관련성을 탐색하는 알고리즘이다. 이러한 알고리즘은 [그림 5]와 같이 궤적이 지나는 모든 지점(혹은 영역)을 확인하여 그에 대한 빈도 등을 분석하고, 각 지점의 연관 관계 및 발생 확률 등을 측정하여 예측하는 시스템에서 사용된다. 특히 이동 객체 궤적에서는 궤적의 출발-도착 (Origin-Destination) 지점 쌍의 연관을 바탕으로 수행되는 알고리즘이 많이 연구되었다. 하나 또는 그 이상의 공간 데이터가 다른 객체들과 지나는 연



[그림 4] 궤적 데이터 클러스터링 및 분류 기법



[그림 5] 궤적 연관 규칙

관 관계를 규칙화 하는 알고리즘 또한 이러한 분류에 속한다.

### 2.3 이동 객체 데이터 마이닝 응용 분야

이동 객체 궤적 데이터 마이닝의 가장 큰 장점은 데이터 마이닝을 통해 습득한 정보를 이용하여 대용량의 이동 객체 궤적 데이터를 효과적으로 관리함으로써 이동 객체 궤적 정보에 대한 접근 및 활용에 따른 비용과 시간을 크게 감소할 수 있다는 것이다. 본 절에서는 이동 객체 궤적 데이터 마이닝을 이용한 대표적인 응용 분야에 대해 소개한다.

**지능형 교통정보 시스템** : 도로 네트워크 공간에서의 이동 객체별 궤적 정보를 분석함으로써 교통사고가 빈번한 위치 분석 및 예측, 도시의 대중교통 노선 증/감축 등의 설계, 증강 현실 시스템 기반 응용 서비스 개발, 위치 기반 광고 서비스 제공에 적용할 수 있다[28]. 또한, 자주 정체되는 주요 도로 구간을 지나는 궤적 데이터를 분석하여 이를 우회하는 노선을 추가하거나 우회 도로를 건설함으로써 교통량의 분산을 유도하는 등 차량 흐름 관리를 수행할 수 있다. 아울러, 새로운 도시를 개발하는 경우 이동 인구의 수나 지형 특징 등이 비슷한 지역의 궤적 정보를 분석하여 도로 설비 계획을 수립할 수 있다. 이러한 시스템에서는 시간 및 공간상으로 유사한 궤적을 찾기 위한 시공간 유사도 측정 알고리즘 및 출발지 목적지 간의 연관성을 추출하기 위한 알고리즘 [18,20,21,22] 등이 적용된다.

**LBS 및 유비쿼터스 환경에서 위치 정보 활용** : 유비쿼터스 기반 지리정보 시스템 및 처리 기술은 대규모의 위치 데이터를 기반으로 데이터베이스를 구축하여 사용자의 개입을 최소화하면서 사용자의 상

황에 적절한 위치 정보 및 유비쿼터스 서비스를 실시간으로 제공해주는 기술을 의미한다. 예를 들어 궤적 데이터의 상황인식 모바일 기기를 이용하여 자신의 위치를 기반으로 제공되는 여행가이드, 캠퍼스 안내 등의 시스템[29-31]이 존재한다. 현재 이러한 서비스를 완벽하게 지원하고 처리하는 시스템은 거의 존재하지 않는 실정이며, 향후 추가적인 기술 개발이 요구된다. 이러한 연구를 통해 유비쿼터스 환경에서 부가적인 U-city 서비스를 창출할 수 있다.

**비즈니스 지리정보시스템(GIS) 설계** : 위치 데이터 이동 궤적의 분석을 통해 수집된 데이터는 마케팅이나 기업전략 수립 분야에서 비즈니스 지리정보 시스템을 위한 응용에서 사용 가능하다[32]. 궤적 데이터의 지역 근접성 기반 클러스터링을 통해 분석한 주요 이동 경로 및 이동 객체의 출발-도착 지점의 연관성 분석 알고리즘을 통해 습득한 지식은 기업의 의사 결정을 위한 매우 중요한 정보로서 지리공간상에서 상권이라고 하는 특정 범위에 대한 통계 정보의 다른 형태로 제공될 수 있으며, 기업이 보다 효율적으로 다양한 LBS를 제공하는 새로운 비즈니스 전략을 제시하는데 도움을 줄 수 있다.

**위성을 이용한 선박 추적 시스템** : 선박 추적 시스템에서는 위성으로 측정된 사진이나 좌표정보를 기반으로 선박의 위치를 탐색 및 저장하고 이를 분류하는 기법을 통칭하며 어장 관리, 해양 오염 경로 예측, 밀입국 및 밀수 등을 관리하기 위한 국경 관리 시스템 등에 사용된다. 이러한 시스템은 해로를 따라 이동하는 선박의 궤적 정보를 수집하여 일반적인 이동 경로를 분석하고, 이동의 목적 및 여러 조건에 따라 궤적 데이터를 분류하여 적용한다. 예를 들어 국경 관리 시스템의 경우, 관측된 선박의 경로가 궤



적 데이터 마이닝을 통해 분석한 기존의 궤적 패턴과 유사하지 않은 경우, 이를 Outlier로 분류하고 면밀히 검토하여 밀입국이나 테러 조직 등의 불법 이동 경로인지 확인 할 수 있다[33]. 또한 궤적 데이터 마이닝을 통해 해양 사고가 다수 발생하는 지점 등을 미리 분석하여 선박 조난 시 구조대를 파급할 지역을 선정하는 등의 의사 결정에도 큰 도움을 줄 수 있다.

### 3. 이동 객체 궤적 데이터 마이닝 기법

본 장에서는 2.2절의 분류 기준을 바탕으로 기존의 궤적 데이터 마이닝 기법을 궤적 데이터 클러스터링 기반 알고리즘, 궤적 데이터 분류 기반 알고리즘 그리고 궤적 데이터 연관 규칙 기반 알고리즘으로 분류한다. 아울러, 각 분류에 해당하는 대표적인 궤적 데이터 마이닝 알고리즘에 대해 기술한다.

#### 3.1 궤적 데이터 클러스터링 기반 알고리즘

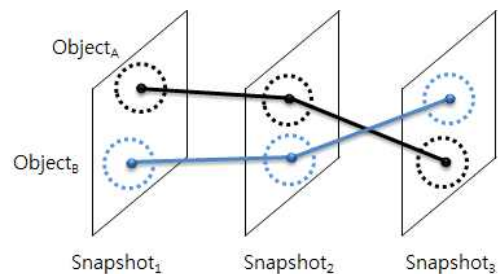
본 절에서는 궤적 데이터를 그룹화 하는 클러스터링 기법을 바탕으로 궤적 데이터 마이닝을 수행하는 기법에 대해 기술한다. 궤적 데이터 클러스터링 기반의 대표적인 마이닝 기법은 J. Yang 과 M. Hu의 연구[1], G. Gidofalvi와 T. Pedersen의 연구[3], S. Qiao et al.의 연구[5] 및 F. Giannotti et al.의 연구[9]가 존재한다.

#### 가. J. Yang 과 M. Hu의 연구[1]

J. Yang 과 M. Hu의 연구에서는 현재 모바일 등의 위치 측정 기기의 자원 한계로 인해 위치 데이터의 정확성이 떨어지며, 따라서 수집된 궤적을 기존의 데이터 마이닝 기법 및 패턴 추출 기법에 적용하

기 어려운 문제점을 지적하였다. 아울러, 기존 궤적 데이터 마이닝 기법은 이동 객체의 속도를 고려하지 않기 때문에 데이터를 정확히 예측하고 패턴을 추출하는데 있어 어려움이 존재함을 인식하였다. 이를 해결하기 위해, 이동 객체의 속도를 고려하여 궤적을 효율적으로 예측하고, 이를 통해 클러스터링 기반의 패턴 추출 알고리즘을 제안하였다. 제안된 TrajPattern 알고리즘은 Normalized Match (NM) 유사도 탐색 기법을 통해 k개의 패턴을 clustering 하는 기법으로, 크게 3단계로 구성된다. 첫째, snapshot의 위치 정보를 [그림 6]과 같은 연속된 궤적으로 정의한다.

아울러, 궤적들의 효율적인 클러스터링을 수행하기 위해 네트워크를 grid 형식으로 표현하고, 이를 통해 각 궤적들을 NM 기법을 이용하여 클러스터링을 수행한다. 클러스터링이 완료되면, 각 클러스터에서는 NM threshold(k 번째까지의 최대 NM값)을 설정하여, 임계값을 기준으로 low 클러스터와 high 클러스터로 분류한다. 둘째, 길이가 가장 작은 패턴부터 길이가 큰 패턴까지 순차적으로 접근하는 방식을 이용하여 클러스터 내 패턴을 추출한다. 가장 길이가 짧은 패턴을 시작으로 패턴을 snapshot 단위 및 시간을 고려하여, 비슷한 궤적을 갖는 패턴을 통합한다. 마지막으로, 각 클러스터에서 선정된 k개의 패턴을 사용자에게 전송한다. 해당 기법은 실제 네



[그림 6] TrajPattern의 궤적 표현

트위크 환경 및 이동 객체의 속도를 고려하여 마이닝을 수행하기 때문에, 마이닝 결과의 정확도가 높은 장점을 지닌다. 또한, NM 기법을 통한 궤적 필터링을 수행하기 때문에, 다른 기법에 비해 상대적으로 낮은 연산 비용을 제공하는 특징을 지닌다.

#### 나. G. Gidofalvi와 T. Pedersen의 연구[3]

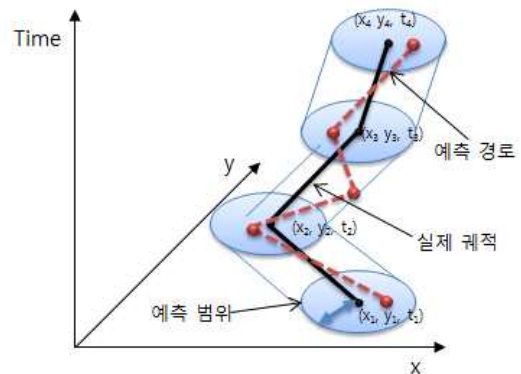
본 연구에서는 빈번하게 발생하는 공유(Sharable) 궤적과 출발지와 목적지를 포함하는 긴 궤적(Long Trajectory)를 모두 고려한 이동 객체 궤적 데이터 마이닝 알고리즘을 제안하였다. 먼저, 공유 궤적의 패턴추출을 위해 지역 마이닝(local mining) 기법을 제안하였다. 이는 빈도 임계값(frequent threshold)과 길이 임계값(length threshold)를 기준으로 궤적 데이터 필터링을 수행하고, 각 궤적을 기준으로 SMFCI(Single most frequent closed itemset)를 생성하여, 유사한 궤적 데이터를 클러스터링하여 패턴을 추출한다. 다음으로, 긴 궤적의 패턴 추출을 위해 전역 LSP 마이닝(global LSP mining) 기법을 제안하였다. 이는 기존의 지역 마이닝 기법에 궤적 간의 거리 개념을 추가하여 패턴을 추출하는 기법이다. 마지막으로, 이 두 가지의 궤적을 동시에 다룰 수 있는 hybrid LSP 기법을 제안하였다. 이는 먼저 전역 LSP를 통해 출발지-목적지에 상응하는 궤적을 찾은 후, 지역 LSP 기법을 통해 유사한 궤적을 통합하여 패턴을 추출하는 방법이다. 아울러, 이를 실제 시스템에 효율적으로 적용시키기 위해 SQL 구문의 질의를 제공한다. 해당 기법은 필터링 기반 지역 LSP 및 거리 기반 전역 LSP 기법을 적용하여 최소한의 연산 비용으로 높은 정확도를 지닌 패턴을 추출할 수 있는 장점을 지닌다. 또한, 궤적 데이터 마이닝을 위한 SQL 질의를 제공함으로써 다양한 응용에 쉽게 적용할 수 있

다.

#### 다. S. Qiao et al.의 연구[5]

S. Qiao et al.의 연구에서는 GPS에서 수집되는 불확실한 데이터에 대해 효율적인 클러스터 생성 및 궤적 예측을 수행할 수 있는 알고리즘을 제안하였다. 이를 위해, 실제 응용에서 수집되는 불확실한 궤적들을 uncertain area라는 특정 포인트 영역으로 설정하여 유사 궤적을 하나의 궤적으로 설정한다 [그림 7].

제안하는 기법에서는 설정된 궤적을 이용하여 클러스터를 생성하고, 이를 통해 궤적 예측을 수행한다. 먼저, 클러스터 생성 단계에서는 궤적 간 포인트의 거리를 비교하여 유사한 궤적을 하나의 클러스터로 구성한다. 클러스터 생성은 다음의 세 단계를 통해 수행된다. 첫째, 궤적 데이터베이스에서 무작위로 기준 궤적을 선정하고, 인접 궤적들과의 거리를 계산한다. 둘째, 만약 거리 임계값 이내에 위치한 인접 궤적의 수가 클러스터 생성 최소 궤적의 수보다 크면, 이를 하나의 클러스터로 구성한다. 셋째, 모든 궤적에 대해 클러스터를 생성하고, 어떠한 클러스터에도 포함되지 않는 궤적은 제거한다. 한편, 궤적 예측 단계에서는 앞서 구성된 클러스터를 이용



[그림 7] Uncertain Area

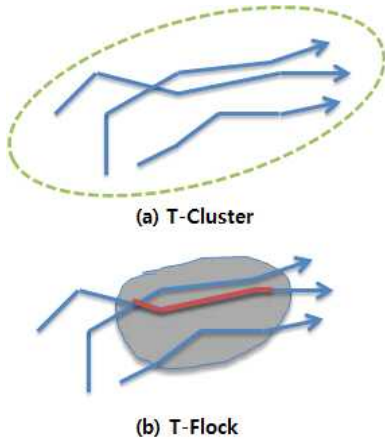
하여 속도-방향 기반 궤적 예측을 수행한다. 즉, 노드(예: 교차로)의 근접한 이동 객체의 속도를 측정하여 어느 방향으로 객체가 진행할 것인지 예측한다. 기존의 빈도 기반 가중치 궤적 예측 기법은 실제 이동 객체의 방향이나 속도 등의 요소를 고려하지 않고 기존 궤적 데이터의 발생 통계에 의한 가중치만을 고려하여 예측하기 때문에, 정확도가 감소하여 이를 실제 응용에 적용하는 것은 적합하지 않다. 이를 기반으로 제안된 궤적 예측 기법에서는 기존 궤적 데이터를 분석하여 각 속도별 방향성 테이블을 생성한다. 따라서 이동 궤적 예측 질의 수행 시, 일정한 노드에 진입하는 이동 객체의 속도를 측정하고, 이를 방향성 테이블과 비교하여 다음 이동 방향 및 속도를 예측한다. 예를 들어, 만약 어떤 자동차가 교차를 진입할 때 빠른 속도로 이동하고 있다면 직진 운동을 할 가능성이 크다. 하지만, 감속 운동을 하는 경우 좌측이나 우측으로 이동할 확률이 높다. 따라서 이러한 환경을 고려하여 방향성 테이블을 선택하고 이동 확률이 높은 방향을 다음 궤적의 이동 방향으로 예측한다. 해당 기법은 모든 궤적들의 포인트를 비교하여 클러스터링을 수행하기 때문에, 사용자에게 높은 정확도를 지니는 패턴을 제공할 수 있다. 아울러, 경로 예측 시 기존 궤적의 통계에 의한 가중치가 아닌 실제 객체의 이동 방향 및 속도를 고려하기 때문에, 사용자에게 높은 경로 예측 정확도를 제공한다. 그러나 클러스터 수행 시 모든 궤적들에 대해 비교 연산을 수행하기 때문에, 높은 연산 비용을 가지는 문제점이 존재한다.

라. F. Giannotti et al.의 연구[9]

F. Giannotti et al.의 연구에서는 대량의 이동 객체 궤적 데이터로부터 마이닝을 수행하는 과정을 Complex Knowledge Discovery Process라 정의하

고, 실제 이탈리아 밀라노(Milan)와 피사(Pisa)에서 수집한 데이터를 이용하여 마이닝을 수행하였다. 또한, 마이닝 수행을 통해 생성된 패턴 등의 결과를 시각화하여 궤적의 밀도 및 시간에 따른 분포가 궤적 데이터 마이닝 결과에 어떠한 영향을 미치는지 보였다. 아울러 이동 객체 궤적으로부터 수집된 정보를 이용하여 대표적으로 수행되는 질의(예: 이동 객체가 빈번히 이동하는 경로 탐색, 사람들의 방문이 빈번한 영역을 지나는 도로에서의 시간 효율적인 이동 경로 탐색, 미래 교통량 예측)에 대한 궤적 데이터 마이닝 일련의 과정을 제시하였다.

아울러 궤적을 효율적으로 처리하기 위해 M-Atlas라는 질의 및 마이닝 언어 구성 시스템을 정의하였다. 이를 위해 궤적데이터 시간별 이동량, 전체 궤적의 길이, 이동 시간 등의 통계정보를 생성하고, 질의에 대한 임계값을 설정하여 이를 만족하는 패턴을 추출하였다. 이때, M-Atlas 수행되는 마이닝은 크게 첫째, 궤적의 전체 이동 경로를 분석하여 질의를 수행함으로써 시간 및 요일 교통량에 따라 궤적에 대한 클러스터를 생성하고 질의에 적용한 경우와 둘째, 궤적의 출발, 도착지점에 대한 빈도 및 이동 경로의 다양성에 대해 고려한 접근 방식으로 분류된다. 클러스터 기반의 마이닝을 수행하기 위해 본 연구에서는 전체 궤적에 대한 통계를 바탕으로 T-Cluster, T-Flock을 생성한다. [그림 8]와 같이, T-Cluster는 전체 궤적의 지리적 인접성을 이용하여 생성하고, T-Flock의 경우 구간에 대한 궤적의 지리적 인접성을 고려하여 생성한다. 따라서 전체 궤적에 대한 특징을 분석하고, 이를 바탕으로 응용에 따른 질의를 수행하는 기법이다. 따라서 시간 기준을 가지고 궤적을 클러스터링 한 경우, 대표적 질의 중 시간대 별로 이동 빈도가 높은 경로를 탐



[그림 8] 궤적 타입 정의

색하는 질의를 수행할 수 있으며, 궤적 중 출현 빈도가 높은 영역을 분석하는 경우에는 사람들이 해당 영역을 접근할 때 주로 이용하는 도로를 찾고 이를 우회하는 도로를 안내하는 응용 등에 적용할 수 있다.

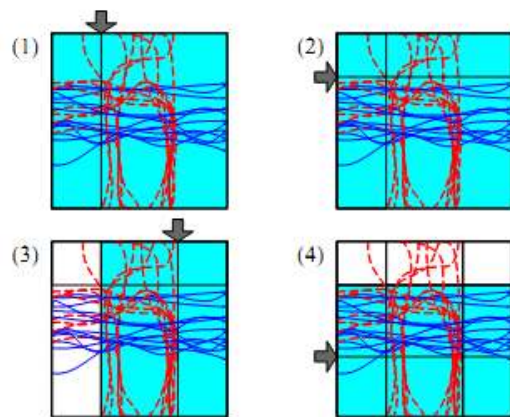
3.2 궤적 데이터 분류 기반 알고리즘

본 절에서는 새로 추가되는 궤적 데이터에 대해 정해진 분류 기법을 기반으로 마이닝을 수행하는 기법에 대해 기술한다. 특히 특정 공유지나 목적지를 기반으로 분류한 궤적에 대해서 패턴을 추출하는 기법에 대해 J. G. Lee 그룹에서 제안한 두 가지 연구 [10,14]에 대해 서술한다.

가. J. G. Lee et al.의 연구[10]

J. G. Lee et al.의 연구에서는 지역 기반클러스터링(Region-based clustering) 기법과 궤적 기반클러스터링(Trajectory-based clustering) 기법을 계층적으로 수행하여 효율적으로 패턴을 분류하는 데이터 마이닝 기법을 제안하였다. 영역 기반 클러

스터링 (Region-based clustering) 단계에서는 궤적들의 동질성을 판별하기 위해, 출발지 혹은 경유지가 비슷한 궤적들의 분류를 수행하며, 궤적 기반클러스터링 (Trajectory-based clustering) 단계에서는 유사 궤적 중 대표적으로 표현될 수 있는 중요한 궤적의 분류를 수행한다. 첫째, 영역 기반 클러스터링은 객체의 궤적을 minimum description length(MDL)[11] 기법을 적용하여, 출발지 혹은 경유지가 다른 궤적이 포함되지 않는 부분을 하나의 region으로 분리하는 기법이다. MDL 기법이란 출발지 혹은 경유지를 기준으로 하여 다른 궤적이 포함되지 않는 부분을 분리하기 위해 수행되는 최소한의 연산 비용을 구하는 기법이다. MDL은 사용자가 궤적을 분류하기 위해 Homogeneity property 혹은 Conciseness property의 정도에 따라 범위를 설정하고, 이에 따라 모든 궤적들은 출발지 혹은 경유지를 기준으로 하나의 Region으로 분류된다. [그림 9]는 영역 기반 클러스터링 기법의 수행 단계를 나타낸다. 그림과 같이 실선과 점선의 출발지 혹은 목적지는 서로 다르기 때문에 이를 MDL을 이용하여 각 region으로 분류한다. 둘째, Trajectory-based 클러스터링은 partition-and-group 기법을 적용



[그림 9] Region-based clustering[10]

하여, Region-based 클러스터링에서 생성된 클러스터링에서 discriminative power가 높은 궤적만을 추출하여 클러스터링을 수행한다. 즉, 기 분류된 궤적들 중에서 대표적으로 표현될 수 있는 중요한 궤적을 판단하고 이를 grouping 한다. 해당 궤적들은 Class-Conscious Grouping 기법[12,13]을 사용하여, 궤적들과의 관계성을 파악한다. Conscious Grouping 기법은 각 궤적마다 distance를 계산하여, 지역적으로 유사한 궤적들 중 distance cost (Homogeneity of an Neighborhood)가 가장 적게 발생하는 궤적을 선택하여 이를 대표 궤적으로 설정한다. 따라서 지역적으로 유사한 궤적을 우선적으로 클러스터링을 수행하고, 클러스터링 내에서 가장 대표성을 띄는 궤적을 추출하여 분류함으로써, 다른 기법에 비해 효율적으로 궤적들을 분류할 수 있다. 해당 기법은 지역적으로 유사한 궤적을 우선적으로 클러스터링하고, 클러스터 내에서 가장 대표성을 띄는 궤적을 추출하기 때문에, 클러스터를 기반으로 유사도가 높은 패턴을 추출할 수 있는 장점을 지닌다. 하지만, 모든 궤적들에 대하여 비교 연산을 수행하기 때문에, 상대적으로 많은 연산 비용을 요구한다.

나. J. G. Lee et al.의 연구[14]

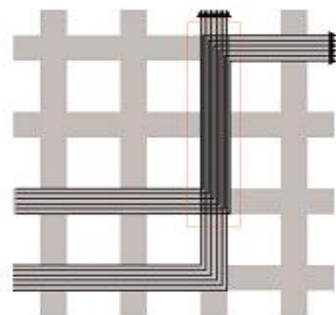
J. G. Lee et al.의 연구에서는 기존의 궤적 데이터 기법에서 사용되는 궤적들의 길이가 길 경우, 패턴 추출 연산 비용이 큰 단점을 해결하기 위해, 전체 궤적을 부분 연속 궤적(partial sequence trajectory)으로 분할하여 분류를 수행하는 궤적 데이터 마이닝 알고리즘을 제안하였다. 제안하는 알고리즘은 궤적 특징 추출 단계 (feature generation), 궤적 특징 선택 단계 (feature selection) 그리고 분류 모델 생성 단계

(classification model construction)로 이루어진다. 첫째, 궤적 특징 추출 단계에서는 궤적들을 Information gain[15]에 따라 분류하고, 최소 발생 빈도 임계값을 통해 분류된 궤적들 중 의미 없는 궤적들을 필터링한다. Information gain은 특정 point를 기준으로 각 궤적들의 엔트로피를 비교하여 유사성을 찾는 지표로 사용되며, (식 1)과 같다.

$$\begin{aligned}
 {}_b C(X)_1 & \left. \begin{matrix} (c|x)=1 \wedge P(c_j|x)_{j \neq i} = 0 \\ P(c_i) - P(x) \\ P(\bar{x}) \end{matrix} \right\} \log \frac{P(c_i) - P(x)}{P(\bar{x})} + \sum_{j \neq i} \left. \begin{matrix} P(c_j) \\ P(\bar{x}) \end{matrix} \right\} \log \frac{P(c_j)}{P(\bar{x})} \\
 & = - (P(c_i) - \theta) \log \frac{P(c_i) - \theta}{1 - \theta} - \sum_{j \neq i} P(c_j) \log \frac{P(c_j)}{1 - \theta}.
 \end{aligned}$$

(식 1)

이 때, x는 궤적의 패턴을 나타내며, ci는 궤적의 특정 포인트를 나타낸다. 아울러, 최소 발생 빈도 임계값은 생성되는 패턴을 의미 있는 데이터로 선택하기 위한 최소한의 발생 빈도를 나타낸다. 둘째, 궤적 특징 선택 단계에서는, F-score 기법을 통해 [그림 10]과 같이 모든 궤적들이 통과하는 도로(예 : 다리, 고가 도로) 같은 궤적을 필터링한다. F-score란 정보 검색 및 문서 분류에 사용되는 기법으로써, 특정 포인트가 존재할 수 있는 확률을 계산하는 기법이다. 따라서 모든 궤적들의 F-score



[그림 10] 궤적 필터링 [14]

값을 계산한 후 F-score가 낮은 패턴 즉, 모든 궤적들이 통과하는 의미 없는 궤적들을 필터링한다. 마지막으로, 분류 모델 생성 단계에서는 선택된 궤적들을 SVM(Support Vector Machine) 분류 알고리즘을 통해 2차 분류를 수행한다. SVM은 벡터 기반의 분류 기법이기에 때문에, 제안된 알고리즘에서는 선택된 궤적들을 벡터로 변환하여, SVM에 적용한다. 해당 기법은 모든 궤적을 부분 연속 궤적으로 분할하여 마이닝을 수행하기 때문에, 기존 기법이 가지는 문제점, 즉, 마이닝 수행을 위한 궤적들의 길이가 길 경우, 마이닝 연산 cost가 큰 문제점을 해결할 수 있다. 또한, information gain, F-score 필터링 기법을 이용하여 유사하지 않은 궤적들에 대해 필터링을 수행하기 때문에, 마이닝 수행 시 적은 연산 비용을 제공하는 장점을 지닌다. 반면, 모든 궤적을 부분 연속 궤적으로 분할하기 때문에, 전체 궤적의 패턴을 원하는 응용, 즉, 출퇴근 경로 분석, 도심 인구 이동 분석 등에는 적용이 불가능한 문제점이 존재한다.

### 3.3 궤적 연관 규칙 기반 알고리즘

본 절에서는 출발-목적지 쌍에 대한 빈도등의 특징을 기반으로 패턴 추출 및 궤적 예측 응용을 수행하는 궤적 연관 규칙 기반 마이닝 알고리즘에 대해 설명한다. 대표적인 연구로는 F. Giannotti et al.의 연구[18], A. Monreale et al.의 연구[20], A. Lee et al.의 연구[21], L. Chen et al.의 연구[22] 마지막으로 F. Giannotti et al.의 연구 [9]가 존재한다.

#### 가. F. Giannotti et al.의 연구[18]

F. Giannotti et al.의 연구에서는 궤적 패턴(T-pattern) 추출을 위해 기존의 연속적인 edge point를 이용하여 표현된 궤적들과 달리

ROI(Region of Interest)[19] 영역을 생성하여 궤적을 표현하고, 패턴을 추출한다. 제안하는 패턴 추출 알고리즘은 ROI 생성 단계, 패턴 추출 단계로 수행된다. 첫째, ROI 생성 단계에서는 궤적들을 분석하여 ROI를 생성한다. 이 때, ROI란 단순한 포인트로 이루어진 집합이 아닌 궤적들 중 의미 있는 지역을 선정하여 생성한 영역을 의미한다. 예를 들어, tourist들은 특정 도시를 여행하고자 할 때, 정해진 명소만을 관광할 뿐만 아니라 도보로 그곳의 주변 지역까지 관광을 한다. 또한 다른 명소를 관광하기 위해 차를 통해 그 곳으로부터 멀리 떨어진 곳으로 이동한다. 이 때 이를 궤적으로 살펴보면 특정 명소와 주변 지역의 명소의 궤적을 하나의 셋으로 표현할 수 있으며, 이를 하나의 boundary 표현하는 것이 ROI가 된다. ROI를 구성하는 방법으로는 static-ROI 와 dynamic-ROI 방법이 존재한다. static-ROI는 T-pattern을 구성하기 위한 전처리 기법으로써, 일반적으로 사람들에게 잘 알려진 장소(예 : 관광지, 유물, 명소) 혹은 객체들의 궤적을 분석하여 궤적들의 밀도가 가장 높은 지점을 기준으로 ROI를 생성하고 이를 확장하는 기법이다. ROI는 기준점이 되는 포인트를 중심으로 거리 임계값을 포함하는 모든 POI에 대해 모든 방향으로 확장하며, 만약 ROI 주변에 frequent가 높은 궤적이 존재할 경우 ROI를 확장한다. 한편, dynamic-ROI는 static-ROI의 단점을 보완하기 위해 제안된 기법으로써, 패턴 추출에 필요한 짧은 궤적들만을 이용하여 ROI를 생성한다. static-ROI는 모든 궤적들에 대해 밀도를 계산하고 ROI를 생성, 확장하기 때문에 많은 연산비용이 발생하는 반면, dynamic-ROI는 궤적들 중 길이가 1인 궤적들을 선별하여 이를 기준으로 ROI를 생성하고, 해당 궤적들 중 연속적인 포인트를 지닌 궤적들에 한해 ROI를 생성하기 때문에,

모든 궤적들에 대해 계산할 필요가 없다. dynamic-ROI는 static-ROI에 비해 궤적의 연산량이 적고, 필요한 궤적에 대해서만 ROI를 생성하기 때문에 효율적이다.

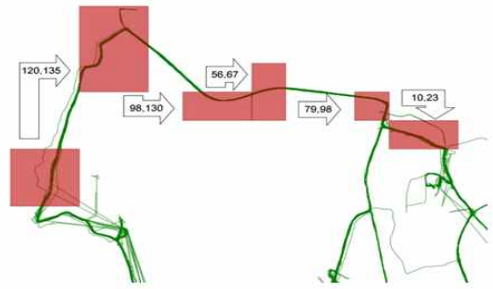
둘째, 패턴 추출 단계에서는 ROI 생성 단계에서 생성된 ROI를 바탕으로 패턴을 추출한다. 모든 궤적에 대해 ROI가 생성되면, 기존의 궤적을 기반으로 이와 유사한 경로를 갖도록 ROI를 연결한다. 즉, 기존 궤적과 유사한 ROI 궤적을 생성하고 이를 바탕으로 식 2와 같이 T-pattern을 추출한다. 이때 A와 B는 ROI를,  $\Delta t$ 는 ROI에서 ROI로 이동하는데 걸리는 시간을 의미한다.

$$\Delta \rightarrow B \xrightarrow{\Delta t_2} B \text{ -----식(2)}$$

[그림 11] 은 실제 도로 네트워크에서의 T-pattern를 나타낸다. 각각의 rectangle은 ROI를 의미하며, 화살표는 ROI에서 ROI로 이동하는 이동성, 화살표 안의 숫자는 각각 ROI를 나가는 시간 및 ROI 들어가는 시간을 나타낸다. 해당 기법은 패턴 추출을 위한 기준을 특정 도로 네트워크의 노드가 아닌 ROI 영역을 기준으로 마이닝을 수행하기 때문에, 여행 경로 추천, 도시 간 인구 이동 등의 응용에 필요한 패턴을 제공할 수 있다. 하지만, ROI를 기반으로 하는 궤적을 재구성할 때, 모든 궤적들에 대해 비교 연산 및 재구성을 수행하기 때문에, 연산 비용이 높은 단점이 존재한다.

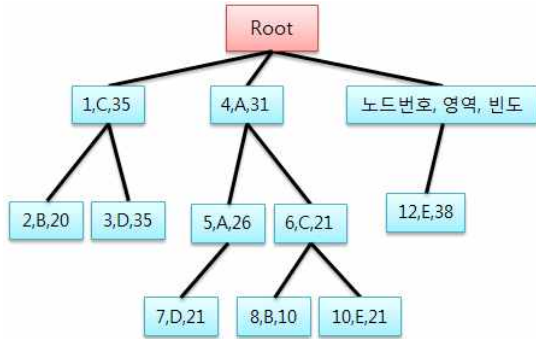
나. A. Monreale et al.의 연구[20]

A. Monreale et al.의 연구에서는 모바일 기기 배터리의 한계로 인해 빈번한 위치 정보 전송이 어려운 문제점을 해결하기 위해 궤적 예측을 수행하는 알고리즘을 제안하였다. 이는 F.Giannotti[18]에서



[그림 11] T-pattern[18]

제안된 T-pattern을 이용하여 T-Pattern tree를 생성하고 이를 이용하여 궤적을 예측하는 알고리즘이며, ROI 생성, T-pattern 생성, T-pattern Tree 생성, 궤적 예측 단계로 구성된다. 첫째, ROI 생성 및 궤적 데이터를 이에 맵핑(mapping)하여 의미 있는 영역의 연속으로 표현하는 작업을 수행한다. 둘째, T-pattern 생성 단계에서는 ROI 생성 단계에서 생성된 ROI를 바탕으로 궤적과 동일하게 이동하는 ROI를 연결하여 T-pattern을 추출한다. 이때 생성되는 T-pattern은 F. Giannotti et al.의 연구[18]에서 정의한 내용과 동일하다. 셋째, T-pattern Tree 생성 단계에서는 추출된 T-pattern들을 이용하여 접두어 트리 형태의 T-패턴 트리를 구축한다. T-pattern tree는 [그림 12]와 같이 표현되며, 이때 트리의 각 노드는 노드 id, ROI id, ROI 간 이동 시간의 정보 {node id, ROI id, t}, 즉, 이동 객체가 빈번하게 방문한 영역들을 나타내고 간선은 영역들 간의 이동을 나타낸다. 마지막으로, 궤적 예측 단계에서는 생성된 T-pattern을 통해 새로운 객체의 궤적과 가장 잘 매치되는 경로를 탐색하고, punctual score기법을 이용하여 점수가 높은 궤적을 예측 궤적으로 검색하여 반환한다. punctual score란 실제



[그림 12] T-패턴 트리

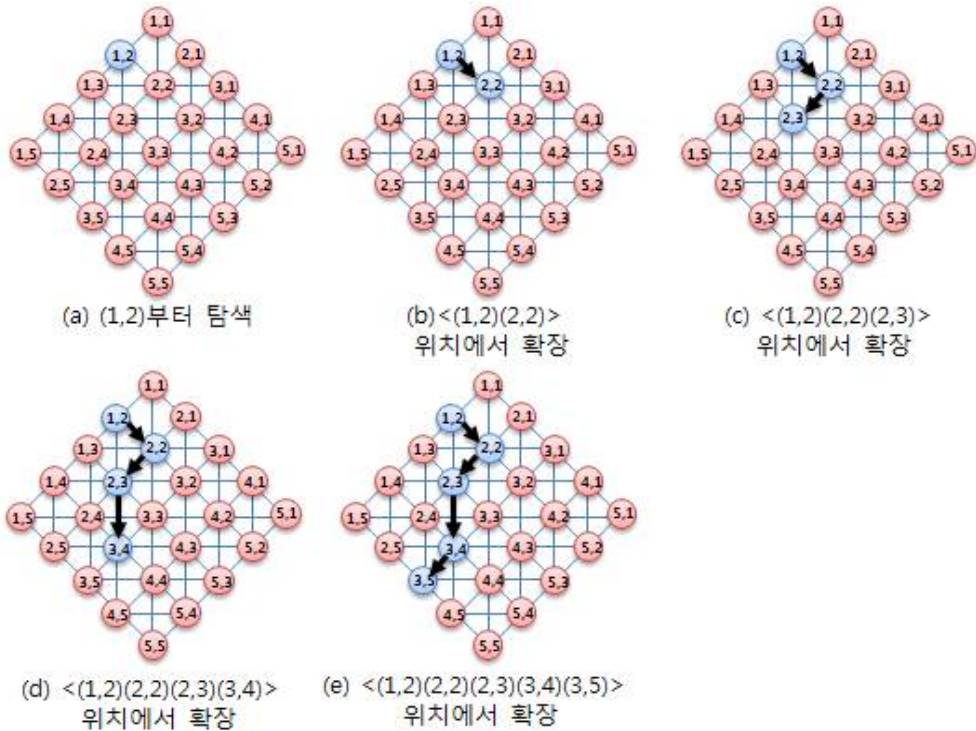
패턴에서 대해 ROI에 도달할 수 있는 확률을 점수화한 수치를 의미한다. punctual score에는 avgScore, sumScore, maxScore가 존재하며, 각 식을 이용하여 값을 구하고 이 중 가장 수치가 높은 값으로 각 패턴을 평가한다. 해당 기법은 패턴 빈도 기반의 패턴 예측이 아닌 ROI 기반의 punctual score 기법을 이용하여 패턴을 예측하기 때문에, 높은 경로 예측 정확도를 지닌다. 그러나 ROI만을 기준으로 T-pattern을 생성하고 이를 기반으로 경로를 예측하기 때문에, ROI에 포함되지 않는 낮은 빈도를 가진 패턴에 대해서는 경로 예측 대상에서 제외되는 문제점이 존재한다.

다. A. Lee et al.의 연구[21]

A. Lee et al.의 연구에서는 추가적인 후보셋 생성에 대한 연산 비용 문제를 해결하기 위해 후보셋 생성 없이 시공간 패턴데이터를 이용하여 효율적으로 패턴을 추출하는 기법을 제안하였다 제안하는 기법은 패턴 정보 리스트 생성 단계, 그래프 기반(graph-based) 마이닝 단계로 구성된다. 첫째, 패턴 정보 리스트 생성 단계에서는 모든 객체의 패턴 포인트에 대해 grid 형태의 mapping graph를 구성한다. 이는 일반적으로 정방향의 도로 네트워크와 유

사하며, 오직 인접 노드와의 연결을 구성하기 때문에, 패턴 추출 시 필요한 탐색 공간을 줄일 수 있다. 또한, 패턴을 추출하기 위한 최소 임계값인 n-support를 기반으로 모든 객체의 패턴에 대해 필터링을 수행하고, 임계값 이상인 패턴에 대해 노드의 패턴을 기준으로 Ti-List에 저장한다. Ti-List를 통해 객체의 패턴을 저장함으로써, 패턴 추출 시 패턴들 간의 지역성을 보장할 수 있다. 둘째, 그래프 기반 마이닝 단계에서는 Ti-List에 저장된 패턴들과 mapping-graph를 이용하여 패턴을 추출한다. 패턴 추출은 깊이 우선 탐색(DFS: Depth First Search) 기법을 통해 진행되며, Ti-List에 있는 패턴들과 mapping graph를 비교하여 패턴을 탐색한다. 해당 패턴에서 갈 수 있는 인접 패턴은 Ti-List에 존재하는 패턴으로 구성되어야 하며, Mapping-graph에서 만약 인접한 패턴이 다수 존재할 경우에는 해당 패턴으로 가기 위해 소요되는 시간이 가장 짧은 패턴을 다음 패턴으로 선정한다. Ti-List와 mapping-graph를 이용하여 패턴을 추출하는 과정은 [그림 13]과 같다. 첫째, Ti-List의 처음 패턴인 (1,2)를 기준으로 mapping-graph를 탐색한다. 인접 리스트가 (2,2), (2,3)의 경우가 존재하며 시간을 고려하여 가장 가까운 패턴, 즉 (2,3)이 다음 패턴으로 선정된다. 둘째, 해당 알고리즘을 재귀적으로 수행하며, (1,2)를 기준으로 하는 패턴은 <(1,2), (2,2), (2,3), (3,4), (3,5)>가 된다. 해당 알고리즘은 기존의 기법과 달리 후보셋의 생성 없이 바로 패턴을 추출하기 때문에 연산 비용을 감소시킬 수 있다. 해당 기법은 도로네트워크의 인접 패턴만을 연결하여 패턴을 추출하기 때문에, 패턴 데이터 마이닝을 위한 탐색을 줄일 수 있으며, 리스트를 통해 인접 패턴 데이터를 저장하기 때문에, 패턴 간의 지역성을 보장할





[그림 13] 패턴 추출의 예

수 있다. 그러나 전체 궤적의 유사성을 고려하지 않고 인접 포인트의 빈도만을 이용하여 패턴을 추출하기 때문에, 다른 기법에 비해 패턴 추출의 정확도가 떨어지는 단점이 존재한다.

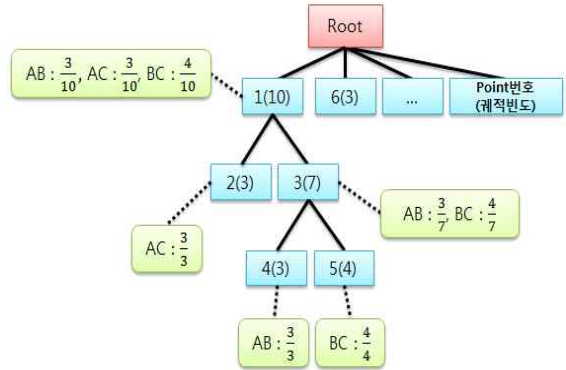
라. L. Chen et al.의 연구[23]

본 연구에서는 기존 도로 네트워크 궤적 마이닝 기법이 자동차에서 측정된 궤적 정보를 기준으로 마이닝을 수행하기 때문에, 목적지의 주변으로 궤적 정보가 생성되어 정확도가 떨어지는 문제점을 제시하였다. 아울러, 본 연구에서는 자동차가 아닌 사용자가 소지한 실제 모바일 기기를 기준으로 궤적 데이터를 생성하며, 예정된 목적지와 기존의 궤적을 바탕으로 마이닝을 수행하여 다음 궤적의 이동 경로

를 예측하는 기법을 제안하였다. 제안된 시스템은 크게 마이닝 모듈과 궤적 예측 모듈로 구성된다. 첫째, 마이닝 모듈에서는 위치 정보를 이용하여 패턴을 추출하며, 총 4단계로 구성된다. 1단계, 모바일 기기로부터 수신 받은 위치 정보를 연속적인 궤적으로 변환하고, 위치 데이터 측정의 오류로 판단되는 궤적의 outlier를 필터링한다. 궤적은 일정 시간 내의 연속적인 포인트로 구성되며, 궤적 속도의 최소 임계값을 정해 데이터 측정의 오류로 판단되는 궤적의 일부분을 삭제한다. 2단계, 시간 기반 클러스터링 기반의 forward-backward 매칭 기법을 사용하여 궤적들을 사용자의 특정 목적지에 따라 분류한다. forward-backward 매칭 기법은 각 궤적 에지의 노드를 이용하여 forward set(FS), backward

set(BS)으로 분류하고, 각 궤적의 FS, BS를 비교하여 같은 노드가 존재할 경우 같은 클러스터로 분류하는 기법이다. 이 때 분류된 각 클러스터는  $\{(origin, destination), Tr_1, Tr_2, \dots, Tr_m\}$ 의 셋으로 구성된다. 3단계, 각 클러스터를 grid network 환경에 적용하여 패턴 추출을 위한 추상화[24,25]를 수행한다. 4단계, 해당 클러스터 내에서 CRPM (Continuous Route Pattern Mining) [26]기법을 통해 패턴을 추출한다. CRPM은 PrefixSpan 알고리즘 [27]을 통해  $n$ -support, 즉 최소 임계값을 설정하여 1차 필터링을 수행하고 시간 임계값을 설정하여 궤적이 특정 노드를 지날 때 시간에 대해 필터링하고 패턴을 추출한다.

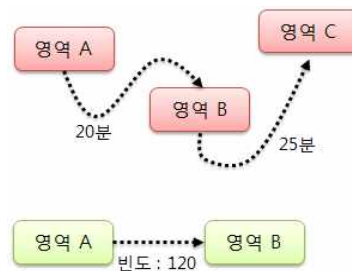
둘째, 궤적 예측 모듈은 추출된 패턴을 이용하여 목적지를 기반으로 궤적을 예측한다. 궤적 예측 모듈은 마이닝 모듈을 통해 생성된 패턴을 이용하여 패턴 트리(pattern tree) 기반 궤적 예측 인덱스를 구성한다. 패턴 트리는 [그림 14]와 같이 연속적인 패턴의 포인트를 루트로부터 top-down 방식으로 저장한다. 아울러, 각 노드에는 현재 노드에서 다음 노드로 갈 수 있는 확률 값을 지니고 있다. 예를 들어, 그림에서 1번 포인트에서 2번 포인트로 갈 확률은  $\frac{3}{10}$ 이며, 3번 포인트로 갈 확률은  $\frac{7}{10}$ 이다. 이러한 방식으로 모든 패턴에 대해 패턴 트리를 구성하며, 이를 기반으로 새로운 객체의 궤적 및 가장 잘 매치되는 트리 상의 경로를 탐색함으로써, 확률이 높은 궤적을 예측 궤적으로 반환한다.



[그림 15] 패턴 트리의 구성

마. F. Giannotti et al.의 연구 [9]

본 연구는 궤적 데이터 마이닝 과정 중에서 궤적의 출발, 도착지점에 대한 빈도 및 이동 경로의 다양성을 고려한 접근 연관성 패턴을 다루는 접근 방식에 대해 기술한다. 본 연구에서는 전체 궤적에 대한 통계를 바탕으로 T-Pattern, T-Flow를 생성한다. [그림 15]와 같이, T-Pattern는 이동 객체 궤적이 지나는 영역(Region)에 대한 시간 정보를 지니고 있으며, T-Flow의 경우 각 구간 이동 객체 이동 빈도에 대한 정보를 저장하고 있다. 따라서 이러한 정보를 이용하여 도시 간 이동이 빈번한 시간 대 및 도로를 탐색하거나 T-Flow정보를 이용하여 구간별 이동 횟수에 대한 임계값을 정의하고 임계 값 이상을 지니는 구간을 정체구간이라 정의하여 질의를 수행한다.



[그림 15] M-Atlas 궤적 정의

### 4. 궤적 데이터 마이닝 기법 분석

3장에서 서술한 기존의 이동 객체 궤적 데이터 마이닝 기법의 특징을 정리하면 <표 1>과 같다. 본 논문에서는 이동 객체 궤적 데이터 마이닝 기법의 분석을 위한 핵심 정보로 데이터 공간, 궤적 표현 정보, 이용 궤적 정보, 응용 분야, 알고리즘 복잡도, 시각화의 6가지 항목을 선정하였다. 첫째, 데이터 공간은 2.1절에서 설명한 것과 같이 각 기법에서 고려

한 이동 객체 궤적이 유클리디언 기반[1,10] 또는 도로 네트워크를 기반[5,9,14,18,20,22]으로 이동하였는지에 대한 여부를 표현한다. 아울러 제안하는 마이닝 기법이 이러한 특징에 제한 없이 모두 적용 가능한 경우 두 가지를 지원한다[3,21]. 궤적의 데이터 공간 특징은 궤적 정보를 생산하는 이동 객체 종류에 영향을 받으며, 유클리디언 공간의 경우 야생 동물의 이동 흔적이나 지구 과학에서 사용되는 관측 데이터 등과 같이 객체 이동 영역의 제한이 적

연구 [reference]	데이터 공간	궤적 표현	이용궤적 정보	응용분야	알고리즘 복잡도	시각화	차별되는 특징
궤적 데이터 클러스터 링 기반	[1]	유클리디언	시공간	전체궤적	패턴추출 궤적예측	$n + \log n$	미제공
	[3]	유클리디언/도 로네트워크	시공간	전체궤적	패턴추출	$O(n + \log m)$	제공 MySQL을 통한 질의지원
	[5]	도로 네트워크	시공간	전체궤적	궤적예측	$O(m)$	미제공
	[9]	도로 네트워크	시공간	전체궤적	패턴추출 궤적예측	$O(n^2)$	제공
궤적 데이터 분류 기반	[10]	유클리디언	공간	부분궤적	패턴추출	$O(n^2)$	제공
	[14]	도로 네트워크	공간	부분궤적	패턴추출	$O(n \log n)$	미제공 부분 궤적지원
궤적 연관 규칙 기반	[18]	도로 네트워크	시공간	전체궤적	궤적예측	$O(m^2)$	제공 ROI정의
	[20]	도로 네트워크	시공간	전체궤적	궤적예측	$O(m^2)$	미제공 패턴트리 정의
	[21]	유클리디언/도 로네트워크	공간	부분궤적	궤적예측	$O(m^2)$	제공 네트워크 Graph 표현
	[23]	도로 네트워크	공간	전체궤적	패턴추출 궤적예측	$O(m^2)$	미제공
	[9]	도로 네트워크	공간	전체궤적	패턴추출 궤적예측	$O(n^2)$	제공

[표 3] 이동 객체 궤적 데이터 마이닝 기법의 특징

은 경우를 포함하며, 사람, 자동차 등의 궤적은 일정 도로를 따라 제한된 이동을 하는 도로네트워크 공간으로 표현된다.

## 5. 향후 연구 방향

본 장에서는 이동 객체의 궤적 데이터 마이닝 연구가 나아가야 할 연구 방향을 제시한다. 첫째, 궤적 데이터의 위치 정보를 보호하는 궤적 데이터 마이닝 알고리즘에 대한 연구가 필요하다. 최근 위치 정보 서비스의 발전과 함께 대량의 위치 데이터가 생산되고, 이를 활용하는 많은 부가 응용 서비스가 활성화되고 있다. 그러나 위치 정보를 생산하는 사용자의 경우 자신의 위치 데이터가 제 3의 서비스 제공자 및 연구 그룹에 제공됨으로 인하여 자신의 개인 정보가 유출될 수 있는 위험을 발견하였다. 특히, 궤적 데이터의 경우 개인정보 보호를 위해 궤적에 사용자의 이름이나 주소 등의 데이터를 표시하지 않고 익명 아이디를 사용한다. 하지만 사용자 궤적의 전체 이동 경로 및 패턴 분석을 통해 방문 빈도가 높은 지점을 집이나 직장 등으로 매핑하여 제 3의 정보와 결합하게 되면 쉽게 자신 정보가 노출될 수 있다. 따라서 이러한 문제를 극복하기 위해 궤적 데이터를 퍼블리싱(publishing) 하기 전에 궤적 은닉화(anonymization)를 수행하여 변형된 궤적정보를 제공하는 기법에 대한 연구가 진행되었다. 그러나 이러한 경우 궤적의 정보가 누락되거나 왜곡되어 궤적 데이터 마이닝에서 잘못된 정보를 생산할 수 있다. 따라서 이러한 문제점을 모두 고려하여 데이터 마이닝에서 의미 있는 데이터를 보존하되, 사용자의 개인 정보를 유출시키지 않도록 하는 개인정보 보호 기반 궤적 데이터 마이닝 기법에 대한 연구가 필수적이다.

둘째, 실제 도로 환경을 고려한 궤적 데이터 마이닝 기법에 대한 심도 있는 연구가 필요하다. 기존의 궤적 데이터 마이닝에 관한 연구들은 주로 유클리디언 공간상에서 이동 객체를 대상으로 하였기 때문에 도로 네트워크를 따라 이동하는 궤적 데이터에 직접 적용하기에는 어려움이 따른다. 최근 도로 네트워크 기반 궤적 데이터 마이닝에 대한 연구가 수행되고 있지만, 궤적 데이터를 표현하기 위한 기준이 부족한 실정이다. 궤적 데이터 위치 좌표 표현의 경우 유클리디언 공간에서는 GPS 등으로 측정된 (x, y) 좌표를 바로 적용하여 표현 가능하지만, 도로 네트워크의 경우 이를 최소한의 오차를 지니도록 하는 도로 네트워크의 에지(edge)에 맵핑하는 연구 등이 추가로 필요하다. 또한 궤적의 각 시점을 나타내는 데 있어서 도로 네트워크의 에지를 이동하는 시간 구간(duration)으로 표현할 것인지, 에지의 특정 시점을 통과하는 시간으로 표현할 것인지에 대한 기준 또한 명확하지 않다. 따라서 도로네트워크 공간에서 이동 객체의 궤적을 정의하기 위한 기준을 제시하고, 궤적 데이터를 분석하여 도로 네트워크 기반 이동 객체 데이터 마이닝을 수행하는 연구가 필요하다.

셋째, 상황 인식 등을 지원하는 다양한 질의 수행을 위한 궤적 데이터 마이닝 연구가 필요하다. 현재 이동 객체 궤적 데이터 기반 데이터 마이닝 알고리즘은 궤적의 공간 정보 및 시공간 정보를 분석하여 서비스를 제공하고 있다. 그러나 지속적으로 성장하는 위치 기반 서비스와 함께 사용자의 요구 사항 또한 다양해지고 있어, 이를 만족하기 위해서는 위치 정보뿐만 아니라 주변 환경을 고려한 정보를 제공할 수 있어야 한다. 예를 들어 날씨에 따른 이동 객체의 이동 패턴을 분석하여 날씨가 좋은 날에는 교외 지역으로 이동하는 사용자의 수가 증가하는 패턴

을 확인했을 경우, 그에 따라 교통 체증이 발생하는 특정 도로를 우회하는 교통 안내 서비스를 제공할 수 있고, 실제 도로네트워크의 상황 즉 공사 중인 도로나 사고 빈발 구간 등을 우회하는 교통 안내 서비스를 제공함으로써 사용자에게 보다 높은 수준의 서비스를 제공할 수 있는 연구가 수행되어야 한다.

## 6. 결론

최근 위치 기반 시스템 및 무선 통신의 발달로 방대한 양의 위치 데이터가 생성되고 있다. 이러한 궤적 데이터는 이동 객체의 행동을 분석 및 예측하기 위한 매우 소중한 정보로 사용될 수 있으며, 이에 따라 궤적 데이터 마이닝을 통한 궤적 데이터의 분석에 대한 관심 또한 매우 높아지고 있다. 기존 궤적 데이터 마이닝 알고리즘을 이용하여 현재 위치 기반 서비스에 필요한 지식을 추출하기 위해서는 기존 궤적 데이터 마이닝 알고리즘을 분석하고, 이를 응용에 적용하기 위한 연구가 필요하다. 따라서 본 논문에서는 기존 궤적 데이터 마이닝 기법을 클러스터링 기반 마이닝 알고리즘, 분류 기반 마이닝 알고리즘, 연관 규칙 기반 알고리즘으로 분류하고 각각의 특징을 기술하였다. 또한, 기존 연구의 분석을 통해 데이터 공간, 궤적 표현 정보, 이용 궤적 정보, 응용 분야, 알고리즘 복잡도, 시각화의 6가지 항목을 선정하고, 이를 바탕으로 기존 연구들을 비교 및 분석하였다. 아울러, 궤적 데이터 마이닝 알고리즘이 나아가야 할 향후 방향을 크게 세 가지 범주로 분류하여 제시하였다. 이러한 내용을 바탕으로 이동 객체 궤적 데이터 마이닝에 대한 연구가 활발히 진행된다면, 궤적 데이터의 분석 비용을 최소화 하면서 사용자에게 수준 높은 서비스를 제공할 수 있는 위치 기반 서비스 개발 등에 크게 이바지 할 것이다.

## 7. 참고문헌

- [1] J. Yang, M. Hu, "TrajPattern: Mining Sequential Patterns from Imprecise Trajectories of Mobile Object", *International Conference on Extending Database Technology*, pp. 664-681, 2006.
- [2] J. Yang, W. Wang, P. Yu, and J. Han. "Mining long sequential patterns in a noisy environment", *Proc. of SIGMOD*, 2002.
- [3] G. Gidófalvi, T. B. Pedersen, "Mining Long, Sharable Patterns in Trajectories of Moving Objects", *Journal of Geoinformatica*, 2009.
- [4] J. Pei, J. Han, and R. Mao. "CLOSET: An efficient algorithm for mining frequent closed itemsets," in *Proc. of DMKD*, pp. 11-20, 2000.
- [5] S. Qiao, C. Tang, M. Chau, "PutMode: Prediction of uncertain trajectories in moving objects databases", *Appl. Intell.* pp. 370-386, 2010.
- [6] Nodelman U, Shelton CR, Koller D "Learning continuous time Bayesian networks." In: *UAI'03: Proceedings of the 19th conference on uncertainty in artificial intelligence. Morgan Kaufmann*, San Francisco, pp 451-45, 2003.
- [7] Nodelman U, Shelton C, Koller D "Continuous time Bayesian networks" In:

- UAI'02: *Proceedings of the 18th conference in uncertainty in artificial intelligence*. Morgan Kaufmann,, San Francisco, pp 378-387, 2002..
- [8] Ester M, Kriegel H-P, Sander J, Xu X "A density-based algorithm for discovering clusters in large spatial databases with noise". *Proceedings of the 2nd international conference on knowledge discovery and data mining*. AAAI Press, Pittsburgh, pp 226-231, 1996.
- [9] F. Giannotti, M. Nanni, D. Pedreschi, F. Pinelli, C. Reanso, "Unveiling the complexity of human mobility by querying and mining massive trajectory data", *Very Large Data Bases*, pp.695-719, 2011.
- [10] J. Lee, J. Han, H. Cheng, "Mining Discriminative Patterns for Classifying Trajectories on Road Networks", *Knowledge and Data Engineering, IEEE Transactions on*, pp.713-726, 2010.
- [11] H. Cheng, X. Yan, J. Han, and C.-W. Hsu, "Discriminative Frequent Pattern Analysis for Effective Classification," *Proc. 23rd Int'l Conf. Data Engineering*, pp.716-725, Apr. 2007.
- [12] Y. W. Chen and C.J. Lin, "Combining SVMs with Various Feature Selection Strategies," *Feature Extraction: Foundations and Applications*, I. Guyon, S. Gunn, M. Nikravesh, and L.A. Zadeh, eds., pp. 315-323, Springer, 2006.
- [13] V.N. Vapnik, *Statistical Learning Theory*. John Wiley & Sons, 1998.
- [14] J. Lee, J. Han, X. Li, H. Gonzalez, "TraClass: Trajectory Classification Using Hierarchical Region-Based and Trajectory-Based Clustering", *Int. Conf. on Very Large Data Base*, pp.1081-1094, 2008.
- [15] P. D. Gr Åunwald, I. J. Myung, and M. A. Pitt. "Advances in Minimum Description Length: Theory and Applications", *MIT Press*, 2005
- [16] J. Lee, J. Han, and X. Li. "Trajectory outlier detection: A partition-and-detect framework", *In Proc. 24th Int'l Conf. on Data Engineering*, pages 140-149, Cancun, Mexico, Apr. 2008.
- [17] J. Lee, J. Han, and K. Whang. "Trajectory clustering: A partition-and-group framework", *In Proc. ACM SIGMOD Int'l Conf. on Management of Data*, pp. 593-604, 2007.
- [18] F. Giannotti, M. Nanni, D. Pedreschi, F. Pinelli, "Trajectory Pattern Mining", *ACM SIGKDD international conference on Knowledge discovery and data mining*, 2007.
- [19] F. Giannotti, M. Nanni, and D. Pedreschi. "Efficient mining of sequences with

- temporal annotations”, In *Proc. SIAM Conference on Data Mining*, pages 346-357. 2006.
- [20] A. Monreale, F. Pinelli, R. Trasarti, F. Giannotti, “WhereNext: A Location Predictor on Trajectory Pattern Mining”, *ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009.
- [21] A. Lee, Y. Chen, W. Ip, “Mining Frequent Trajectory Patterns in Spatial-temporal Databases”, *Information Sciences: an International Journal*, 2009.
- [22] L. Chen, M. Lv, G. Chen, “A system for destination and future route prediction based on trajectory mining”, *Pervasive and Mobile Computing*, pp.657-676, 2010.
- [23] H. Jeung, H.T. Shen, X. Zhou, “Mining trajectory patterns using hidden Markov models”, in: *Proc. of International Conference on Warehousing and Knowledge Discovery*, pp.470-480, 2007
- [24] Q. Ye, L. Chen, G. Chen, “Predict personal continuous route”, *Journal of Zhejiang University*, pp.221-223, 2009.
- [25] L. Liao, D. Fox, H. Kautz, “Extracting places and activities from GPS traces using hierarchical conditional random fields”, *International Journal of Robotics Research*, 119-134, 2007.
- [26] J. Pei, J. Han, B. Mortazavi, H. Pinto, “Mining sequential patterns by pattern-growth: the Prefix Span approach”, *IEEE Transactions on Knowledge and Data Engineering*, pp.424-440., 2004
- [27] 이충호, 안경환 외, “u-GIS 공간정보 기술 동향”, *한국 전자통신 연구소, 전자통신동향분석 제 22권 제 3호*, 6월, 2007
- [28] 박경린, “ITRC의 텔레매틱스 요소기술 연구 현황” *Special Report, TTA 저널/정보통신 표준화소식*, 2009
- [29] Griswold, “Active Campus: Experiments in Community-Oriented Ubiquitous Computing,” *Journal of Computer*, pp.73-81, 2004.
- [30] Pashtan, “Personal Service Areas for Mobile Web Applications,” *Journal of IEEE Internet Computing*, 2004, pp.34-39.
- [31] T. Gu, “Toward an OSGi-Based Infrastructure for Context-Aware Applications,” *IEEE Pervasive Computing*, pp.66-74, 2004.
- [32] B. Lee, “Defining of Trade Area using Spatial Data Mining Technique in Business GIS”, *The Journal of GIS Association of Korea*, Vol.11, No.2, pp.171-184, July 2003
- [33] J. Lee, J.Han, X. Li, “Trajectory Outlier Detection: A Partition-and-Detect Framework”, *International Conference on Data Engineering*, pp140-149, 2008



장 미 영

2009년 전북대학교 컴퓨터  
공학과 졸업(학사)

2011년 전북대학교 컴퓨터  
공학과 졸업(석사)

2011년 - 현재 전북대학교

컴퓨터공학과 박사과정

관심분야 : 공간 데이터베이스, LBS k-NN 질의 처  
리 알고리즘, 위치 정보 보호, 궤적 데이터 마이닝

관심분야 : 공간 네트워크 데이터베이스, 하부저장  
구조, 클라우드 컴퓨팅, 데이터베이스 아웃소싱 등



윤 민

2008년 전북대학교 컴퓨터  
공학전공 졸업(학사)

2010년 전북대학교 컴퓨터  
공학과 졸업(석사)

2011년 - 현재 전북대학교

컴퓨터공학과 박사과정

관심분야 : 공간 데이터베이스, 클라우드 컴퓨팅, 데  
이터베이스 아웃소싱



장 재 우

1984년 서울대학교 전자  
계산기 공학과(공학사)

1986년 한국과학기술원  
전산학과(공학석사)

1991년 한국과학기술원

전산학과(공학박사)

1996년 - 1997년 Univ. of Minnesota, Visiting  
Scholar

2002년 - 2004년 Penn State Univ., Visiting  
Scholar

1991년 - 현재 전북대학교 IT 정보공학과 교수