

로그데이터의 시공간 데이터마이닝 및 시각화 연구동향*

조나혜** · 강영옥***

A Research Trends about Spatio-temporal Data Mining and Visualization of Log Data*

Nahye Cho** · Youngok Kang***

요약 : 최근 모바일과 웹 사용의 증가로 일상생활에서 기록되는 로그데이터를 다양하게 분석하여 의미있는 정보와 지식을 이끌어내고자 하는 연구가 증가하고 있다. 웹과 모바일 기기로부터 생성되는 로그데이터는 시공간적인 정보를 담고 있으며, 데이터를 다차원적으로 탐색하고 시각화하여 기존에 분석하지 못했던 다양한 의미를 찾을 수 있음이 확인되고 있다. 본 연구에서는 시간과 공간 정보를 가지고 있는 로그데이터를 다차원적으로 탐색하고, 의미를 분석하는 데이터마이닝과 시공간 데이터를 시각화하여 의미를 도출하고자 하는 시각화 관련 연구들을 분야별, 연구방법별로 분석하여 연구동향을 살펴보고 의미를 찾고자 한다.

주요어 : 로그데이터, 시공간 데이터마이닝, 시각화, 연구동향

Abstract : Recently increasing the usage of mobile and web, the researches which are trying to draw a meaningful information and knowledge in numerous different ways by analyzing the log data recorded in our daily life are growing very fast. Many researchers have studied to explore the patterns and the meanings by visualizing the log data in multi-dimensional ways because the log data generated from a number of sensors mounted on the mobile device shall contains a temporal and spatial information. The goal of this study is searching a implication and exploring the researches related to both data mining and visualization that analyze the log data which has spatio-temporal information multi-dimensionally by research area and method.

Key Words : Log data, Spatio-temporal data mining, Visualization, Research trend

I. 서론

정보화 사회가 급진전하면서 개인, 정부기관, 그리고 대기업 등에서 대량의 데이터를 만들어내고 공개하고 있다. 예를 들면 대형마트의 상품 바코드 데이터, 대중교통카드 및 택시의 GPS 데이터, 인터넷 홈페이지 접속 시 서버에 기록되는 데이터 등이 있다.

특히 최근에는 모바일 사용의 증가로 일상생활에서 기록되는 로그데이터를 다양하게 분석하여 기존에 알지 못했던 의미있는 정보를 발견하고자 하는 연구가 많이 이루어지고 있

다. 모바일 기기에 탑재된 많은 센서로부터 생성되는 로그데이터는 시공간적인 정보를 담고 있으며, 이러한 데이터를 다차원적으로 탐색하고 시각화하여 의미를 도출해낼 수 있다. 기존의 지리 분야에서는 공간 데이터를 중심으로 연구가 활발히 진행되었지만, 시간을 고려한 공간 데이터에 대한 연구는 최근에 들어 다양하게 시도되고 있다(Vu *et al.*, 2008). 이미 해외에서는 시공간 데이터를 최신의 분석 기술과 융합하여 다양한 분야에서 연구하고 있다.

본 연구에서는 시간과 공간 정보를 가지고 있는 로그 데이터를 다차원적으로 탐색하고 시공간 데이터마이닝 및 이를

*본 연구는 본 연구는 2014년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2014S1A5B6037290)

**이화여자대학교, 사회과교육과, 박사수료(Ph.D. Student, Department of Social Studies, Ewha Womans University, 881202@hanmail.net)

***이화여자대학교, 사회과교육과, 교수(Professor, Department of Social Studies, Ewha Womans University, ykang@ewha.ac.kr)

시각화한 한 연구들을 분석하여 그 동향을 살펴보고자 한다.

II. 로그데이터 연구분야

최근 디지털 경제의 확산으로 기존의 데이터에 비해 그 규모가 방대하고, 생성주기가 짧은 대규모 데이터를 빅데이터라고 하며, 이러한 빅데이터 환경에서 다양한 일반 사용자들이 생성하는 SNS, 휴대폰, UCC 등 문자 외에 위치정보, 수치정보, 동영상정보 등 다양한 데이터들이 수집되고 유통되고 있다. 빅데이터는 흔히 3V라고 하는데 데이터의 규모(Volume), 데이터의 생성주기(Velocity), 데이터의 다양성(Variety)에 있어 특성이 있다는 것이다(강만오 등, 2012). 그러나 빅데이터는 단순히 많은 양의 데이터만을 뜻하는 것이 아니라 대규모의 데이터를 분석하여 그 속에서 발견하기 힘든 유용한 정보를 찾아내는 것에 핵심이 있다. 이러한 측면에서 로그데이터는 빅데이터의 일종이라고 볼 수 있다.

최근 모바일과 웹 사용의 증가로 일상생활에서 기록되는 로그데이터의 양이 급증하면서, 이를 통해 의미있는 분석을 하고자 하는 분야가 다양해지고 있다. 이러한 로그데이터 연구는 국내외에서 웹사이트의 효율적인 활용과 고객 마케팅 및 서비스를 위해 일찍부터 마케팅 분야(변시우, 2002; Jun, 2002; Sujatha and Punithavalli, 2012; Kim, 2014), 정보검색 분야(Park and Joo, 2013; Nicholas *et al.*, 2014; Shin *et al.*, 2014)에서 많은 연구가 진행되고 있으며, 의료, 게임관련 분야, 금융 이상거래 탐지, 교육 및 학습 분석학 분야까지 다양하게 로그데이터를 분석하고자 하는 시도가 이루어지고 있다.

의료 분야에서는 판단하기 어려운 질병을 진단하고 모니터링하기 위해 로그데이터를 연구하고(Pierleoni *et al.*, 2014; Kim and Bae, 2015), 온라인 및 모바일 게임에서는 해킹 시도나 부정사용을 방지하기 위해 게임 로그 데이터를 분석하는 연구들이 있다(이기용·정규만, 2014). 금융거래에서도 일반적으로 보이지 않는 결제를 발견하기 위해 결제 로그 분석 및 데이터 마이닝을 이용하여 실시간 이용자 정보와 결제 정보를 분석하여 이상거래를 식별하고자 한 연구가 있다(정성훈 등, 2015). 교육 분야에서는 온라인 교육 사이트 품질향상, 새로운 입시 전략 기획, 학습자의 학습 패턴과 학업 성취도 간의 관련성을 분석하는 연구들이 진행되어 왔다(김윤경, 2002; 최승배 등, 2009; 조일현·김정현, 2013). 뿐만 아니라 최근에는 사용자의 개인 행동 패턴이나 감정을 인식하고, 사회적 관계를 분석(Min and Cho, 2011; Min *et al.*, 2013; Singh *et al.*, 2013;

Mafrur *et al.*, 2015)하는 등 다양한 연구로 확대되고 있다.

로그데이터의 시공간적 연구가 가장 많이 이루어지고 있는 분야는 교통카드 등 카드사용 데이터로부터 개인 이동 패턴, 생활 패턴을 분석하여 라이프 스타일에 따른 행동반경을 예측하고 그에 맞는 아이템을 추천하는 연구들이다. Thiagarajan *et al.*(2009)은 WIFI 센서 로그를 분석하여 버스 이동 경로 정보를 승객들의 참여로부터 데이터를 수집하고 버스 도착 시간과 이동경로 시간을 예측하는 시스템을 개발하였다.

Ferreira *et al.*(2013)연구에서는 택시 데이터를 독자적인 시공간적 시각화 기법과 통계 그래프를 동시에 사용하여 분석함으로써 택시의 이동 패턴, 그리고 이와 관련된 도시민 삶의 패턴과 도시 라이프를 연구하였다. 택시에는 많은 센서가 탑재되어 있기 때문에 이러한 도시민의 경제적 활동, 이동 패턴 등을 통해 도시 라이프를 확인할 수 있었다. 이 연구에서는 'Taxi Viz'라는 시스템을 개발하고 새로운 모델을 제안하였는데, 이 시스템을 이용하여 한 도시 내에서 2011년과 2012년의 택시 이동경로 시각화를 통해 택시의 출발지점과 목적지간의 이동량 패턴을 비교분석하였다. 평소와 다른 이동 패턴이 보이는 시점들을 발견하고 이에 대한 원인을 분석해본 결과, 자연재해에 따른 결과임을 확인할 수 있었다. 2011년과 2012년의 허리케인의 영향에 따른 결과도 각각 차이를 보이고 있는데, 그 차이를 사회경제적으로 비교분석하고 해석함으로써 복잡한 택시데이터를 한눈에 시각화하여 좀 더 빠르게 실시간으로 도시 라이프를 이해할 수 있는 가능성을 열어 준 연구라 할 수 있다.

서울 수도권 지하철 교통망 승객 흐름을 시각화한 연구도 있다(김호성 등, 2010). 서울 지하철 승객의 대용량 교통카드 트랜잭션 데이터베이스를 분석하여 지하철 교통망을 구성하고, 교통카드 데이터 마이닝을 통해 최단 경로 상의 승객 흐름을 확인하였다. 또한, 역 별 위치를 노드로 표현하고, 역으로 유입되는 승객수와 역에서 유출되는 승객수를 표시하여 승차차 승객수 계산, 그리고 하차 승객수에서 승차 승객수를 빼서 역 주변에 머무는 인구를 확인하고, 시간에 따른 밀도분석을 통해 역 주변 유동인구 밀도를 시각화하여 이동패턴을 분석하였다.

외출을 두려워하는 장애인의 심리적 부담감을 덜어주기 위해 로그데이터 연구가 시도된 사례도 있다. 서정민 등 (2010)과 Hwang *et al.*(2012)은 장애인 라이프 로그를 기록하고 분석할 수 있는 시스템을 자체적으로 개발하여 제안하였다. 스마트 폰이나 모바일 기기의 로그 기록들을 활용하여 사용자의 라이프 로그를 저장하고 패턴을 학습한 뒤, 가속도 센서 등

으로 일상 패턴에서 나타나지 않는 사고나 위험상황을 감지하도록 알고리즘을 개발하였다. 라이프로그의 모니터링을 통해 장애인의 상태를 확인할 수 있도록 하였다.

범죄 분야에서는 범죄자와 연루자들의 휴대폰 정보를 통해 시간대별 이동 경로를 추적하여 범죄자의 정확한 위치를 찾아내는 연구가 주로 이루어지고 있다. Nakaya and Yano (2010)는 2003년 1월~2004년 12월까지 수집된 교토 부의 주요 범죄 데이터를 시공간적으로 시각화 하여 그 범죄 패턴을 분석하였다. GIS 소프트웨어인 'ArcGIS' 및 GPS 트래킹 소프트웨어인 'GeoTime' 툴을 사용하여 시공간경로(Space Time path), 그리고 시공간커널밀도추정(Space-time Kernel Density Estimation)과 시공간 스캔통계(Space-time Scan Statistics) 방법을 통해 시공간적으로 높은 밀도의 클러스터가 발견되는지 확인하였다. 3D형태의 핫스팟 분석을 수행한 결과 교토부 내에서 가쓰라 클러스터와 모모야마 클러스터 사이에 서로 교차적인 범죄 클러스터가 발생하는 것을 확인할 수 있었다.

III. 로그데이터의 데이터 마이닝 연구

웹 및 모바일 등으로부터 수집되는 로그데이터는 시공간적 특성을 가지고 있기 때문에 시공간적인 데이터마이닝 기법을 사용하여 다양한 의미 분석이 가능하다. 이러한 시공간 데이터를 활용하여 데이터마이닝을 통한 패턴 추출, 의미 있는 정보 발견, 미래 예측 등의 분석을 진행하고 또 다차원적으로 데이터를 분석하여 시각화 하는 연구가 많이 이루어지고 있다.

데이터마이닝에 대한 수많은 연구들은 30년 전부터 해외에서는 연구되어 왔다(Andrienko *et al.*, 2003). 초기 데이터마이닝 연구는 큰 데이터(large data)로부터 관련된 정보를 추출하는 새로운 메커니즘을 개발하는 것이었다. 그러나 점점 더 거대하고 다양한 종류의 데이터 특히 시공간적인 관계를 가진 데이터가 생성되면서 더욱 복잡한 계산을 요구하게 되었다.

로그데이터 관련 주요 연구들을 활용한 데이터 분석 방법론 별로 정리한 내용은 표 1과 같다. 표 1에 나타난 바와 같이 연관분석, 군집분석, 분류분석, 트렌드 분석, 회귀분석 등이 대표적으로 사용되었다(Koperski *et al.*, 1996; Yang and Webb, 2002). 연관분석(association analysis)은 대량의 데이터에 숨어 있는 연관규칙(association rule)을 찾는 방법이다(이정진, 2011). 연관 분석을 통한 결과들은 명시적으로 나타나지 않지만 유의미한 패턴들이나 규칙들이며 향후 분석을 위한 좋은 실마리를 제공할 수 있다(이건학, 2004).

연관 분석을 위한 대표적인 알고리즘으로 선형적 알고리즘(apriori algorithm)이 있다. 선형적 알고리즘은 트랜잭션 데이터베이스(transactional databases)에서 빈번하게 발생하는 아이템 세트와 연관규칙학습을 위한 알고리즘이다. 이는 데이터베이스의 빈번한 개별 항목을 식별하고 아이템 세트가 데이터베이스에서 충분히 자주 나타나는 만큼 더 큰 항목 세트로 확장하여 진행한다. 선형적으로 결정된 빈도 아이템 세트 데이터베이스의 일반적인 경향을 강조하며 이는 연관 규칙을 결정하기 위해 사용될 수 있다. 특히 이 알고리즘은 시장 바구니 분석과 같은 영역에서 사용되기도 한다. 예를 들어 Shin *et al.*(2014) 연구에서는 동시에 사용되는 정보들의 연관규칙(if-then form)을 발견하기 위해 선형적 알고리즘을 사용하였는데, NDSL 과학 기술 정보 검색 시스템에서 국제특허 분류 기준인 IPC(International Patent Classification; 표 2) 필드로 분류된 세션들 중에서 B(treatment manipulation, transportation)와 F(mechanical engineering, lighting, heating, weapon, explosive) 세션이 높은 관련이 나타나면 이는 B세션의 정보를 검색할 때 F세션의 정보도 같이 검색하는 경향이 높다는 것을 발견할 수 있음을 보여주고 있다. 이견학(2004)의 연구에서는 교통사고의 공간적 패턴과 이웃한 공간과의 공간적 연관성을 분석하기 위해 선형적 알고리즘을 사용하였다. 이 연구에서는 교통사고 지점으로부터 500m 이내에 있는 객체들을 교통사고와 인접한 공간 객체로 간주하고, 사고 유형별 클러스터링한 군집 유형별 이웃한 공간객체와의 도로시설, 문화교육시설, 산업시설, 서비스시설, 의료후생시설, 주택시설, 행정기관, 지형, 하천과의 연관성 분석을 수행하였다.

연관분석에서 또 다른 대표적인 알고리즘으로 베이지 네트워크(bayesian networks) 방법이 있다. 베이지 네트워크는 랜덤 변수의 집합과 방향성 비순환 그래프를 통하여 그 집합을 조건부 독립으로 표현하는 확률의 그래픽 모델이다. 예를 들어, 베이지 네트워크는 질환과 증상 사이의 확률관계를 나타낼 수 있다. 증상이 주어지면, 네트워크는 다양한 질병의 존재 확률을 계산할 수 있다. 베이지 네트워크에서는 추론과 학습을 수행하기 위한 효과적인 알고리즘이 존재한다. 불확실성 하에 문제를 표현하고 해를 구할 수 있는 베이지 네트워크의 일반화를 영향 다이어그램이라고 부른다.

Min and Cho(2011) 연구에서는 이 방법을 사용하여 사회적 관계 및 성격을 분석하였다. 이 연구에서는 모바일에 수집되는 로그(Call, SMS, PIMS, Bluetooth etc)로부터 베이지 네트워크 분석을 통해 사용자의 감정 상태나 행동에 따른 개인적인 맥락(감정, 바쁨, 스트레스, 활동)과, 연락 패턴에 따른

표 1. 로그데이터를 이용한 데이터마이닝 기법 및 알고리즘

Knowledge Type	Algorithm	Field	Author
Association	• Apriori algorithm	정보검색 서비스에서 사용자 로그에 기반한 검색어 분류 및 추천 연구. 교통사고의 공간적 패턴과 이웃한 공간과의 공간적 연관성을 분석.	Shin <i>et al.</i> (2014) 이건학(2004)
	• Binary decision tree • Genetic programming	사용자의 콘텐츠 소비 기록에서 시간적 응집 단위를 고려한 콘텐츠 추천 방법을 제안.	Kim(2014)
	• Association rules	웹 로그를 분석하여 사용자들의 웹 페이지 방문순서와 세션 구분등의 정보를 이용하여 연관규칙을 통한 분석을 수행.	김석기 등(2001)
	• Bayesian networks	사용자의 사회적 관계를 분석하여 사회적 맥락 및 개인적 맥락에 따른 그들의 관계를 관리하고 추천하는 시스템(SmartPhonebook)을 개발.	Min and Cho(2011)
Clustering	• Spread clustering	로그에 기초한 키워드 추출 방법을 통해 가장 많은 빈도의 중복되는 키워드 셋들 간에 높은 연관관계를 가지는 키워드들을 클러스터링 함으로써 spread에 기초한 클러스터링 방법은 제안.	Park and Joo(2013)
	• Grid clustering	모바일 기기로부터 수집되는 사용자의 로그데이터를 통해 사용자의 POI 지점을 발견.	Montoliu and Gatica-Perez (2010); Montoliu <i>et al.</i> (2013)
	• DBSCAN clustering algorithm	모바일로부터 수집되는 여러 가지 로그 기록들을 이용하여 사용자의 기분 상태를 상황-인지(context-awareness)를 통해 추론할 수 있는 시스템을 개발.	LiKamWa <i>et al.</i> (2011)
	• K-means	교통사고의 공간적 패턴을 분석한 연구로써 교통사고 군집 및 그 특성을 분석.	이건학(2004)
	• Density analysis - Nearest neighbor - Kernel analysis	진주시 교통사고의 계절별, 주야간별, 토지용도, 그리고 교통사고 유형에 따른 시계열적 공간분포 특성을 분석.	성병준 · 유환희(2014)
	-	Web Usage Mining은 웹 서버로부터 사용자의 접근 패턴을 자동으로 발견하고, 사용자의 navigation 패턴을 예측.	Sujatha and Punithavalli(2012)
Classification	-	웹 로그를 분석에 기초하여 클러스터링을 이용한 웹 페이지 세션 구분 및 분류.	김석기 등(2001)
	• SVM • NN • RF	일상에서 느끼는 사람들의 행복감을 모바일 로그 데이터로부터 자동적으로 인식할 수 있는 방법을 연구. - data: call log, sms and Bluetooth proximity data - 모바일로부터 얻을 수 있는 1) 사람들의 모바일 폰을 통해 알 수 있는 활동, 추가적인 지표인 2) 날씨 정보, 그리고 3) 개인의 성향(특성)을 통해 행복감을 자동적으로 인식	Bogomolov <i>et al.</i> (2013)
	• NN • SVM • RF • GMB • Decision trees	스마트 폰으로부터 수집되는 활동 기록, 날씨 상태, 그리고 성격적 특성 등으로부터 일상 생활에서 느끼는 스트레스와 스트레스 받지 않는지 정보를 자동적으로 분류할 수 있는 시스템 개발. - data: people activities (smartphones), weather conditions, personality traits	Bogomolov <i>et al.</i> (2014)
	• SVMs • Decision trees • Naïve Bayes	스마트폰으로부터 수집되는 데이터를 마이닝 하여 사회적 관계의 삶의 양상 가족과 업무, 가족과 사회적 관계, 업무와 사회적 관계 등을 분류하고 여러 기법을 통해 그 결과를 비교. - data: contact list, call log, and SMS log, (and Facebook)	Min and Cho(2011)
	• Correlation analysis • Multiple regression • SVM classifier	스마트폰으로부터 자동적으로 추출된 행동 특성과 자가보고 형태의 5가지 성격유형의 관계를 분석. - data: GPS, call logs, and Bluetooth (SMS logs, Call logs, App logs, BT logs, Profile logs)	Chittaranjan <i>et al.</i> (2013)
	• Naive bayes	면대면 상호작용, 전화, SMS 로그를 통해 측정된 사회적 행동들이 소비행동을 예측하는데 활용될 수 있음을 확인하고, 사회적 관계 패턴을 분석하여 소비 행동을 분류하는 것이 보다 더 지출 행동에 대한 예측효과가 높은 것을 발견함.	Singh <i>et al.</i> (2013)
Trends Analysis	• Floyd shortest path • Density analysis	서울 지하철 승객의 대용량 교통카드 트랜잭션 데이터베이스로부터 지하철 교통망을 구성하고 분리된 네트워크로부터 최단경로 상의 승객 흐름과 역 주변 유동인구 밀도분석 및 시각화.	김호성 등(2010)
	• Sequential pattern analysis	웹 로그를 분석하여 사용자들의 웹 페이지 방문의 순서에 따른 방문 패턴을 발견	김석기 등(2001)
	• 빈발한 시퀀스 패턴 - All_MOP - Max_MOP	빈발한 시퀀스 패턴을 탐색하기 위한 기법으로 새로운 알고리즘을 제안하고, 이동 객체 데이터(모바일 환경)로부터 모든 빈발한 패턴과 최대 패턴들을 탐색하여 기존 알고리즘 DFS_MINE와 비교평가.	Vu <i>et al.</i> (2008)
Regression Analysis	• 시공간 순서패턴 - MINE_MOP - MovingSeqCreation - AllMOP	버스 이동 경로 정보를 승객들의 참여로부터 데이터를 수집하고 버스 도착 시간과 이동 경로 시간 예측하는 시스템을 개발.	Thiagarajan <i>et al.</i> (2009)
	• Regression analysis	고객의 구매행태를 분석하는데 있어서 기존의 사이트 평가 지표로 사용되는 있는 페이지뷰의 한계를 보완하기 위해 고객의 사이트내 체류시간, 이벤트 참여여부 등 새로운 지표를 제안.	Jun(2002)
	• 기술통계분석 • 상관분석 • 중다회귀분석	학습자 시간관리 전략(학습시점 간격의 규칙성, 총 학습시간, 학습 접속 횟수)을 독립변수로 학업성취도를 종속변수로하여 중다회귀분석을 통해 수행하여 학습성취도와의 관계를 분석.	조일현 · 김정현(2013)

사회적 맥락(관련 활동, 관계 유형, 친밀함, 관계)을 추출하는 방법을 제안하고 있다. 여기서 베이스 네트워크 모델은 모바일 환경에서 불확실한 것들(감정, 바쁨, 스트레스, 활동, 관련된 활동, 관련된 유형, 친밀, 관계)을 다루는데 효율적으로 이용될 수 있음을 보여주고 있다.

군집분석(clustering analysis)은 여러 집단의 데이터들이 섞여 있고 각 데이터의 소속집단을 모르는 경우 유사한 속성을 갖는 데이터의 군집을 찾는 기법이다. 마케팅 분야(Sujatha and Punithavalli, 2012)에서는 웹 로그 데이터로부터 사용자의 패턴을 발견하여 잠재적인 사용자를 분류하고, 비슷한 관심사를 가진 그룹으로 클러스터링한다. 클러스터링된 결과는 미래의 사용자 요구를 예측하는데 활용된다. 김석기 등(2001) 연구에서는 웹 사용자들의 방문 정보를 개개인의 그룹으로 구성하는 세션(session) 구분 작업을 통해서 실제 방문객의 방문 특성을 구분할 수 있음을 확인하였고 이에 대한 정확도를 비교 평가하였다. 정보검색 분야(Park and Joo, 2013)에서는 밀접하게 관련된 키워드 검색을 위해 클러스터링 방법을 사용하며, 높은 관련성 있는 키워드들의 클러스터를 찾는 데 주로 활용되었다.

군집 분석에서 특히, 공간적 정보를 가진 데이터의 경우, 그리드 클러스터링(grid clustering), k-평균(k-means) 방법, 최근 거리(nearest neighbor), 커널 밀도 분석(kernel density analysis), DBSCAN 알고리즘 등을 사용하여 공간적인 변수 또는 거리, 밀도 등에 기초한 분석을 많이 하고 있다(Agrawal and Srikant, 1994). Montoliu *et al.*(2013)과 Montoliu and Gatica-Perez(2010)는 모바일로부터 수집된 사용자 로그와 위치정보를 이용하여 그리드(grid) 기반 클러스터링 알고리즘을 사용하여 사용자 위치정보를 그룹화 하여 머무른 지점(stay points)을 통해 머무른 장소(stay regions)를 발견하고 최종적으로 사용자의 POI (point of interest)를 예측하거나 추천하였다.

이건학(2004)은 교통사고 데이터에 GIS 기법을 활용하여 교통사고의 공간적 패턴과 이웃한 공간과의 공간적 연관성을 연구하였다. 공간적 패턴은 k-평균 알고리즘을 사용하여 군집분석을 수행하였다. k-평균 군집화는 거리에 기반을 둔 클러스터링 기법으로 기준점에서 가까운 곳의 데이터들을 하나의 군집으로 묶는 방법이다. k-평균 군집화는 n개의 데이터를 k개의 클러스터로 분할하는 방식으로, 각 점들이 가장 가까운 클러스터에 속하도록 한다. 따라서 이 방법은 데이터를 분류하거나 기존의 데이터를 학습하여 새로운 데이터를 인식하는 작업에 유용하다. 교통사고 개별 속성(사고내용, 사고 유형,

발생 요일, 월, 주야, 사고자 성별, 도로 형태, 차도폭, 음주운전, 토지이용, 날씨, 사망자수, 사고자 연령 등)을 클러스터링하여 군집을 4개로 분류 하였고 각 군집별 특성을 분석하여 기술하였다. 성병준·유환희(2014)는 진주시 14개의 동, 2013년 1월 1일~12월 31일까지 교통사고 자료 데이터를 이용하여 밀도분석 및 커널분석 기법을 사용한 군집분석을 수행하였고, 교통사고를 크게 시가지, 공업단지, 주거지 등 3개의 공간군집으로 분류하고, 3개의 군집에 대해 차대 사람보다 차대차의 교통사고발생량이 많았음을 분석하였다.

LiKamWa *et al.*(2011)는 모바일에 기록된 로그(어플 사용 기록, 전화 사용기록, 이메일이나 문자 메시지, 웹 사용 기록, 캘린더 입력과 위치정보 변화 등)를 DBSCAN (Density Based Spatial Clustering of Applications with Noise) 알고리즘을 사용하여 사용자의 행동의 특성을 클러스터링하여 기분이나 감정을 패턴화할 수 있는 모델을 개발하였다.

분류분석(classification analysis)은 소속집단을 알고 있는 데이터를 이용하여 모형을 만들어서 소속집단을 모르는 데이터들의 집단을 결정하는 기법이다. 분류분석은 다양한 분야에서 응용되고 있는데, 의사가 환자를 진찰하여 그 진찰기록으로 어느 병인지 분류하는 일, 이메일 제목을 이용하여 스팸 메일을 가려내거나 백화점 매장을 방문한 고객이 상품을 구매할 고객인지 아닌지를 구별하는 일 등에 활용되고 있다(이정진, 2011). 최근에는 감정 및 스트레스를 인식하거나 사회적 관계 및 성격을 분류할 때도 활용되고 있다. 대표적으로 SVM(Support Vector Machines), RF (Random Forests) 분류 기법이 있으며, NN (Neural Networks), 의사결정나무(decision trees) 알고리즘과 함께 주로 사용한다.

서포트 벡터 머신(support vector machine, SVM)은 기계 학습 분야 중 하나로 패턴 인식, 자료 분석을 위한 지도 학습 모델이며, 주로 분류와 회귀 분석을 위해 사용한다. 랜덤 포레

표 2. 국제특허분류 (IPC: International Patent Classification)

Section	Description
A	Commodities
B	Treatment manipulation, transportation
C	Chemical, metallurgy
D	Textile, paper
E	Fixed structure
F	Mechanical engineering, lighting, heating, weapon, explosive
G	Physics
H	Electricity

스트(random forest) 알고리즘은 분류, 회귀 분석 등에 사용되는 앙상블 학습 방법의 일종으로, 훈련 과정에서 구성된 다수의 결정 트리로부터 검출, 분류(부류), 그리고 예측(회귀 분석) 등 다양한 어플리케이션으로 활용된다.

감정 및 스트레스를 연구하는 분야에서는 모바일 폰으로부터의 전화 및 SMS 기록, 개인적 활동 정보, 날씨 정보 뿐 아니라 wifi, 블루투스, GPS 등으로부터 수집된 다양한 로그 데이터를 기반으로 일상생활에서 느끼는 사람들의 행복감이나 스트레스, 감정 등을 자동적으로 인식할 수 있는 방법을 모색하는 연구(Bogomolov *et al.*, 2013; 2014; Seneviratne *et al.*, 2014), 또는 스마트 폰으로부터 개인과 사회적인 행복감 수준을 측정하는 연구도 있다(Muaremi *et al.*, 2012). 이 연구들은 공통적으로 SVM 분류기법을 사용하였고, 각 연구별 특징에 따라 NN, RF, 의사결정트리 알고리즘을 활용하여 각각의 행복감, 스트레스, 감정 등을 분류하였다.

사회적 관계 및 성격을 연구한 분야에서는 모바일 로그 기록(전화, SMS 기록, 블루투스, GPS, 페이스 북 등) 데이터를 SVM 분류기법 외에 통계적 기법(상관분석, 다중회귀분석)과 함께 분석하여 사용자의 성격적 특성(외향적, 쾌활한, 성실한, 정서적 안정성, 그리고 경험에 대한 개방성)으로 자동적으로 분류하였다(Chittaranjan, 2013). Singh *et al.*(2013) 연구에서는 사회적 상호작용 정보를 휴대전화의 로그 기록(call logs; SMS logs; and proximity data)을 통해 확인하고 이것을 사용자의 지출행동을 분류하는 근거로 활용하였다. 로그 기록별 평균과 차이점을 4개의 피쳐(NumEvents, NumContacts, IterationDiversity, EnagementTop3)로 특징화 하고, 새로운 사업에 대한 탐색, 충성스러운 고객이 될 수 있는지의 여부, 과소비 경향이 높는지 낮은지에 따른 분류를 하였다. 각 분류 태스크의 서브셋에 대한 추가적인 분류작업과 교차검증 등을 통해 높은 정확도 값을 가지는 피쳐(features)들을 분류하였더니 사회적-모바일 피쳐들이 소비의 다양성, 충성심, 과소비에 각각 69%, 69%, 77%의 값을 가지는 것을 알 수 있었다. 이는 휴대전화 로그 기록 기반의 사회적 상호작용 패턴이 개인 기반의 지출행동에 대한 예측 능력을 제공할 수 있음을 보여주는 것이라 할 수 있다. Min and Cho(2011)은 모바일 폰으로부터 사용자의 텍스트 메시지나 연락처 기록 등의 데이터를 SVM, 의사결정나무, NB (Naive Bayes) 알고리즘을 사용하여 커뮤니케이션 강도, 반복성, 사용한 도구 또는 매체, 그리고 시간적 경향에 따라 관계를 분류하였다. 연구 결과 여러 라이프 양상 중 사람들의 일반적인 측면은 크게 가족, 직장, 그리고 사회적 관계에 따라 분류되는 것을 확인하였다. 특히,

SVM 알고리즘은 가족과 일, 일과 사회적 관계, 그리고 가족과 사회적 관계를 분류하는데 활용되었으며, 규칙기반 모델인 의사결정나무와 확률적 모델인 NB 알고리즘을 통한 결과 값(정확도)을 서로 비교하였다.

트렌드 분석방법은 시계열 데이터마이닝 방법에 속한다. 시계열 데이터마이닝(time series data mining)은 시간 변화에 따라 기록한 데이터의 의미를 찾아내는 과정을 말하며, 주식시장의 일별 주식이 변동 데이터, 회사의 일별 매출액 데이터, 환자의 일별 치료 상황 데이터 등이 있다. 시계열 데이터에서 유용한 정보를 찾는 이 분석 방법은 크게 추세 분석, 유사성 분석, 주기성 분석, 순차패턴 분석 등이 있다.

추세(trend) 분석은 시간이 경과함에 따라 관측값이 어떤 경향을 갖는 경우로, 직선형태, 이차곡선의 형태 또는 S자 형태 등 다양한 추세유형이 있다. 추세 요인은 장기간에 걸쳐서 시계열에 나타나는 소비행태의 변동, 인구변동, 인플레이션 등의 경향을 파악하기 위한 것이라 할 수 있다. 유사성분석은 주식시장의 분석, 의학 데이터 분석, 공학 데이터 분석 등에 주로 이용되는데, 시계열 데이터 전체의 유사성 분석도 할 수 있지만, 부분 시계열만의 유사성만을 분석하는 경우도 있다.

주기성 분석은 시계열 데이터에서 주기적으로 발생하는 패턴들에 대한 탐색을 의미한다. 예를 들어, 시간별 전력 소비, 시간별 교통량 등은 주기적 패턴을 보인다고 할 수 있다. 김호성 등(2010)은 지하철 교통카드로부터 수집된 데이터를 이용하여 지하철 교통망에서 시간대별, 어느 장소에 유동인구가 밀집되는지를 직관적으로 확인할 수 있도록 시각화하였다. 장지훈(2009)은 모바일로 부터 수집된 시공간 데이터를 순차 패턴 마이닝(apriori algorithm)을 통해 개인별 빈번 패턴, 여러 사람 간 빈번 패턴 등을 분석하여 사용자들의 생활 패턴을 발견하고, 이에 대한 연관성 규칙을 만들어 시스템에 적용하였다. Vu *et al.*(2008)은 이동 객체 데이터(모바일 환경)로부터 모든 빈발한 패턴과 최대 패턴들을 탐색하기 위해 All_MOP, Max_MOP 알고리즘을 제안하였으며, 기존 알고리즘인 DFS_MINE와 비교평가를 수행하였다. 시간에 따라 순서화된 연관성의 조합을 이용하여 모든 가능한 빈발 패턴을 탐사한 후, 빈발 패턴 중 최대 빈발 패턴 탐사에 의한 시공간 규칙을 확인하였다.

회귀분석(regression analysis)기법은 마케팅 또는 교육 및 학습사이트 부분에서 많이 활용되고 있다. Jun(2002)은 인터넷 쇼핑을 사용자 행태분석을 위해 로그데이터를 활용하였다. 3개의 독립변수(구입 동기/계기, 체류시간, 빈도: 고객 체류 시간, 페이지 클릭수와 이벤트참여 여부)와 외재변수(인지

된 위험, 가격)를 통해 제품을 구매할 고객인지를 판단하는 회귀분석을 수행하였다. 조일현·김정현(2013)은 LMS에 수집된 로그데이터를 통해 학습자 시간관리 전략(학습시점 간격의 규칙성, 총 학습시간, 학습 접속 횟수)을 독립변수, 학습성취도를 종속변수로 하여 중다회귀분석을 사용하였다.

관련 연구들을 분석해본 결과, 시공간적 정보를 가진 데이터는 군집분석 및 트렌드 분석 방법을 주로 활용하는 것을 확인할 수 있었다(서정민 등, 2010; Nakaya and Yano, 2010). 연관규칙 방법은 공간적인 데이터 셋의 패턴 발견에 유용하기 때문에 자주 활용되는 방법이지만 공간패턴 구조를 발견하기 어렵고, 문제의 크기가 증가함에 따라 데이터가 기하급수적으로 증가하며, 사용된 모델에 크게 의존하고, 거의 발생하지 않는 항목 식별이 어렵다는 단점이 있었다.

국내에서는 시공간 데이터를 공간 데이터마이닝 또는 시계열 데이터마이닝에 초점을 맞춰서 연구한 사례들도 있으며, 최근에는 시공간을 동시에 고려한 마이닝 기법이 제안되고 있다. 박진형 등(2009)은 자동원격검침(AMR) 시스템에서 수집되는 전력 사용량 데이터의 특성을 분석하기 위해 시공간 데이터 마이닝 기법을 사용하였다. 고객의 전력 사용 목적에 따른 군집 분석을 위해 분할 군집화 기법을 적용하였고, 3차원 큐브 마이닝 기법을 사용하여 시간 속성과 공간 속성에 대한 패턴을 탐색하였다. 3차원 큐브 마이닝 기법을 적용하기 위해서 CubeMiner 분석기술을 사용하여 수집되는 속성들을 모두 만족하는 빈발패턴을 찾고자 하였다. 추가적으로 다양한 시간 도메인에서의 주기 패턴 발견을 위해 캘린더 패턴 마이닝 기법을 이용하였다. 비슷한 사례로 이현규 등(2009)은 GIS 및 시공간 마이닝 기술을 이용한 GIS 기반 새로운 우편 마케팅 기법을 제안하였다. 강남구와 송파구 지역의 내외부 데이터를 사용하여 'GeoLifestyle'을 세분화하기 위해 군집 분석(k-means 방법을 사용)을 하였고, 시공간 큐브 마이닝(Cube Miner 기법을 사용)을 통해 구매 및 소비 성향에 따른 시공간 패턴을 추출하였다. 앞의 연구와 마찬가지로 보다 정확한 미

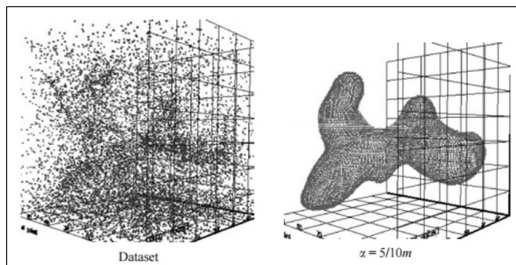


그림 1. 시 공간 데이터 마이닝 (Böhlen, 2003)

래 예측을 위해 캘린더 시간패턴기법을 추가로 적용하였다.

뿐만 아니라 해외에서는 시공간 데이터를 시각화함으로써 의미있는 지리적 패턴을 발견하고자 하는 연구들이 많았다. 또한, 시공간적 정보를 가진 데이터에 대해 데이터마이닝 기술을 적용하고자 할 때는 기존에 데이터마이닝의 대표적인 기법인 클러스터링 또는 연관규칙 방법을 공간적인 데이터 셋에 맞게 수정·적용하여 의미있는 정보를 발견하거나, 공간적 데이터마이닝 기술인 클러스터링 또는 연관규칙 방법 외에 시간적인 데이터마이닝 기술(트렌드 분석)을 일부 접목하여 시공간적 정보를 분석하고자 하는 연구도 있었다. 이외에도 Compieta *et al.*(2007)의 연구와 같이 새로운 분석 기법들(visual mining 등)을 통해 3차원 형태로 시각화하여 데이터를 보다 즉각적으로 상호작용하고(그림 1), 시공간적으로 탐색하여 분석하고자 한 사례도 있다.

IV. 로그데이터의 시공간 시각화 연구

로그데이터의 시공간적 시각화(그림 2)는 시각화를 통해 데이터의 특성을 분석할 수 있는 킷을 제공하기도 한다(Rinner, 2004). Miller(1991)는 시간지리학 개념에 기반하여 GIS의 3차원 표현기법을 활용하여 개인통행자료의 시공간적 패턴을 효과적으로 표현하는 지리적 시각화 환경을 개발하였다(그림 3). 기존의 시공간입방체(space-time cube¹⁾)를 개선하여 시공간경로(space-time path), 정거장과 활동꾸러미(station and activity bunch), 시공간프리즘(space-time prism)과 같은 시간지리학의 주요개념을 표현하였다. 시공간적 제약을 받는 인간활동이 시간의 경과에 따라 공간상에 남기는 자취

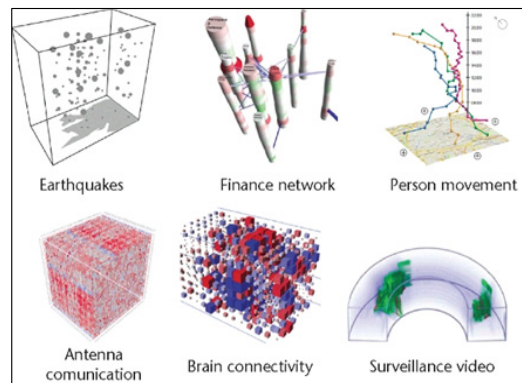


그림 2. 다양한 경로로 수집된 로그데이터의 3D 그래프 형태로 시각화한 사례 (Rinner, 2004)

를 시공간경로라고 지칭하였다. 사람의 이동경로는 주로 2차원 평면에 표시될 수도 있지만, 여기에 시간 축을 추가하여 이동경로상의 시간대별 위치를 3차원 공간상에 표시하면 시공간경로가 된다(그림 3a). 시공간경로에서 특정지점에 수렴이 발생할 경우, 이를 정거장이라 명명하였는데, 정거장은 예를 들어 학교, 집, 식당과 같이 일정 시간동안 활동이 집중되는 장소이며, 시공간경로 상에서는 시간 축에 거의 평행한 직선 형태로 표시하였다. 개별 시공간경로는 특정 지점에서 교차하는데, 이 교차점에 나타나는 원통(cylinder) 모양의 지점이 정거장에 해당하고, 여기에서 발생하는 활동들을 활동꾸러미라고 지칭하였다(그림 3b). 시공간프리즘은 이용가능시간 내에서 시공간경로 상에 표현되는 시공간적 범위를 말한다(그림 3c).

시간지리학에서의 시간은 거리와 마찬가지로 측정 가능한(time as distance) 개념으로서(Vasiliev, 1997), 시간에 따른 개인의 이동은 거리를 나타내는 x, y축과 시간을 나타내는 z축으로 구성된 3차원 공간상에 표현될 수 있다(그림 4). 다양한 분야에서 이러한 시공간적 데이터를 수집하여 이를 3차원으로 시각화하고 주요한 패턴을 연구한 논문들이 많이 있다. 주요한 몇 가지 사례 연구들을 살펴보면 다음과 같다.

Demšar and Virrantaus (2010)는 핀란드 만에서의 선박의 이동궤적 데이터(AIS: Automatic Identification System)를 시공간 밀도 형태로 시각화 하고 선박유형에 따른 시공간 패턴을 연구하였다. 2008년 하루 또는 한 달간의 AIS 데이터를 3D형태로 다양하게 시각화 하고 탐색하여 분석하였다. 하루 동안의 여객선 데이터를 전통적인 space-time cube와 밀도 분석이 추가된 space-time cube로 시각화하고, 장단점을 기술하였다. 또한, 한 달간의 여객선 이동경로를 시각화하여 기존에 쉽게 파악하기 어려웠던 지역별 이동 패턴이나 밀도정보를 손쉽게 시각화하고 패턴을 발견할 수 있음을 보여주었다(그림 5).

Fang and Lu(2011)는 텍사스 휴스턴 지역의 공기 오염 데이터를 space-time cube를 통해 시각화하고 이를 분석하였다. 두 가지의 방법-LUR (Land Use Regression)과 보간(spatial interpolation)-으로 하루 동안의 공기 오염(오존)을 시간대별로 시각화하고 정확도를 비교분석하였다.

재난 재해 분야에서도 로그데이터의 시공간적 분석을 시도한 연구들이 있다. Gatalisky *et al.*(2004)는 터키 서부의 Marmara 해에서 발생한 지진을 Space-time cube를 통해 시각화하고 시공간적 패턴을 분석하였다. Space-time cube를

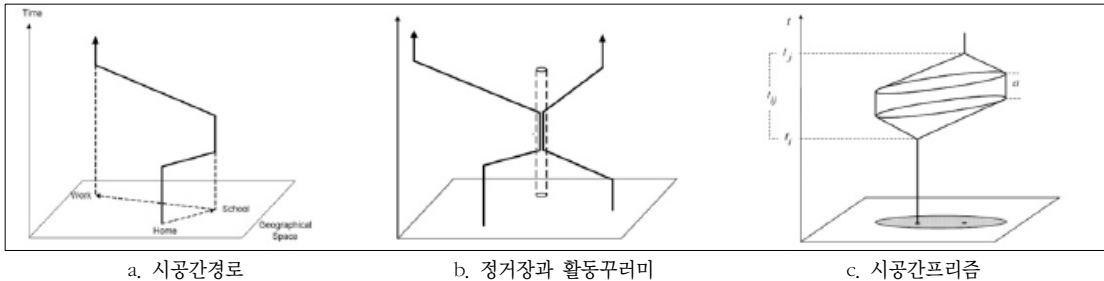


그림 3. 시간지리학의 주요개념 (Miller, 2005)

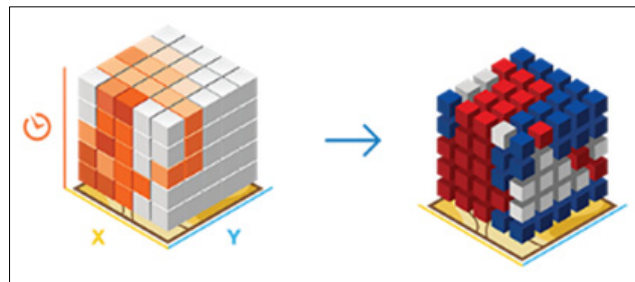


그림 4. ArcPro의 Space-Time Cube

(<http://desktop.arcgis.com/en/arcmap/10.3/tools/space-time-pattern-mining-toolbox/learnmorecreatethecube.htm>)

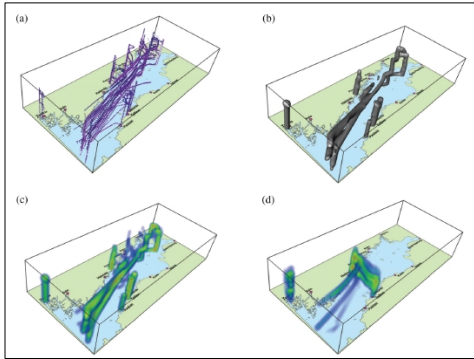


그림 5. 여객선의 이동 패턴을 3D로 시각화한 사례 (Demšar and Virrantaus, 2010)

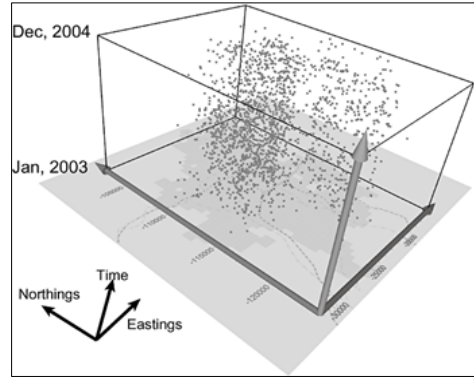


그림 7. 2003년 1월~2004년 12월까지의 소매치기 데이터를 시각화한 사례 (Nakaya and Yano, 2010)

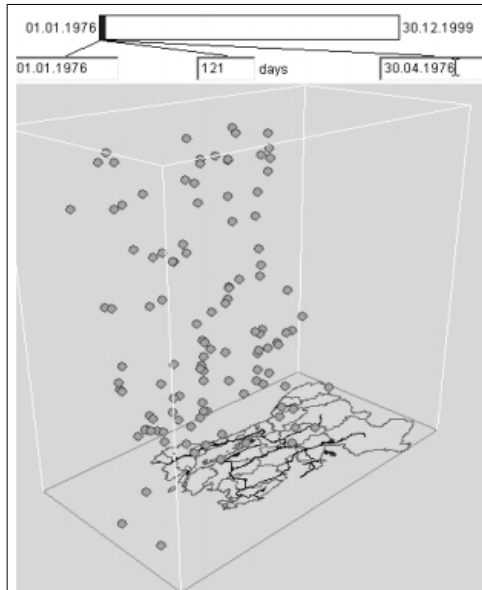


그림 6. 4개월간 지진데이터를 space-time cube를 사용하여 시각화한 사례 (Gatalasky et al., 2004)

통해 짧은 간격으로 가까운 지역에서 지속적으로 발생하는 지진을 탐색하고 시각화하였다. 그림 6은 1976년~1999년까지 Marmara(터키 서부) 지역에서 발생한 지진 데이터 10,550건을 보여준다. 기본적으로 지진은 순간적인 사건으로 인식하지만, 이 연구에서는 지진 발생에도 시공간적 패턴이 발견될 수 있음에 가정을 두고 수행되었다. Sim et al.(2013)은 재난 재해 및 범죄와 관련된 패턴 연구를 하였는데 시공간 스캔 통계(space-time scan statistic)를 3D로 시각화하여 범죄의 유형별 시공간적 패턴의 특성을 탐색하였다.

Nakaya and Yano(2010)는 2003년 1월~2004년 12월까지

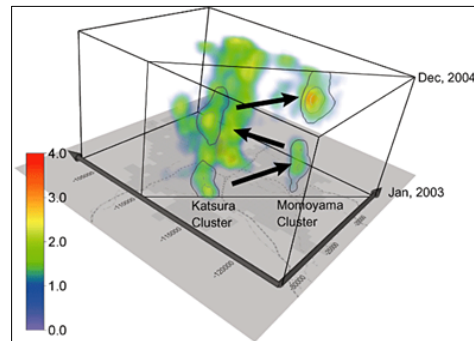


그림 8. 교토부 내의 범 죄 클러스터 발생 패턴을 시각화한 사례 (Nakaya and Yano, 2010)

의 소매치기 데이터(1,855건)를 시각화하였다(그림 7). 2D와 비교하여 3D로 각 포인트들을 표현하여도 의미있는 시공간적 패턴을 인식하기가 어렵다는 단점을 인식하고, 커널밀도 추정(KDE)과 시공간 스캔통계(Space-time scan statistics)를 통해 시공간적으로 높은 밀도의 클러스터가 발견되는지 확인하였다. 또한, 그림 8과 같이 3D형태의 space-time cube에 핫스팟 분석을 추가로 수행하여 교토부 내에서 가츠라 클러스터와 모모야마 클러스터 사이에 서로 교차적인 범 죄 클러스터가 발생하는 것을 확인하였다.

로그데이터의 시공간적 연구들을 살펴본 결과, 단순한 포인트 매핑은 빅데이터를 다루기에 어려울 뿐더러, 공간적 핫스팟이나 패턴을 발견하기 어려운 것으로 확인된다. 커널밀도추정(KDE)과 같은 핫스팟 분석 기법은 다양한 포인트 데이터의 밀도를 분석하는 가장 널리 알려진 기술이며, 포인트들 간의 밀도나 클러스터를 쉽게 확인할 수 있으나, 시간 데이터를 표현하기 어렵다는 단점이 있다. 이에 반해 3D 지리적 시

각화(3D geovisualization)는 공간적 패턴, 공간적 관계 및 트렌드를 보여줄 수 있다. 즉, 복잡한 시공간적 패턴을 직관적으로 이해하기 쉽게 표현할 수 있다는 장점이 있다. 3차원의 공간 데이터를 시각화하는 방법인 시공간 스캔통계(space-time scan statistic)는 잠재적인 클러스터의 원통으로 표현되며, 원통은 공간(위치)과 시간(높이)을 포함하는 지리적 차원으로 표현된다. 또 다른 방법인 시공간 입방체(space time cube)는 시공간 데이터를 큐브에 지도화하여 표현하는 시각화 기술(Kristensson *et al.*, 2009)이며, 시공간 밀도(space time density)와 같은 밀도 분석 및 핫스팟 분석을 통해 시공간적 패턴 발견을 용이하게 할 수 있다.

V. 결론

최근 다양한 채널로부터 일상생활에서 기록되는 로그데이터를 수집할 수 있게 되면서 다양한 유형의 로그데이터를 분석하여 기존에 알지 못했던 의미있는 정보를 발견하고자 하는 연구가 많이 이루어지고 있다. 로그데이터들은 시공간적인 정보를 함께 담고 있으며, 이러한 데이터를 다차원적으로 탐색하고 시각화 할 수 있다. 기존의 로그데이터 연구에서는 시공간적 차원을 동시에 고려한 연구는 부족했으며, 특히 로그데이터의 공간적 연구 및 시각화는 지리학 분야에서도 새롭게 연구될 수 있는 주제가 될 것이다.

따라서 본 연구에서는 최근에 시간을 고려한 공간 데이터의 다양한 시각화 및 데이터마이닝 기술에 대한 연구동향을 살펴보고, 크게 세 가지의 함의를 얻을 수 있었다.

첫째, 마케팅, 정보 검색 뿐만 아니라 의료, 교육 분야 등 다양한 분야에서 로그데이터를 분석하고자 하는 연구들이 이루어지고 있는 것을 알 수 있었다. 특히, 최근에 모바일과 웹 사용의 증가로 쉽게 개인차원의 정보(personal data) 수집이 가능해졌고, 개인의 통행 패턴, 개인의 성향이나 사회적 관계 등을 이용한 서비스를 수행하는 분야에서 이러한 연구가 많이 이루어지는 것을 알 수 있었다.

둘째, 실시간에 가까운 데이터들이 기록되는 로그데이터는 많은 양의 데이터가 기록되며, 이러한 대용량의 시공간적 정보를 가진 로그데이터를 최근에는 시공간을 동시에 고려한 데이터마이닝 기법을 적용하고자 한 연구들이 등장하고 있으며, 기존의 데이터마이닝 방법들을 다양하게 접목하거나 새로운 분석기법들을 통해 데이터를 보다 즉각적으로 상호작용하고, 시공간적으로 탐색하여 분석하는 연구들도 많이 이루어

지고 있음을 알 수 있었다. 특히, 공간정보 분야에서는 시공간적 정보를 가진 데이터에 대해 주로 군집분석이나 트렌드 분석 방법을 활용하는 것을 확인할 수 있었다. 기존에도 공간 정보에 대해 군집분석을 시도한 사례가 있었고, 시간적 패턴을 확인하고자 할 때 주로 트렌드 분석방법을 적용한 사례가 있었다. 또한, 시공간 데이터를 동시에 분석하고자 할 때, 각 분석방법의 장점을 살려서 새로운 분석 방법인 3차원 큐브마이닝 기법, visual mining 등을 적용한 사례도 확인할 수 있었다. 다른 데이터마이닝 기법에 비해 이 방법들이 시공간 패턴을 확인하기에 유용한 방법임을 알 수 있었다.

셋째, 로그데이터가 가지는 시공간적 정보를 다차원적으로 탐색하고 시각화 하고자 하는 연구들이 증가하고 있음을 알 수 있었다. 특히 기존에 2D로 지리적 시각화(geovisualization)하여 표현할 때는 시간적 정보나 클러스터 정도를 확인하기 쉽지 않았던 반면, 3D 지리적 시각화(3D geovisualization)는 공간적 패턴, 공간적 관계 및 트렌드를 보여줄 수 있다. 즉, 복잡한 시공간적 패턴을 직관적으로 이해하기 쉽게 표현할 수 있다는 장점이 있었다.

최근에 많은 연구들에서 로그데이터를 수집하여 시공간적으로 시각화할 수 있는 시스템을 개발하거나 기존의 툴들을 사용하여 시각적으로 표현하고 그 특징을 발견하는 것만으로도 충분한 의미가 있음을 시사하고 있다. 그리고 발견된 특징이나 주요한 패턴이 있다면 향후 다양한 분야에서 이를 활용하고 또 다른 연구를 수행하는데 기반이 될 수 있음을 보여주고 있다.

註

- 1) 시공간 입방체(Space-Time Cube)를 시각화 하고 분석할 수 있는 소프트웨어는 대표적으로 ESRI 社の ArcPro 소프트웨어에서 제공하는 시공간 입방체 툴(create space-time cube, Visualizing the Space Time Cube in 2D or 3D)이나 GeoTime 소프트웨어, 또는 파이프로세싱(pyprocessing) 등이 있다.

참고문헌

- 강만오·김상락·박상무, 2012, “빅 데이터가 여는 미래의 세상,” 정보과학회지, 30(6), 25-32.
 김석기·안정용·한경수, 2001, “웹 로그(WEB LOG) 데이터

- 분석 방법에 관한 연구,” 응용통계연구, 14(2), 261-271.
- 김윤경, 2002, “효과적인 Cyber 교육 시스템 구축을 위한 Web Log 분석 활용 방안 연구,” 이화여자대학교 석사학위논문.
- 김호성·박중수·이금숙, 2010, “서울 수도권 지하철 교통망 승객 흐름의 시각화,” 한국콘텐츠학회논문지, 10(4), 397-405.
- 박진형·이현규·신진호·류근호, 2009, “GIS-AMR 시스템에 서 시공간 데이터마이닝 기법을 이용한 전력 소비 패턴의 분석 및 예측,” 정보처리학회논문지D, 16(3), 307-316.
- 변시우, 2002, “Log data 분석을 통한 웹사이트의 효율적 관리 방안-ABC Jeans 사 사례 분석을 중심으로,” 인터넷비즈니스 연구, 3(1), 99-124.
- 서정민·문형만·박창선·황정환·김병주, 2010, “모바일 센서네트워크를 활용한 장애인 라이프로그의 분석 및 위험 감지 시스템에 관한 연구,” 한국컴퓨터정보학회, 18(2), 311-314.
- 성병준·유환희, 2014, “진주시 교통사고 공간적 군집분석,” 2014 한국지형공간정보학회 춘계학술대회, 169-172.
- 이건학, 2004, “GIS 와 공간 데이터마이닝을 이용한 교통사고의 공간적 패턴 분석,” 대한지리학회지, 39(3), 457-472.
- 이기용·정규만, 2014, “MapReduce 기술을 이용한 게임 로그 데이터 분석 기법,” 한국컴퓨터게임학회논문지, 27(1), 19-25.
- 이정진, 2011, 「R, SAS, MS-SQL을 활용한 데이터마이닝」, 파주: 자유이카데미.
- 이현규·최용훈·정 훈·박중홍, 2009, “GIS-우편 마케팅 시스템에서 Geo-Lifestyle 군집화 및 시공간 데이터 큐브를 이용한 구매, 소비 성향 예측,” 한국공간정보시스템학회 논문지, 11(4), 80-89.
- 장지훈, 2009, “시공간 데이터를 이용한 데이터마이닝,” 서울대학교 박사학위논문.
- 정성훈·김하나·신영상·이태진·김휘강, 2015, “결제로그 분석 및 데이터 마이닝을 이용한 이상거래 탐지 연구 조사,” 정보보호학회논문지, 25(6), 1525-1540.
- 조일현·김정현, 2013, “학습분석학을 활용한 e-러닝 학습성과 추정 모형의 통계적 유의성 확보 시점 규명,” 교육공학연구, 29(2), 285-306.
- 최승배·강창원·조장식, 2009, “웹 로그데이터를 이용한 대학 입시 지원자 형태 분석,” 한국데이터정보과학회지, 20(3), 493-504.
- Agrawal, R. and Srikant, R., 1994, Fast algorithms for mining association rules, *Proceedings, 20th International Conference Very Large Data Bases VLDB*(Vol. 1215), September, 487-499.
- Andrienko, N., Andrienko, G., and Gatalsky, P., 2003, Exploratory spatio-temporal visualization: an analytical review, *Journal of Visual Languages and Computing*, 14(6), 503-541.
- Bogomolov, A., Lepri, B., and Pianesi, F., 2013, Happiness recognition from mobile phone data, *Proceedings, Social Computing(SocialCom), 2013 International Conference*, September, 790-795.
- Bogomolov, A., Lepri, B., Ferron, M., Pianesi, F., and Pentland, A.S., 2014, Pervasive stress recognition for sustainable living, *Proceedings, Pervasive Computing and Communications Workshops(PERCOM Workshops) 2014 IEEE International Conference*, March, 345-350.
- Böhlen, M., Bukauskas, L., Eriksen, P.S., Lauritzen, S.L., Mažeika, A., Musaeus, P., and Mylov, P., 2003, 3D visual data mining-goals and experiences, *Computational Statistics & Data Analysis*, 43(4), 445-469.
- Chittaranjan, G., Blom, J., and Gatica-Perez, D., 2013, Mining large-scale smartphone data for personality studies, *Personal and Ubiquitous Computing*, 17(3), 433-450.
- Compieta, P., Di Martino, S., Bertolotto, M., Ferrucci, F., and Kechadi, T., 2007, Exploratory spatio-temporal data mining and visualization, *Journal of Visual Languages & Computing*, 18(3), 255-279.
- Demšar, U. and Verrantaus, K., 2010, Space-time density of trajectories: exploring spatio-temporal patterns in movement data, *International Journal of Geographical Information Science*, 24(10), 1527-1542.
- Fang, T.B. and Lu, Y., 2011, Constructing a near real-time space-time cube to depict urban ambient air pollution scenario, *Transactions in GIS*, 15(5), 635-649.
- Ferreira, N., Poco, J., Vo, H.T., Freire, J., and Silva, C.T., 2013, Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips, *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2149-2158.
- Gatalsky, P., Andrienko, N., and Andrienko, G., 2004, Interactive analysis of event data using space-time cube, *Proceedings, Information Visualisation 8th International*

- Conference* 2004, July, 145-152.
- Hwang, M.G., Lee, S.M., and Seo, J.M., 2012, A Study on Life-log analysis and monitoring system for disabled person using smart media, *Journal of the Korea Society of Computer and Information*, 17(8), 99-106.
- Jun, H.C., 2002, A study on web shopping mall user behavior using log files: W. cosmetic mall case, Unpublished Master Dissertation, Department of Business Administration, KAIST.
- Kim, J.K., 2014, A Novel Recommendation Approach Based on Chronological Cohesive Units in Contents Consuming Logs, Unpublished Ph.D. Dissertation, Department of Electricity and Electronic Engineering, Sungkyunkwan University.
- Kim, J.H. and Bae, J.S., 2015, Diagnosis analysis inpatient log data based on process mining, *Proceedings, the Spring Conference on Korean Institute of Industrial Engineers*, 1027-1038.
- Kristensson, P.O., Dahlbäck, N., Anundi, D., Björnstad, M., Gillberg, H., Haraldsson, J., and Ståhl, J., 2009, An evaluation of space time cube representation of spatiotemporal patterns, *IEEE Transactions on Visualization and Computer Graphics*, 15(4), 696-702.
- Koperski, K., Adhikary, J., and Han, J., 1996, Spatial data mining: progress and challenges survey paper, *Proceedings, ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery*, June, Montreal, Canada, 1-10.
- LiKamWa, R., Liu, Y., Lane, N.D., and Zhong, L., 2011, Can your smartphone infer your mood, *Proceedings, PhoneSense Workshop*, November, 1-5.
- Mafrur, R., Nugraha, I.G.D., and Choi, D., 2015, Modeling and discovering human behavior from smartphone sensing life-log data for identification purpose, *Human-centric Computing and Information Sciences*, 5(1), 1-18.
- Miller, H.J., 1991, Modelling accessibility using space-time prism concepts within geographical information systems, *International Journal of Geographical Information System*, 5(3), 287-301.
- Miller, H.J., 2005, A measurement theory for time geography, *Geographical Analysis*, 37(1), 17-45.
- Min, J.K. and Cho, S.B., 2011, Mobile human network management and recommendation by probabilistic social mining, *IEEE Transactions on Systems, Man, and Cybernetics, Part B(Cybernetics)*, 41(3), 761-771.
- Min, J.K., Wiese, J., Hong, J.I., and Zimmerman, J., 2013, Mining smartphone data to classify life-facets of social relationships, *Proceedings, 2013 Conference on Computer Supported Cooperative Work*, February, 285-294.
- Montoliu, R., Blom, J., and Gatica-Perez, D., 2013, Discovering places of interest in everyday life from smartphone data, *Multimedia Tools and Applications*, 62(1), 179-207.
- Montoliu, R. and Gatica-Perez, D., 2010, Discovering human places of interest from multimodal mobile phone data, *Proceedings, 9th International Conference on Mobile and Ubiquitous Multimedia*, December, 12.
- Muaremi, A., Arnrich, B., and Tröster, G., 2012, A survey on measuring happiness with smart phones, *Proceedings, 6th International Workshop on Ubiquitous Health and Wellness(UbiHealth 2012)*.
- Nakaya, T. and Yano, K., 2010, Visualising crime clusters in a space-time cube: an exploratory data-analysis approach using space-time kernel density estimation and scan statistics, *Transactions in GIS*, 14(3), 223-239.
- Nicholas, D., Clark, D., Jamali, H.R., and Watkinson, A., 2014, Log usage analysis: what it discloses about use, information seeking and trustworthiness, *International Journal of Knowledge Content Development and Technology*, 4(1), 23-37.
- Park, N.H. and Joo, K.H., 2013, Log based keyword extraction and spread based clustering for an efficient information searching, *International Journal of Software Engineering and Its Applications*, 7(6), 201.
- Pierleoni, P., Pernini, L., Belli, A., and Palma, L., 2014, An android-based heart monitoring system for the elderly and for patients with heart disease, *International Journal of Telemedicine and Applications*, 10.
- Rimmer, C., 2004, Three-dimensional visualization of activity-travel patterns, *Proceedings, the Munster GI Days*, 231-237.
- Seneviratne, S., Seneviratne, A., Mohapatra, P., and Mahanti, A., 2014, Predicting user traits from a snapshot

- of apps installed on a smartphone, *ACM SIGMOBILE Mobile Computing and Communications Review*, 18(2), 1-8.
- Shin, S.M., Yi, T.S., and Kim, H.S., 2014, An analysis of a convergence utilization field of technology based on usage log in information service, *Proceedings, International Conference on Convergence Technology*, July, 967-968.
- Sim, S., Walker, W.C., Cook, J.R., Doyle, R., and Keys-Mathews, L., 2013, Exploratory spatial-temporal visualization of hurricane impacts on crime events in miami, Florida, *Proceedings, 26th International Cartographic Conference(ICC)*.
- Singh, V.K., Freeman, L., Lepri, B., and Pentland, A.S., 2013, Predicting spending behavior using socio-mobile features, *Proceedings, Social Computing(SocialCom) 2013 International Conference*, September, 174-179.
- Sujatha, V. and Punithavalli, A., 2012, Improved user navigation pattern prediction technique from web log data, *Procedia Engineering*, 30, 92-99.
- Thiagarajan, A., Ravindranath, L., LaCurts, K., Madden, S., Balakrishnan, H., Toledo, S., and Eriksson, J., 2009, VItrack: accurate, energy-aware road traffic delay estimation using mobile phones, *Proceedings, 7th ACM Conference on Embedded Networked Sensor Systems*, November, 85-98.
- Vasiliev, I., 1997, Mapping time, *Cartographica: The International Journal for Geographic Information and Geovisualization*, 34(2), 1-51.
- Vu, T.H.N., Lee, B.J., and Ryu, K.H., 2008, Discovery of frequent sequence pattern in moving object databases, *The KIPS Transactions: Part D*, 15(2), 179-186.
- Yang, Y. and Webb, G.I., 2002, A comparative study of discretization methods for naive-bayes classifiers, *Proceedings, PKAW(Vol. 2002)*.
- 교신: 강영욱, 03760, 서울시 서대문구 이화여대길 52, 이화여자대학교 사회과학교육과 지리전공(이메일: ykang@ewha.ac.kr)
- Correspondence: Youngok Kang, Department of Social Studies, Ewha Womans University, 52 Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Republic of Korea (Email: ykang@ewha.ac.kr)
- 투 고 일: 2016년 11월 7일**
심사완료일: 2016년 11월 30일
투고확정일: 2016년 12월 1일

