

이동데이터 시간분석을 통한 이동양태 파악

홍수찬*, 송하윤*

*홍익대학교 컴퓨터공학과

e-mail : suchanhong@naver.com

hayoon@hongik.ac.kr

Investigating Cyclic Pattern of Mobility through Analysis of Geopositioning Data

Suchan Hong *, Ha Yoon Song *

*Dept. of Computer Science, Hongik University

요 약

사람은 한 장소를 방문할 때 순환 패턴이 있으며, 이 패턴에 여러 싸이클의 경향이 있다. 요즘은 스마트폰 및 기타 휴대용 장치로 개인 이동성 데이터를 수집하는 것이 가능하다. 이러한 장치는 다양한 위치 데이터를 수집하고 여러가지 방법으로 분석할 수 있게 해준다. 위치 수집기를 기반으로 지구 위치 데이터에서 추출된 사람의 이동성 모델을 수립하고, 위치 클러스터를 방문자의 순환 패턴을 조사할 수 있다. 수년 동안 수집된 개인의 이동성 모델을 토대로 클러스터 재방문 시간을 계산 후 분석하여 그래프로 시각화하였다. 시간 순서의 위치 클러스터와 방문 클러스터에 대한 위치 데이터는 1 분 단위로 측정된다. 전체 데이터 방문 횟수는 15 분마다 정규화하고, 자원 봉사자의 다양한 지리적 위치 데이터 셋에 대해 방문의 순환 패턴은 자기 상관, 자기 공분산 및 재방문 시간으로 살펴볼 수 있다.

1. 서론

최근 4 차 산업혁명과 IoT, 스마트 시티가 대두되고 있고 이로 인하여 인간의 위치정보를 기반으로 한 인간의 이동 모습에 관한 모델 또한 필요해졌다. 인간의 이동패턴에 관한 모델을 다양한 산업 환경에 적용함으로써 이윤을 창출하려는 노력이 다양해지고 있고, 이에 대한 기초연구가 중요시 되었다. 기술의 발달로 인하여 휴대용 위치정보 수집 장비와 스마트폰의 사용, IoT 에 대한 관심이 대중화되면서 인간의 위치정보를 수집하는 일이 점점 용이해졌고, 수집한 인간의 이동정보를 통해 인간의 이동패턴을 모델화할 수 있었다.

모델링 한 인간의 이동정보를 기반으로 개인의 위치와 군집간의 관계를 파악이 가능해졌다. 인간은 시간에 따라 한 장소에 머물러있거나 다른장소로 이동하는 형태를 가지고 있다. 인간의 이동정보에 시간이 포함되어 있기 때문에 인간의 위치데이터와 시간을 활용하여 위치에 재방문하는 시간에 대해 분석하고자 하였다.

본 논문에서는 인간의 이동데이터를 시간 분석하여 이동 양태를 파악하기 위해 일정기간 동안 수집한 자원봉사자 CDY 의 위치정보와 수집한 위치정보를 사용하여 만든 개별 통합 이동성 모델을 사용하였다. 위치정보는 <날짜, 시간, UNIX time, 위도, 경도>로 이

루어져있고, 개별 통합 이동성 모델의 정보는 <군집 번호, 군집 중심의 위도, 군집 중심의 경도, 군집의 반지름>으로 이루어져있다. 모든 위치 정보 중 군집 내에 위치한 데이터를 바탕으로 시간에 따른 이동 양태를 분석하였다. 자원봉사자 CDY 는 위치정보를 수집하기 위하여, Sports Tracker[7]앱과 Garmin[8]에서 제작한 GPS 수집기를 이용하였다.

본 논문은 다음과 같이 구성된다. 2 장에서는 위치 데이터를 사용한 기존연구들에 대해서 설명하며 3 장에서는 시간 분석을 통한 이동양태 예측 방법에 대해서 소개할 것이다. 이어서 4 장에서는 이동 양태 예측방법을 사용하여 나온 실험 결과에 대해서 설명하며 5 장에서는 결론에 대해 서술하면서 논문을 마무리 할 것이다.

2. 기존연구

인간의 위치에 관한 연구는 여러 가지 특징과 관련해서 지속적으로 연구 해왔다. 기존에 진행된 연구로는 EM (Expectation-Maximization) algorithm[5]을 통해서 클러스터 즉, 군집을 생성하는 연구[1,6]와 개별 이동성 모델을 기반으로 집단 이동성 모델을 생성하는 연구[2], 사람의 성격과 자주 방문하는 장소의 상관관계 [3], 위치와 군집 간의 분석[4] 등 여러 연구가 존재하였지만 인간의 이동 데이터와 시간에 대한 연구는 진

행되지 않았다. 따라서 시간과 관련하여 인간의 이동 패턴을 찾는 본 연구는 기존에 시도하지 않았다는 점에서 큰 의미가 있다.

3. 시간분석을 통한 이동양태 예측

본 연구에서 실시한 위치데이터를 통한 인간의 이동 패턴을 분석하기 위해 시간이라는 데이터를 활용하여 상관관계를 분석하려한다. 시간에 따라서 한 장소(군집)에 머물러 있거나 다른 장소로 이동하는 특징을 기반으로 시간과 장소에 따른 상관관계를 파악하려한다.

시간과 장소에 따른 상관관계가 존재하는지 파악하기 위해서 군집에 머물러 있는 시간대에 대해 일정한 규칙이 있는지 파악하는 것이 필요해졌다. 따라서 기존에 진행된 연구 중 EM 알고리즘을 통해 나타난 개별 통합 이동성 모델을 가지고 군집을 추출해 내었고 이중 군집에 방문하는 시간대를 확인하였다. 결과적으로 일정 시간대에 방문한다는 것을 확인하였고 연구를 진행하였다

시간에 따른 위치의 상관관계를 따져보기 위하여 Autocorrelation 과 Autocovariance 분석을 사용하였다. Autocorrelation 은 동일한 하나의 시계열 내에서 시간변수(τ)와의 상관계수를 측정하는 것으로 수식은 다음과 같다.

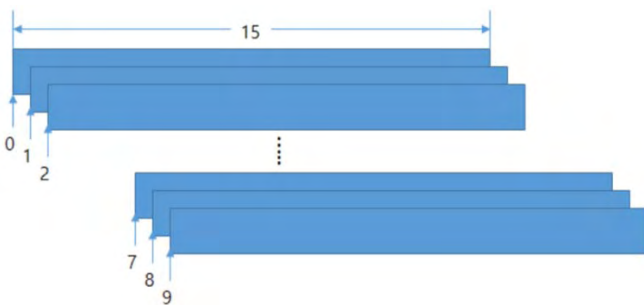
$$R(t, \tau) = E[X(t) \cdot X(t + \tau)]$$

Autocovariance 또한 동일한 하나의 시계열 내에서 시간변수(τ)와의 공분산을 측정하는 것으로 수식은 다음과 같다.

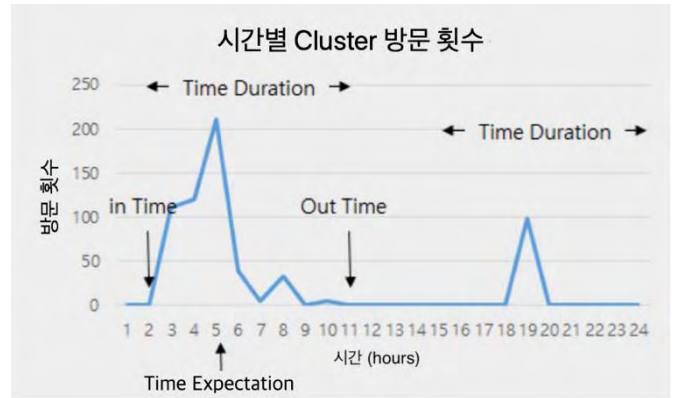
$$C_X(t, \tau) = Cov[X(t), X(t + \tau)]$$

$$C_X(t, \tau) = R_X(t, \tau) - \mu_t \mu_{t+\tau}$$

Autocorrelation 과 Autocovariance 를 구하기 위해 일정한 시간 배열을 사용해야한다. 기존에 군집을 생성하는 연구에서 15 분 이상 머물러있을 경우에 군집에 포함하였기 때문에 time slice 는 15 분으로 가정하고 연구를 진행 하였다. Time slicing 을 그래프로 그리면 다음과 같다.



(그림 1) Time Slice 에 관한 대략적 그림



(그림 2) 시간별 Cluster 방문 횟수에 따른 In Time, Out Time, Time Expectation 의 정의

위 그림에서 나온 $E[X(t)]$ 와 $E[X(t + \tau)]$ 를 사용하여 autocorrelation 과 autocovariance 를 계산한다.

본 연구를 진행하는데 사용된 데이터는 2013 년 5 월 ~ 2017 년 11 월의 이동 데이터이다. 데이터의 양이 방대하여, 각각의 correlation 과 covariance 를 구하기 위한 연산시간이 길어졌고, 이를 해결하기 위하여 병렬 프로그래밍, 즉, GPGPU 프로그래밍을 사용하여 연산 시간을 단축시켰다. 실험에 사용된 컴퓨터 사양은 인텔 i7 CPU, 16GB RAM, nVidia 980Ti GPU 이다.

4. 재방문 시간과 군집 간의 관계에 대한 분석

우선 재방문 시간에 따른 의미 여부를 확인하기 위하여 상관계수 값에 대하여 우선 분석하였다. 상관계수의 값에 대한 의미에 대하여 분석하면 다음 표처럼 해석할 수 있다.

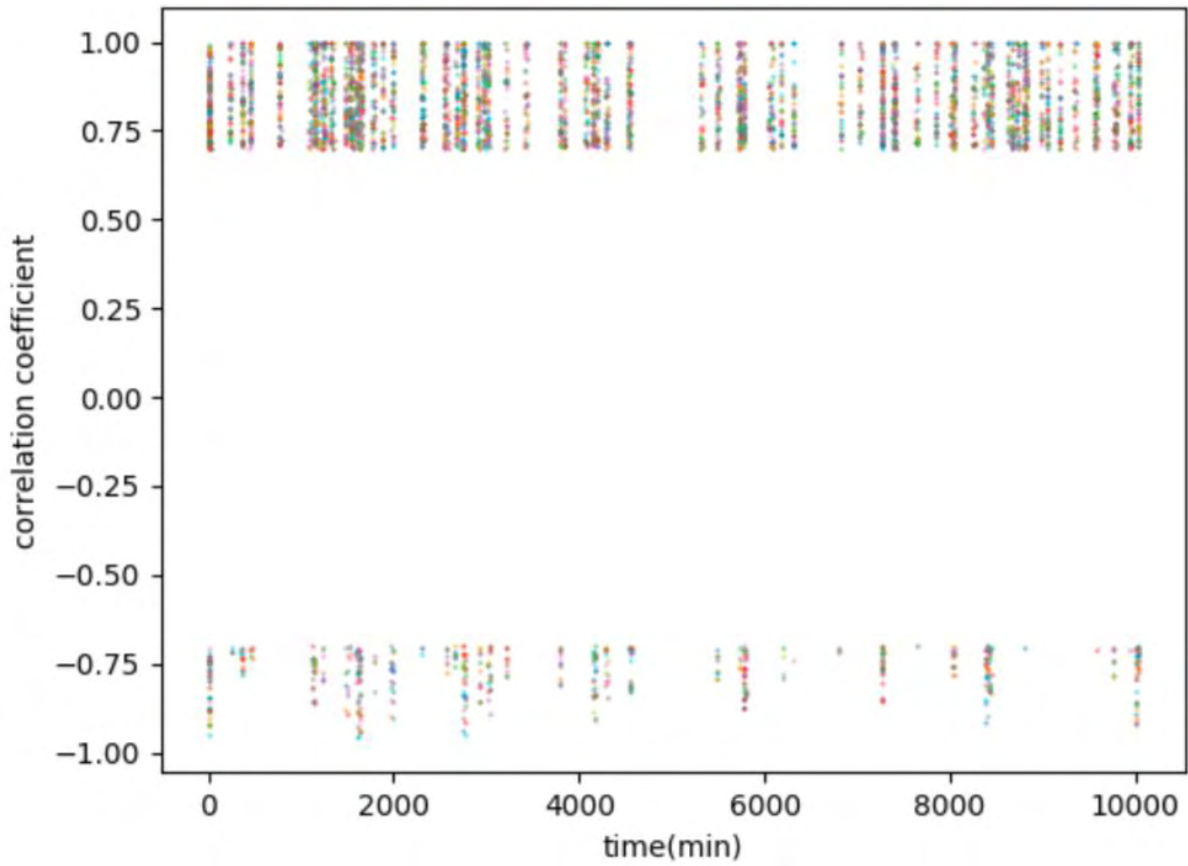
<표 1> 상관계수의 의미

상관 계수의 범위	의미
-1.0 < r < -0.7	매우 강한 음(-)의 상관 관계
-0.7 < r < -0.3	강한 음(-)의 상관 관계
-0.3 < r < -0.1	약한 음(-)의 상관 관계
-0.1 < r < 0.1	상관 관계 없음
0.1 < r < 0.3	약한 양(+)의 상관 관계
0.3 < r < 0.7	강한 양(+)의 상관 관계
0.7 < r < 1.0	매우 강한 양(+)의 상관 관계

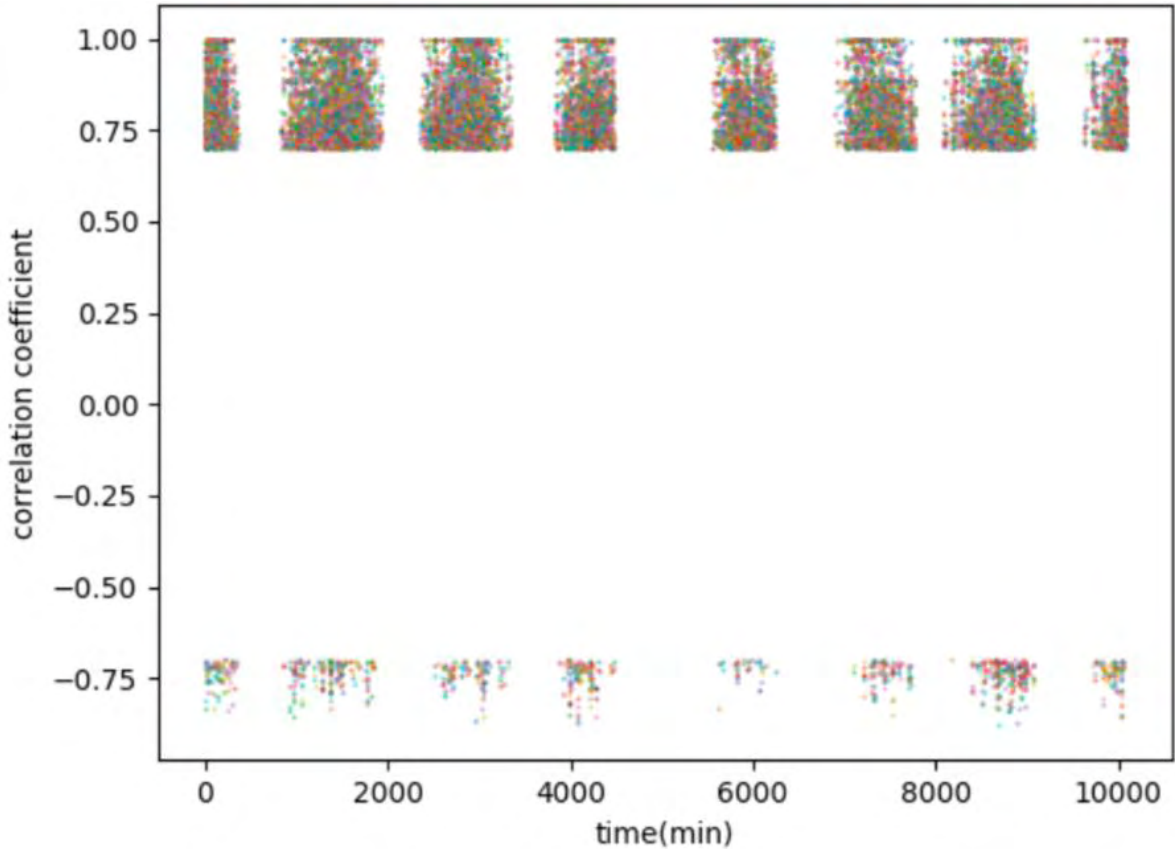
표 1 에서 매우 강한 음(-)의 상관관계의 의미를 가지는 상관계수의 값이 -0.7 보다 작거나 같고 - 1.0 보다 크거나 같은 경우와 매우 강한 양(+)의 상관관계의 의미를 가지는 상관계수의 값이 0.7 보다 크고 1.0 보다 작거나 같은 경우에 대하여 그래프를 그려 확인을 하였다.

그림 3 와 그림 4 은 CDY 가 데이터를 수집한 후 10 주차의 시간에 따른 상관관계를 나타낸 그래프이다. 그림 3 는 Cluster 1 에 대해, 그림 8 은 Cluster 2 에 대하여 나타내었는데 그래프를 해석해보면 Cluster 1 의 재방문 시간의 주기는 확실히 정해져 있지 않고 수시로 방문했다고 볼 수 있다.

그림 4 의 경우에는 총 8 개의 부분으로 구분되어



(그림 3) CDY 의 데이터 수집 10 주차의 Cluster 1 에 대한 상관관계 그래프



(그림 4) CDY 의 데이터 수집 10 주차의 Cluster 1 에 대한 상관관계 그래프

있는데 그림 3 의 해석과 비슷하게 해석할 수 있다. 재방문 주기가 약 하루단위, 즉, 1440 분 단위로 재방문하였다고 볼 수 있다.

대부분의 결과들을 보면 상관관계가 0.7 보다 크고 1.0 보다 작거나 같은 부분과 상관관계가 - 0.7 보다 작거나 같고 - 1.0 보다 크거나 같은 부분의 비어 있는 부분이 비슷한 것을 확인할 수 있다. 이를 통해 내릴 수 있는 결론은 음의 상관관계와 양의 상관관계 자체가 비슷하게 나타난다고 판단할 수 있다. 하지만 각각의 그림들에서 상관관계가 0.7 보다 크고 1.0 보다 작거나 같은 부분에는 딱 차 있지만 - 0.7 보다 작거나 같고 - 1.0 보다 크거나 같은 부분에서는 - 1.0 에 가까운 부분이 적게 나타나는 것을 확인할 수 있다. 이를 통해 내릴 수 있는 결론은 음(-)의 상관관계보다는 양(+의 상관관계가 더 자주 나타난다고 판별 할 수 있고, 양의 상관관계는 존재하지만 음(-)의 상관 관계가 나타나지 않는 데이터들의 시간이 더 의미 있는 시간이라고 판별할 수 있다.

<표 2> SQuARE 와 ISO/IEC 9126, ISO/IEC 14598 사이의 관계

순위	Cluster 1	Tau (min)	Cluster 2	Tau (min)
1	10332	1	2196	1
2	3724	2	961	1719
3	2060	3	830	2
4	1077	4	635	3301
5	943	9	634	1328
6	918	11	573	3
7	903	10	477	25
8	879	5	477	15
9	874	8	473	40
10	832	7	466	4
11	824	6	463	47
12	771	12	461	65
13	747	14	460	43
14	736	13	444	63
15	698	15	438	86
16	601	16	434	337
17	578	17	434	45
18	566	20	430	49
19	538	19	428	39
20	530	18	424	58
21	518	10070	423	117
22	499	21	423	24
23	488	10078	415	70
24	486	10077	411	37
25	486	10069	410	67

표 2 는 강한 상관계수로 분류된 상위 25 개의 결과를 나타낸 것이다. Cluster 1 과 Cluster 2 를 서로 비교하여 상위 25 개를 표로 표현하였는데, Cluster 1 의 경우 시간의 차이 τ 가 1 분~21 분이 상위 25 개 안에 포함되었고 10069 분, 10070 분, 10077 분, 10078 분이 포함된 것을 볼 수 있다. 상위 25 개만 봤을 때 방문 후 머물러 있는 확률이 높다고 볼 수 있다. Cluster 2 의 경우 시간의 차이 τ 가 Cluster 1 에 비하여 좀 더 다양하게 나타나는 것을 볼 수 있다. 예를 들면, Cluster 2 는 방문 후 1719 분 뒤에 다시 방문 한 경

우가 많고, 3301 분 뒤에 재방문 한 경우가 많다고 해석할 수 있다. 따라서 Cluster 2 의 경우 Cluster 1 보다 더 의미 있는 재방문 시간을 가진다고 볼 수 있다.

5. 결론

본 논문에서는 이동 데이터를 시간 분석하여 이동 양태가 어떤 형태로 나타나는지에 대하여 파악하는 방법을 제시하였다. 수집한 데이터를 시간에 따라 time slicing 한 뒤 각각의 시간에 따라 autocorrelation 과 autocovariance 를 계산하였다. 계산 결과에 따른 기준을 상관계수가 1.0 이하 0.7 초과일 경우에 대하여 강한 양(+의 상관계수로, 상관계수가 - 0.7 이하 - 1.0 이상일 경우에 대하여 강한 음(-)의 상관계수로 구별하였고, 이를 이용하여 시간에 따른 상관계수의 그래프를 그려 그래프의 의미를 파악하였다. 그래프에서 강한 양의 상관계수가 나오는 구간에 강한 음의 상관계수가 나오는 시간은 의미 있는 데이터라고 해석하였고, 표를 통해 군집별로 재방문 빈도가 높은 시간을 찾아내었다.

이러한 순환 패턴 분석은 다양한 위치기반 서비스 (LBS)를 설계하는데 유용하게 쓰일 것이다. 향후 연구론 사람들의 이동 데이터 및 지리적 위치 데이터를 IoT 와 같은 기기로 수집후 동일한 분석을 하는 것이다. 개인 이동성 패턴이 비슷한 주기적 패턴을 나타내지만 사람마다 다를 수 있다.

Acknowledgement

이 연구는 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행되었다. (NRF-2017R1D1A1 B03029788)

참고문헌

[1] H. Kim, H. Y. Song. Daily life mobility of a student: From position data to human mobility model through expectation maximization clustering. Multimedia, Computer Graphics and Broadcasting. 2011

[2] 김동엽, 송하윤. 집단 이동성 모델 도출 기법. 한국정보과학회 학술발표 논문집, 2015

[3] 이은별, 송하윤, 사람 성격과 자주 방문하는 장소의 상관관계 분석. 한국 정보과학회 학술발표논문집, 2015,

[4] H. Y. Song, D.Y. Choi. Defining Measures for Location Visiting Preference. Procedia Computer Science, 2015

[5] H. O. Hartley, Maximum likelihood estimation from incomplete data, Biometrics, 1958

[6] H.Y. Song, Probabilistic Space-Time Analysis of Human Mobility Patterns. In: WSEAS TRANSACTIONS on COMPUTERS, 2016

[7] Larue, Michael; Pheasant, David. Active sports tracker and method. U.S. Patent Application No 11/609,634, 2006.

[8] Garmin gpsmap 62s. <https://buy.garmin.com/en-US/US/on-the-trail/discontinued/gpsmap-62s/prod63801.html>; 2015.