

개별 이동성 모델부터의 집단 이동성 모델 도출

김동엽*, 최동연**, 송하운
홍익대학교 컴퓨터공학과

e-mail : *hatddum@gmail.com, **dy3459@naver.com, hayoon@hongik.ac.kr

Modeling Group Mobility from Individual Mobility Model

Dong-Yup Kim*, Dong-Yun Choi**, Ha Yoon Song

*Department of Computer Engineering, Hongik University

요 약

본 논문에서는 사람의 이동정보인 위치데이터를 바탕으로 위치분석(Location Analysis)을 통해 나타난 개개인의 이동성 모델을 바탕으로 각 개인의 이동성 모델에서 나타난 군집들의 관계를 분석해 개인이 속한 집단의 이동성 모델을 만든다. 집단 이동성 모델은 개인의 모델에서 나타난 군집을 이용하는데, 각 군집을 만드는데 필요한 위치 정보들과 군집의 중심, 군집간의 거리의 값을 계산하여 새로운 통합 군집을 만든다. 새로 만드는 군집은 각 특징에 따라 Micro Cluster, Macro Cluster의 2가지로 분류하였다.

실제 수년간 수집한 2명의 통합 개인 이동성 모델을 바탕으로 집단 이동성 모델을 생성한다. 집단 이동성 모델 생성에는 R Language를 사용하였고 결과 모델을 지도상에 표시할 수 있다.

1. 서론

다양한 산업과 학문분야들이 발전을 하면서 실제적인 인간 이동모습에 관한 모델의 필요성이 점차 대두되어왔다. 정확한 인간의 이동패턴에 관한 모델은 다른 학문 분야에서의 정확한 결과를 유도할 수 있게 할 뿐만 아니라 다양한 산업 환경 혹은 공익에서도 또한 유용하게 쓰일 수 있다. 기술의 발달로 인해 휴대용 위치정보 수집 장비나 혹은 스마트폰의 사용이 대중화되면서 사람의 이동정보인 위치정보들을 수집하는 일이 점점 용이해졌고, 이렇게 수집한 인간의 이동정보를 통해 우리는 인간의 이동패턴을 모델화 할 수 있다.

본 논문의 목적은 이러한 개개인의 이동 모델을 가지고 그룹의 이동 모델을 도출하여 사용자가 분석하려는 집단의 이동 패턴을 파악하는 것이다.

2. 데이터 수집 및 분석

그룹 이동성 모델을 만들기 위해서는 먼저 그룹에 속하는 각 개인의 이동성 모델이 필요하다. 각 개인이 수집한 위치 정보 data를 EM Algorithm (Expectation-Maximization Algorithm, 약자 EM)을 사용하여 일일 이동성 모델에서 Cluster 생성하였다. EM Algorithm에 대하여 대략적으로 설명하면 확률 모델에 관측 불가능한 변수들이 포함되어 있는 경우 최대우도함수(maximum likelihood method)를 사용하여 이에 적합한 변수를 찾는 방법이다. 이 알고리즘은 기존의 확률을 기반으로 하여 더 좋은 확률을 찾는 계산을 반복하는 구조로 이루어져 있다.

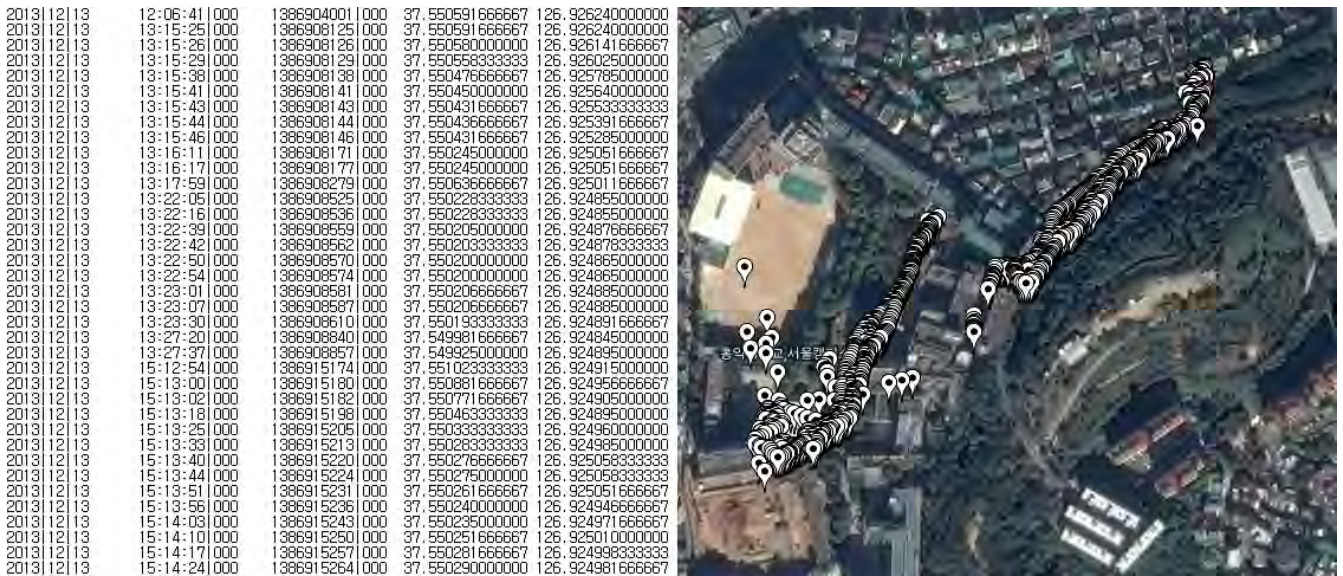
기존 연구로 진행되었던 알고리즘을 이용해서 개인의 이동성 모델을 만든다.[1] 개인의 이동성 모델이 생성하는 과정을 간단하게 설명하면 우선 개인의 위치 데이터를 수집한다. 위치 데이터는 스마트폰의 'Sport Tracker' Application과 위치 정보 수집 장치인 Garmin을 이용하였다. 수집된 데이터에는 기록된 시간과 위치의 정보(위도, 경도 등)가 저장되는데 연구에 필요한 정보인 위도, 경도, 시간 정보만을 데이터에서 추출한다. 추출된 데이터를 이용해서 개인의 일일 이동성 모델을 EM 알고리즘을 이용해서 생성하고[2], 일일 이동성 모델들을 이용하여 통합 개인 이동성 모델이 생성된다.

<그림 1>의 좌측은 수집한 데이터를 한 번 가공한 것이고 우측은 가공된 데이터를 지도 위에 표시한 모습이다. 가공된 데이터를 분석하여 일일 개인 이동성 모델을 만들 수 있다. <그림 2>는 수년간 수집한 위치 데이터를 이용하여 만든 일일 이동성 모델을 종합한 통합 개인 이동성 모델을 만든 결과이다.

3. 집단 군집화 방법

집단 이동성 모델을 생성하기 위해서 Cluster간의 거리를 구해야 하기 때문에 지표면 위의 두 점 사이의 거리를 구하는 방법이 필요하다. 지구는 평면이 아닌 구 형태를 가지고 있기 때문에 우리는 Haversine 공식을 이용한다. Haversine 공식은 <함수1>과 같다.

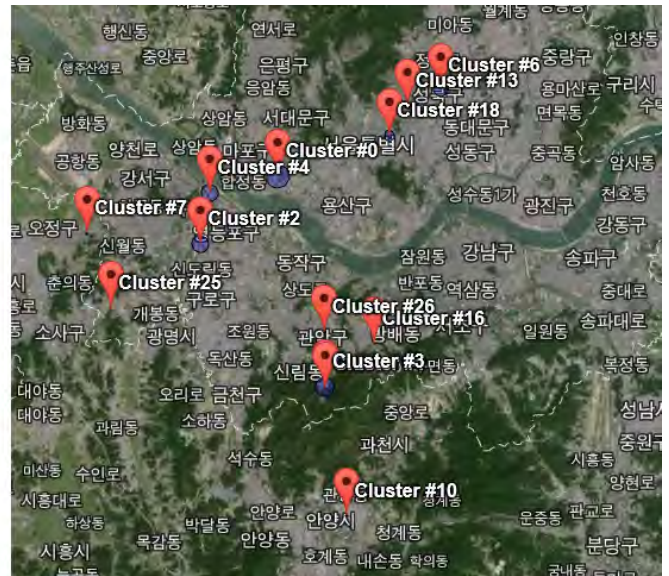
Haversine 공식은 구 위에 위치하는 두 지점 사이의 최단거리를 구하는 공식이다.[3] 이 공식을 이용해서 지구라



<그림 1> 가공된 초기 데이터(좌)와 지도 위에 표현한 데이터(우)

```

Number of Data : 0 Collected Time(hour) : 329.187222
-----
Cluster# 0 Accuracy Value : 1.000000
Center : 37.551797313501 126.925956648099
Stddev : 0.000000000000 0.000000000000
Max Distance : 0.632087
Mean Distance : 0.243226
TimeRatio : 0.408175
Count : 42826
Stay time(hour) : 134.366111
-----
Cluster# 1 Accuracy Value : 0.939824
Center : 37.454363739648 126.953435048668
Stddev : 0.000000000000 0.000000000000
Max Distance : 0.076024
Mean Distance : 0.015535
TimeRatio : 0.135208
Count : 16301
Stay time(hour) : 44.508611
-----
Cluster# 2 Accuracy Value : 0.972124
Center : 37.520625549470 126.881046248044
Stddev : 0.000000000000 0.000000000000
Max Distance : 0.599544
Mean Distance : 0.032641
TimeRatio : 0.019130
Count : 3834
Stay time(hour) : 6.297222
-----
Cluster# 3 Accuracy Value : 0.556403
Center : 37.455159337932 126.952996280526
Stddev : 0.000000000000 0.000000000000
    
```



<그림 2> 수집한 데이터의 통합 이동성 모델(좌)과 지도위에 표현한 데이터(우)

는 구체 위에서 두 장소 사이의 거리를 측정할 수 있다.

$$\begin{aligned}
 latitude &= (latitude2 - latitude1) \times \pi / 180 \\
 longitude &= (longitude2 - longitude1) \times \pi / 180 \\
 latitude1 &= latitude1 \times \pi / 180 \\
 latitude2 &= latitude2 \times \pi / 180 \\
 a &= \sin(latitude/2)^2 + \\
 &\cos(latitude1) \times \cos(latitude2) \times \sin(longitude/2)^2 \\
 c &= 2 \times a \times \tan(\sqrt{a}, \sqrt{1-a}) \\
 d &= R \times c
 \end{aligned}$$

<함수 1>

<함수 1>의 R 은 지구 반지름 값이고, π는 원주율 값이다. 다만 위 공식에서 지구가 완전한 구가 아닌 타원형으로 이루어진 구이기 때문에 반지름 값은 위치에 따라 변한다. 이로 인해 오차율이 생길 수 있다. 본 논문에서는

위치에 따라 심하게 변하는 경도의 값에 변화를 주었다. 위도의 경우 1°당 거리가 심하게 변하지 않지만 경도의 경우엔 북극 또는 남극에 가까워 질 경우 거리가 굉장히 짧아지기 때문이다.

지표면 위의 두 점사이의 거리를 구할 수 있게 되면, 두 사람의 개인의 이동성 모델에 있는 Cluster를 모두 각각 비교하여 Cluster간의 중심 사이의 거리를 구할 수 있다. 중심 사이의 거리를 비교하여 Cluster의 거리가 서로 닿을 정도로 가까우면 두 Cluster를 이용하여 새로운 Micro Cluster와 Macro Cluster를 생성한다.

Micro Cluster는 두 Cluster에 공통적으로 속하는 부분에 포함되는 위치 정보를 모두 수집하여 군집시킨다. 생성방법은 2개의 Cluster의 중심과 반지름의 값을 가져온 뒤, 각 Cluster내부에 포함된 점을 분류한다. 분류한 값이 서로 다른 클러스터 내부에 포함되는지 판단한 뒤, 내부에

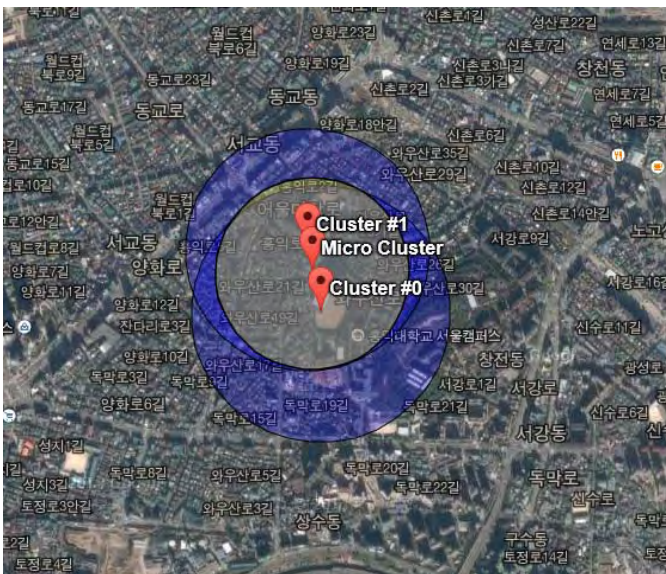
포함된 점만 가지고 새로운 리스트를 생성한다. 이렇게 생성된 리스트를 가지고 리스트에 속해있는 점들의 위도의 평균과 경도의 평균으로 Micro Cluster의 중심을 구한다. Micro Cluster의 반지름은 중심으로부터 리스트의 각 점까지의 거리를 Haversine공식을 통하여 구한 뒤, 그 점 사이의 평균과 최댓값을 더하여 Micro Cluster의 반지름의 길이로 정한다. 원을 생성할 때에는 각도를 1°씩 늘려주며 총 360개의 새로운 점을 생성하여 연결하였다. <그림 3>의 왼쪽 그림이 서로 다른 두 명의 Micro Cluster 다.

Macro Cluster는 두 Cluster중 어느 것이라도 포함되는 위치정보를 모두 수집하여 군집시킨다. 생성방법은 2개의 Cluster의 중심과 반지름의 값을 가져온 뒤, 각각의 Cluster내부에 포함된 점을 분류한다. 중심의 위치와 반지름의 길이, 원을 생성하는 방법은 Micro Cluster와 동일하다. <그림 3>의 오른쪽 그림이 서로 다른 두 명의 Macro Cluster다.

<함수 2>는 새로운 Cluster의 반경을 구하는 공식이다. *clat*, *clon*은 새로운 Cluster 중심의 위도와 경도를 나타낸다. *edgelat*과 *edgelon*는 반경의 각 1도에 해당하는 점의 위도, 경도이고 *i*는 1에서 360의 각도 값을 갖고 *i*에 0.01745329251 값을 곱해서 각 각도에 해당하는 라디안 값으로 변환한다.

```
lonangle <- 6378 * cos(clat * 0.01745329251)
            * 2 * pi/360
latangle <- 40000/360
edgelat <- clat * cos(i * 0.01745329251)
            * maxdis * (1/latangle)
edgelon <- clon * sin(i * 0.01745329251)
            * maxdis * (1/lonangle)
```

<함수 2>



4. 집단 이동성 모델 생성

집단 이동성 모델에 사용될 개인 이동성 모델은 2013년 5월 20일부터 2014년 4월 30일까지 실험참가자의 위치 data를 이용한다, 수집된 위치 정보 data의 개수는 총 6445737개이다.

분석에 사용할 언어는 Big-data Analysis에 적합한 R language를 이용했다. 프로그램은 연구실의 서버 컴퓨터를 사용해서 작업시간을 최소화 했다. 결과는 kml 형식으로 출력하여 지도 위에 나타내는데 새로 생성되는 Micro Cluster는 노란색, Macro Cluster는 빨간색으로 영역을 표시해 구분한다. 각 Cluster의 중심은 빨간 표시로 나타내고 우측에 Cluster의 이름을 표시한다.

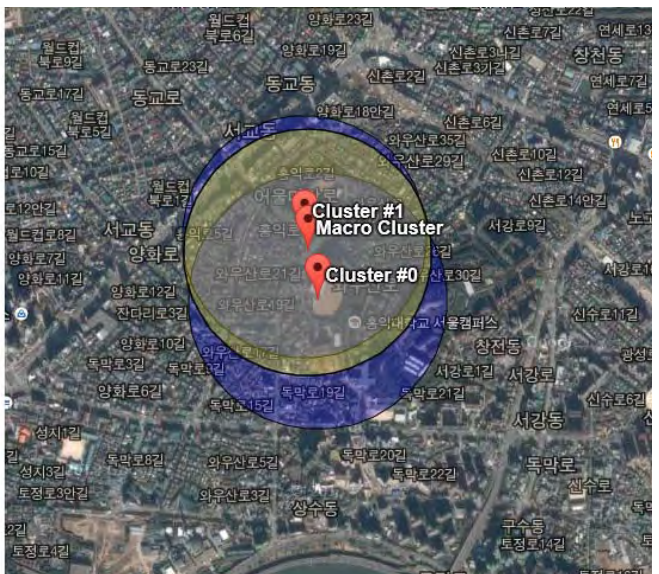
분석에 사용할 통합 모델들은 <그림 4>와 같다. <그림 4>를 살펴봤을 때 집단 이동성 모델에서 새로운 Cluster 쌍이 5개가 생성 될 것이라는 것을 예측하고 실제로 분석을 한 결과 서울 홍익대학교, 덕평 자연휴게소, 제주도의 3곳, 총 5쌍의 Cluster가 생성되었다.

<그림 5>는 결과 모델(좌)과 새로 생성된 Micro Cluster와 Macro Cluster 중 홍익대학교 부근에 생성된 Micro Cluster와 Macro Cluster(우)이다. Macro Cluster와 Micro Cluster는 예상했던 크기와 비슷한 크기로 출력되었다. 실제로 출력 된 결과의 데이터에서는 분석에 사용된 Cluster가 표현되지 않지만, <그림 5>에서는 편의를 위해 표현해 주었다. 파란색과 초록색이 분석에 사용된 Cluster이다.

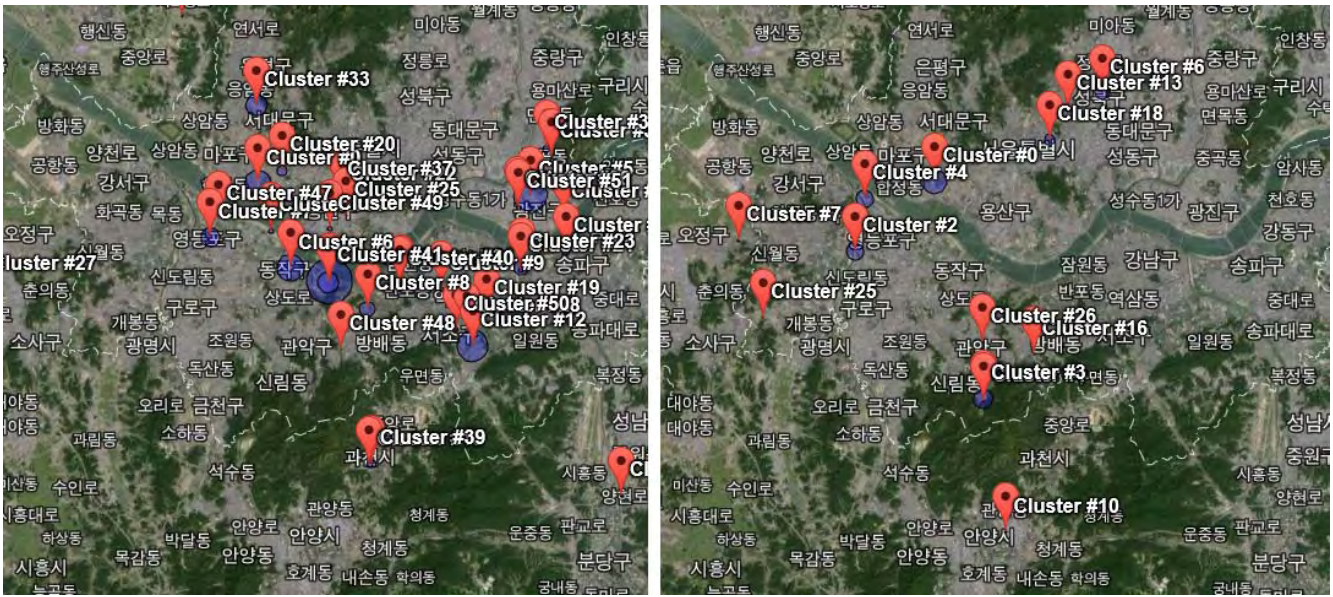
5. 결론 및 향후 연구 진행방향

본 논문에서는 각 개인이 수집한 위치 정보를 분석한 통합 개인 이동성 모델을 이용해서 집단 이동성 모델을 만들 수 있었다.

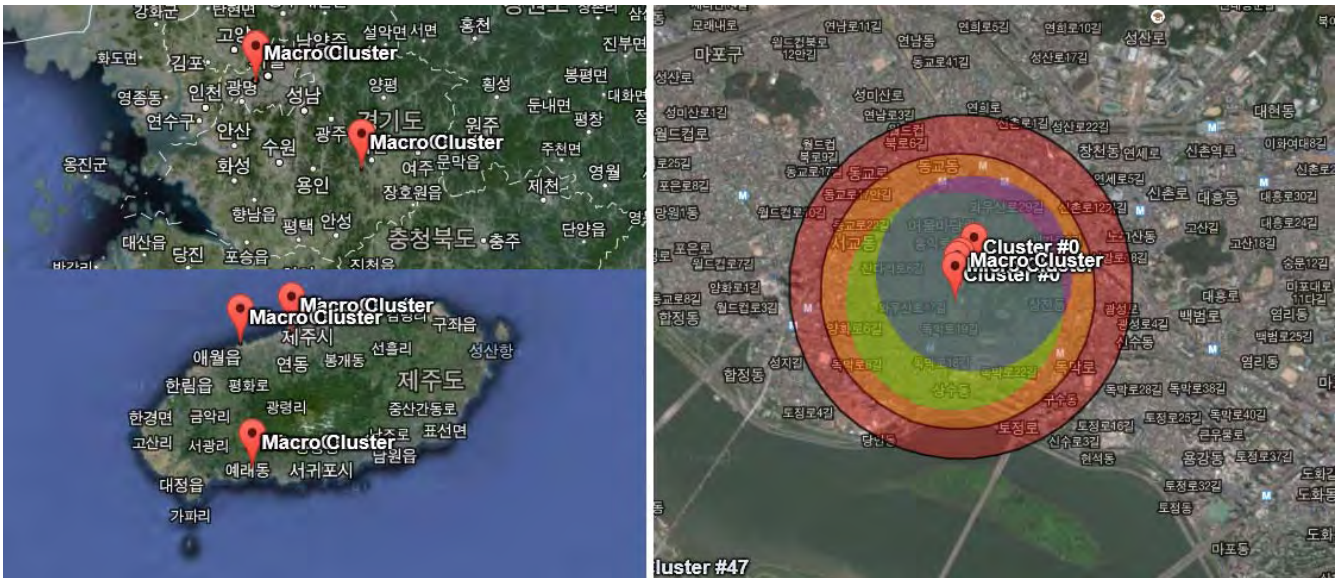
새로 생성되는 Micro Cluster와 Macro Cluster의 중심 위치는 분석하기 전 예측했던 위치에 정확히 나타났으나 각 Cluster의 반경의 넓이가 예측보다 넓게 생성되었다.



<그림 3> Micro Cluster(좌)와 Macro Cluster(우)



<그림 4> 2명의 통합 개인 이동성 모델



<그림 5> 통합 이동성 모델(좌)과 모델의 홍익대학교 부근(우)

본 연구에서는 데이터 필터링 없이 모든 위치 데이터를 사용했기 때문에 Cluster 주변에 이동 중인 위치 데이터까지 분석에 활용되어 Cluster가 넓게 생성된 것이다.

앞으로 본 연구가 나아가야 할 방향은 첫째, 알고리즘 최적화를 통해 분석 시간을 단축시켜야 할 것이며, 둘째, 위치 데이터를 입력받을 때 각 Cluster를 생성할 때 실제로 사용된 위치 데이터들만을 사용해야 할 것이며, 셋째, 현재 연구는 2명의 모델을 이용하여 집단 이동성 모델을 생성했으나 2명 이상의 개인 모델에 적용시켜야 할 것이다.

6. Acknowledgement

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임.

(NRF-2012R1A2A2A03046473)

참고문헌

- [1] H. Kim, H.Y. Song, Daily life mobility of a student: From position data to human mobility model through expectation maximization clustering, in: Multimedia, Computer Graphics and Broadcasting, Vol. 263 of Communications in Computer and Information Science, 2011, pp. 88-97.
- [2] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via EM algorithm, Journal of the Royal Statistical Society, Series B (Methodological), vol. 39, no. 1, pp. 1-38, 1977
- [3] C. C. Robusto, "The Cosine-Haversine Formula", The American Mathematical Monthly, Vol. 64, No. 1, Jan., 1957.