

# 개인의 위치에 기반하는 장소 방문 선호도를 추출하는 시스템에 관한 연구

송하윤 김도연  
 홍익대학교 컴퓨터공학과  
[hayoon@hongik.ac.kr](mailto:hayoon@hongik.ac.kr), [dodo@g.hongik.ac.kr](mailto:dodo@g.hongik.ac.kr)

## A System on Establishing a Favored Location Measurement based on Personal Geopositioning Data

Ha Yoon Song, Do-Yeon Kim  
 Dept. of Computer Engineering, Hongik University

### 요 약

DB로부터 위치 데이터를 추출해오면, location visiting measure를 통해 한 개인의 장소 방문 선호도를 측정한다. 이 과정에서 위치 데이터 클러스터링 및 개인의 이동성 파악을 위한 6가지 조치를 한다. 위치 데이터 분석 결과로 얻은 이동성 통계는 LBS(Location Based System)를 포함한 다양한 분야에 적용될 수 있다. 연구의 최종 목표는 새로운 위치 데이터의 수집에 따른 방문 선호도 변화를 적용할 수 있도록 하는 location visiting measure의 자동화 시스템을 구축하는 것이다.

### 1. 서론

최근 GPS, 스마트폰, 스마트 위치와 같은 디바이스를 통한 위치 확인 매커니즘이 급속도로 발전하고 있다. 이를 통해 개인의 위치 데이터를 수집하는 것이 너무도 간단하고 쉬워졌다. 수집한 위치 데이터의 분석을 바탕으로 특정 장소에 대한 선호도와 같은 개인의 이동성 패턴을 알 수 있는 이동성 모델이 생긴다면, 학문 및 산업 분야에서 활용이 가능할 것이다.



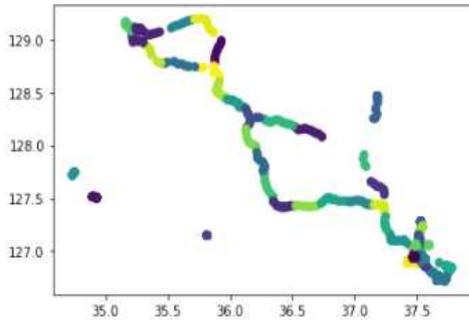
[그림 1] 연구 전체 흐름도

연구에서 작성될 시스템의 전체적인 흐름도는 [그림 1]과 같다. 이동데이터(Geopositioning Data)는 스마트폰, GPS 수신기, IoT 디바이스로 수집되고 DB에 저장된다. 저장된 위치 데이터를 불러와 Location Visiting Measure를 진행하여서 한 개인의 방문 빈도수, 방문 기간과 같은 이동성 통계를 내어 방문 선호도를 측정하고, 측정 결과 또한 DB에 저장한다. 추가로 visualizer를 통해 지도상에 분석 결과를 시각화해 볼 수 있다.

### 2. 이동데이터 추출 및 클러스터링

스마트폰이나 다른 IoT 디바이스를 통해 수집된 개인의 위치 데이터는 Data preprocessing 과정을 거친 후 DB에 들어간다. DB로부터 추출한 위치 데이터를 클러스터링 코드에 접목해 클러스터링을 진행한다. 클러스터링 결과로는 사용자 이름, 클러스터 개수, 위도, 경도, 최대 거리, 평균 거

리, 머문 시간, 요소 개수 등을 알 수 있다. 현재 관련 분야의 연구자가 이동데이터를 수집하여 클러스터링한 결과를 다시 DB에 저장하는 과정을 자동화하는 것에 대한 연구를 진행 중이어서 location visiting measure에는 적용하지 않았다.



[그림 2] 전국적으로 분포되어있는 위치데이터 모습

데이터는 홍익대학교 연구실 소속 연구자가 수집한 과거 2013년 1월부터 12월까지의 1년 치 이동 데이터를 사용했고 현재 DB에 저장되어있다. 프로그래밍 언어는 Python을 사용했고, DB로부터 경도, 위도와 같은 필요한 정보들을 추출해 K-Means 클러스터링을 진행했다. [그림 2]는 구현한 결과를 지도에 나타낸 것으로 해당 데이터가 전국적으로 넓게 분포되어 있음을 확인했다.



[그림 3] 동일 위치에 해당하는 여러 개의 클러스터

[그림 3]에 나타나 있는 동일 장소의 여러 개의 클러스터는 서로 다른 시간에 동일 장

소를 여러 번 방문했음을 의미한다. 개인이 클러스터에 상주하는 기간과 얼마나 자주 클러스터를 방문하는지에 대한 정보는, 클러스터 특정 영역의 방문 기간 및 방문 빈도를 분석함으로써 얻을 수 있다.

### 3. 장소 방문도 측정 및 이동성 통계

개인의 이동성에 대한 통계는 선행 연구 논문에 따라 다음 6가지의 기본적인 조치를 사용한다.

#### A. 지역 내 위치 방문 빈도(LVFIA)

LVF(위치 방문 빈도) = 위도, 경도가 동일한 특정 위치1의 카운트 (1)

CVF(클러스터 방문 빈도) = 위치1을 포함한 클러스터 내 총 위치 데이터 수 (2)

$$LVFIA = \frac{LVF}{CVF} \quad (3)$$

#### B. 위치 방문 빈도 순위(ROLVF)

TLVF(총 위치 방문 빈도) = 총 위치 수 (전체 클러스터 수) (4)

$$ROLVF = \log \frac{TLVF}{LVF} \quad (5)$$

#### C. 지역 방문 빈도 순위(ROAVF)

$$ROAVF = \log \frac{TLVF}{CVF} \quad (6)$$

#### D. 내부지역 방문기간(LVDIA)

LST(위치 체류시간) = 특정위치 체류시간 (7)

CST(클러스터 체류시간) = 위치 1을 가지는 클러스터 내 모든 위치별 총 체류시간 (8)

$$LVDIA = \frac{LST}{CST} \quad (9)$$

#### E. 위치 방문 기간 순위(ROLVD)

TLST(총 위치 체류시간) = 기존 클러스터 내 모든 위치 총 체류시간 (10)

$$ROLVD = \frac{TLST}{LST} \quad (11)$$

F. 지역 방문 기간 순위(*ROAVD*)

$$ROAVD = \log \frac{TLST}{CST} \quad (12)$$

LVFIA, ROLVF, LVDIA, ROLVD는 경도 및 위도와 같은 위치를 기반으로 하고, ROAVF, ROAVD는 위치 클러스터를 기반으로 계산한다. 앞서 언급된 6가지 기본 조치의 일부 조합은 의미가 있다. LVFIA × ROLVF 와 LVDIA × ROLVD의 두 가지 조합을 추가로 설정할 수 있다. 특정 위치1에 대한 이러한 조합은 순위와 빈도를 곱하고 순위와 방문 기간을 곱하는 것이기 때문에 위치1의 중요성을 의미한다. 이러한 조합이 높을수록 모든 위치 중에서 개인에게 더 중요한 위치1이 되게 된다.

위의 조치를 보유하고 있는 실제 데이터 세트에 적용하여 이동성 통계를 내기 위한 실험을 진행했다. 우선 위치데이터를 클러스터링 한 결과를 분석해보았다.

C	latitude	longitude	
16	37.548005	126.919997	1
	37.548040	126.919042	1
	37.548100	126.918075	1
	37.548217	126.922820	1
	37.548260	126.919250	1
	37.548292	126.917288	1
	37.548300	126.919972	1
	37.548325	126.922953	1
	37.548355	126.917062	1
	37.548360	126.919660	1
	37.548375	126.916987	1
	37.548385	126.916982	1
	37.548390	126.916985	2
	37.548395	126.916987	1

[그림 4] 클러스터16에 해당하는 위치 정보

[그림 4]는 클러스터 번호 16에 해당하는 위치 데이터의 일부이다. 대부분이 1개에서

2개의 위치가 겹치고 있는데, 이는 어떤 장소로 이동하는 과정에서 수집된 데이터들을 의미한다. 이러한 현상은 한 개인의 장소 방문 선호도를 측정하는데 불필요한 정보라고 할 수 있다. 보유하고 있는 데이터에 불필요한 정보들이 일부 비중을 차지하고 있음을 확인하였고, 이러한 데이터를 사전에 제거하는 전처리 작업이 필요할 것으로 판단된다.

4. 기대 효과 및 향후 계획

방문 빈도수 및 지속시간 측정과 같은 개인의 장소 방문 선호도에 대한 통계는 위치기반서비스(LBS)를 구축하는데 활용할 수 있다. LBS는 이동통신망을 기반으로 사람이나 사물의 위치정보를 활용한 모든 응용 소프트웨어 서비스를 의미한다. LBS는 빅데이터 분석, 안전, 교통정보, 지도 및 주변 정보, 광고 마케팅, 엔터테인먼트, IoT 서비스 등이 있다.

향후 해결해야 할 과제는 우선 무의미한 데이터를 제거했을 때, 남은 데이터의 양이 너무 적다는 것이다. 전국보다는 좁은 범위에서 장기간 수집된 데이터 세트를 사용하여 실험을 끝까지 진행된다면, 이동성 통계에서 더욱더 유의미한 결과를 볼 수 있을 것으로 생각된다. 또한, DB와의 연결 관련 문제도 해결해야 한다. DB에 새롭게 데이터가 들어오며 따라 이동성 통계가 갱신되어 새로운 결과를 재생산 후 DB에 저장하는 과정을 자동으로 처리하는 시스템 구축에 관한 연구가 추가로 필요하다.

이 연구는 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행됨 (NRF-2019R1F1A1056123)

참고문헌

[1] Dong Yun Choi & Ha Yoon Song (2021),  
“A Method to Measure the Degree of the  
Favorite Location Visiting of Mobile Objects”,  
JSCE, Volume 15, pp. 47-57

[2] Hyunuk Kim, Ha Yoon Song (2011),  
“Daily Life Mobility of a Student: From  
Position Data to Human Mobility Model  
through Expectation Maximization Clustering”,  
FGIT-MulGraB, pp. 88-97

[3] Ha Yoon Song (2016), “Probabilistic  
Space-Time Analysis of Human Mobility  
Pattern”, WSEAS TRANSACTIONS on  
COMPUTERS, Volume 15, pp. 223-238