

A Method for Group Mobility Model Construction and Model Representation from Positioning Data Set Using GPGPU

Song Ha Yoon[†] · Kim Dong Yup^{††}

ABSTRACT

The current advancement of mobile devices enables users to collect a sequence of user positions by use of the positioning technology and thus the related research regarding positioning or location information are quite arising. An individual mobility model based on positioning data and time data are already established while group mobility model is not done yet. In this research, group mobility model, an extension of individual mobility model, and the process of establishment of group mobility model will be studied. Based on the previous research of group mobility model from two individual mobility model, a group mobility model with more than two individual model has been established and the transition pattern of the model is represented by Markov chain. In consideration of real application, the computing time to establish group mobility mode from huge positioning data has been drastically improved by use of GPGPU comparing to the use of traditional multicore systems.

Keywords : Group Mobility Model, Clustering, Parallel Computing, Markov Model, R, GPGPU

GPGPU에 기반하는 위치 정보 집합에서 집단 이동성 모델의 도출 기법과 그 표현 기법

송 하 윤[†] · 김 동 업^{††}

요 약

인간의 위치 이동 데이터를 모바일 기기에서 수집한 위치 정보를 이용해 얻을 수 있게 되면서, 위치 정보를 어떻게 이용할 수 있는지 그 활용 방안이 중요시 되고 있다. 이 연구에 앞서 위치 정보에 포함된 위치 정보와 시간 정보를 이용한 개인 이동성 모델 도출 연구가 선행되었다. 이동성 모델의 개념을 집단으로 확장하여 특정 집단에 속한 사람들의 개인 이동성 모델을 이용한 집단 이동성 모델을 도출하는 방법에 대해서 연구했고, 두 명의 개인 이동성 모델을 이용한 집단 이동성 모델과 그 모델을 표현하는 Markov 모델을 생성할 수 있었다. 본 논문에서는 세 명 이상의 개별 이동 모델을 포함하는 사람의 이동성 모델을 생성하고 집단 모델 내 군집간의 확률 기반 Markov 모델을 도출하는 방법에 대해 소개한다. 또한 GPGPU 기법을 통해 생성 시간을 줄이는 기법을 이용하여 실용화를 고려하였다.

키워드 : 집단 이동성 모델, 군집화, 병렬 컴퓨팅, Markov Model, R, GPGPU

1. 서 론

개인이 갖고 있는 모바일 기기에서 위치 정보를 수집할 수 있게 되면서, 개인의 위치 정보를 이용한 연구가 대두되고 있다. SNS(Social Network Service), 모바일 어플리케이션에

서는 다양한 위치 기반 서비스를 제공하고 있고, 자동차 네비게이션, 전자 발찌 등 그 활용 범위가 점점 확대되어 가고 있다. 위치 정보를 이용하여 사람 또는 사물의 이동 패턴을 분석할 수 있으며 그 이동 패턴의 활용 가능성은 위치 정보 보다 훨씬 더 크고 다양한 분야에서 이용가능하다. 이 논문의 선행 연구는 개인이 수집한 위치 정보를 이용해서 개인 이동성 모델을 도출했었다[1]. 연구의 범위를 개인에서 집단으로 확장하여 집단의 이동성 모델을 만들 것이다. 특정 집단 모델을 만들 수 있게 되면 연령별, 직종별, 성별 등 사용자가 원하는 사람들의 공통된 이동 패턴을 구할 수 있다. 기존 연구에서는 2 명의 집단 이동성 모델을 생성할 수 있었다.

* 이 연구는 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행됨(NRF-2016K1A3A7A08952097).

† 중신회원 : 홍익대학교 컴퓨터공학과 교수

†† 비 회 원 : 홍익대학교 컴퓨터공학과 석사과정

Manuscript Received : June 23, 2016

First Revision : September 20, 2016

Accepted : September 20, 2016

* Corresponding Author : Song Ha Yoon(librarymusik@gmail.com)

하지만 3 명 이상의 집단 이동성 모델을 생성하기 위해서는 수많은 위치 정보를 효율적으로 분석하는 방법이 필요하다. 그래서 GPU 장치를 활용한 병렬 처리 연산을 사용한다. 또한 생성한 집단 이동성 모델의 Markov 모델을 이용해서 모델 내 군집간의 이동 확률을 구할 수 있다. 2장에서는 관련된 연구를 소개한다. 3장에서는 수집한 위치 데이터에 대해서 다루고, 4장에서는 집단 이동성 모델을 도출하는 방법을 다룬다. 5장에서는 Markov 모델을 도출하고 6장에서는 실험 결과를 보인다. 7장에서 결론을 다루고 마친다.

2. 관련 연구

인간이나 사물의 이동 패턴에 대한 분석에 관한 다양한 연구들은 여러 분야에서 다양하게 수행되어져 왔다. 심리학적 측면에서는 GPS 수집 장치를 이용하여 좀 더 정밀하게 사람 개개인의 움직임과 심리적 특성을 연구해야 함을 보이는 연구가 있었다[2]. 통계·물리학 측면에서는 휴대폰의 기지국 정보를 바탕으로 인간의 이동이 중심부로부터 Power law 법칙을 따름을 보였고, 이런 방식으로 인간의 이동을 93%까지 예측이 가능하다고 말하고 있다[3]. 다른 연구로는 GPS의 이동 궤적을 바탕으로 계층적인(Hierarchical) CRFs (Conditional Random Fields)를 사용하여 사람이 있는 장소(place)와 활동성(activities)을 추출한 연구도 있었다[4]. 또 기존의 사용자의 현재 위치에 기반한 위치기반(location-based) 서비스 방식이 아니라 사용자의 다음 위치의 예측을 통한 서비스를 제공하기 위해 Foursquare 데이터를 이용하여 사람의 다음 위치를 추정하는 방법에 대한 연구도 있다[5]. 그러나 지금까지의 위치 분석 방법에 관한 연구들은 주로 분석 대상이 단일 대상에 한정된 것인 반면에, 본 연구는 하나 이상의 이동 패턴들을 통합하여 이동 패턴들이 포함된 집단의 이동 패턴을 도출해내는 데 목적을 두고 있다. 집단의 이동 모델을 도출해 낼 수 있으면 분석 시 사용자가 분석에 사용할 대상만을 지정해서 집단으로 설정하고 그 집단의 이동 패턴을 알아낼 수 있다. 이는 분석에 연령, 성별, 나이, 직업 등 인간이 갖는 특정한 성질에 따른 이동 패턴을 도출해 낼 수 있다는 장점을 갖는다. 그러나 기존의 연구 중 이러한 주제를 다루고 있는 연구는 존재하지 않는다. 따라서 다수의 인원이 수집한 위치 정보를 통해 나타난 개인 이동성 모델을 가지고 여러 명의 집단 이동성 모델을 도출해내는 방법에 대해서 논의하고자 한다.

이 연구를 진행하기 앞서 진행했던 사전 연구는 위치 정보로부터 각 개인의 이동성 모델을 도출해 내는 연구를 진행했다. 개인 이동성 모델은 기대치 최대화 알고리즘을 이용해서 구한다[6]. 개인 이동성 모델의 결과로 수집한 기간 동안 사용자가 자주 방문했던 위치들의 Cluster들의 정보와 Cluster간의 이동 확률이 출력된다. 이 외에도 연구실에서는 위치 정보를 활용하여 장소의 주목할만한 지점을 찾는 연구

[7], 위치의 속도를 이용해서 에러를 검출하는 연구[8], 파티클 필터를 이용한 이동 상태 분류[9], Big Five Personality를 이용한 위치 예측[10], 성격과 선호 장소간의 상관 관계 분석[11] 등 다방면으로 연구를 진행 중이다.

3. 위치 데이터 수집

집단 이동성 모델의 분석을 위해서는 많은 양의 위치 정보가 필요하다. 실험에 사용할 위치 정보를 수집하기 위해 연구 지원자를 모집하여 개인이 이동하는 위치 정보를 제공받는다. 지원자는 스마트폰 어플리케이션인 'Sports-Tracker'와 전용 GPS 리시버인 'Garmin' 시리즈를 사용하여 위치 정보를 수집한다. 수집 장치는 GPS 신호를 이용해서 초당 한 개의 위치 정보가 기록된다. 위치 수집은 24 시간동안 지원자가 이동한 경로를 모두 기록하는 것을 원칙으로 하며 개인 정보 보호를 위해 수집한 정보를 닉네임을 이용하여 그룹화한다.

수집된 데이터에서 분석에 사용되는 위도, 경도, 시간 정보만을 추출해서 새로운 raw data set를 만들어 사용한다. 데이터 수집은 지속적으로 이루어지고 있으며 실제 연구에 사용하는 데이터는 본인을 포함한 9 명의 지원자가 13년 5월부터 15년 7월까지 최대 2년에서 최소 1년 동안 수집한 위치 데이터 총 23,712,540 개를 이용해서 집단 이동성 모델을 생성했으며 분석 시간 비교에는 모든 참가자 23 명의 위치 데이터 총 53,643,505 개를 사용한다. 각 수집 장치에서 모은 위치 정보들은 웹이나 또는 기기에서 다운로드받을 수 있다.

4. 집단 이동성 모델 도출 방법

집단 이동성 모델을 도출하기 위해서는 우선 집단에 속하는 개인의 이동성 모델과 수집 기간 동안의 위치 데이터가 필요하다. 개인 이동성 모델의 결과물을 가지고 여러 사람의 집단 모델을 만들기 전에 먼저 두 사람의 집단 모델을 생성하는 알고리즘을 구현했다[12]. 먼저 기존 연구인 두 명의 집단 이동성 모델을 생성하는 방법을 간단히 소개하고 그 다음 세 명 이상의 집단 이동성 모델을 생성하는 방법에 대해 설명한다.

4.1 두 명의 집단 이동성 모델

두 사람의 집단 모델의 구현 방법은 알고리즘 1과 같다. 입력 값은 두 명의 위치 정보와 개인 이동성 모델이다.

1, 2 줄과 같이 각 위치 정보에서 분석에 사용할 위도, 경도, 시간 정보와 개인 모델에서 군집의 중심 위도, 경도, 반지름 값들을 추출한다. 3-5 줄과 같이 두 사람의 개인 모델의 군집들을 서로 비교한다. 6-10 줄은 양쪽 군집 모두에

Algorithm 1. Modeling Group Mobility

```

Require: Position Dataset x 2
         Individual Mobility Model x 2
1 // filter latitude, longitude, radius in each clusters
2 Extract position data1 and position data2
3 for Clusters_of_Person1 do
4   for Clusters_of_Person2 do
5     if Distance(cluster1, 2) ≤ radius1 + 2 then
6       /* filter position data belonged to two clusters
7         together */
8       Union filtered position data1, position data2
9       micro cluster's center is Average(cluster1, 2)
10      micro cluster's radius is Max(union_distance)
11      /* filter position data belonged to cluster1 and
12        cluster2 */
13      Union filtered position data1, position data2
14      macro cluster's center is Average(cluster1, 2)
15      macro cluster's radius is Max(union_distance)
16    end if
17  end for
18 end for
19 return micro cluster, macro cluster
    
```

Algorithm 1. A Algorithm of Mobility Model of Group of Two

포함되는 위치 정보들을 추출하여 새로운 Micro Cluster를 생성한다. 11-15 줄은 어느 한쪽 군집이라도 포함되는 위치 정보를 추출하여 새로운 Macro Cluster를 생성한다. 새로운 군집의 중심은 두 가지 방법으로 구할 수 있는데, 한 가지는 포함된 위치 정보의 평균값을 사용하는 것이고 다른 하나는 합치는 두 군집의 중심 위치 값을 사용하는 것이다. 우리는 두 번째 방법을 사용한다. 이유는 첫 번째 방법을 사용했을 경우 예외적인 상황, 즉 두 군집에 공통적으로 포함되는 위치 정보가 편중됐을 때 그 주변에 중심이 생기고 반지름의 크기가 비효율적으로 커질 수 있기 때문이다. 결과는 새로 생성된 집단 군집의 중심 위치, 반지름, 포함된 위치 정보 개수 등이 기록된 텍스트 파일과 군집들을 지도 위에 시각화한 파일이 출력된다.

Fig. 1은 실제로 1년 동안 위치 데이터를 수집한 실험 참가자 두 명의 집단 이동성 모델을 도출한 시각화 파일이다.

Fig. 2는 Fig. 1의 두 명의 개인 이동성 모델에서 나타난 군집들 중 서로 겹치는 영역이 있는 군집들만 나타낸 것이다. 그림의 위쪽은 왼쪽부터 홍익대학교, 덕평 휴게소이고, 아래쪽은 제주도에서 나타난 곳으로 왼쪽부터 숙소 근처, 제주 공항, 박물관이다. 이를 통해서 집단 모델이 각 개인이 이동한 곳들 중에 공통적으로 방문한 장소에 군집을 생성한다는 것을 검증하였다.

4.2 세 명 이상의 집단 이동성 모델

세 명 이상의 집단 이동성 모델을 만들기 위해서는 무엇보다도 엄청난 양의 위치 정보를 처리할 수 있는 연산 처리 능력이 필요하다. 우리는 이런 문제점을 해결하기 위해서



Fig. 1. A Visualization of Mobility Model with a Group of Two



Fig. 2. Locations of Overlapped Clusters by Different Individual Mobility Models

알고리즘에 GPU를 활용하기로 결정했다[13].

GPU를 사용하기 위해서 Nvidia에서 제공하는 CUDA를 R Language에 접목시키는 방법을 사용한다. R Language에서 CUDA로 프로그래밍된 연산 함수를 R Language에서 호출해서 사용한다[14]. 우리의 알고리즘 중에 대부분의 처리 시간을 차지하는 부분은 위치 정보와 군집간의 거리를 계산하고 군집 내에 포함되어 있는지, 포함되는 경우 그 위치의 위도, 경도 정보를 추출하는 부분이다.

R Language에서 행렬로 선언된 위치들의 위도, 경도들을 CUDA 함수에서 배열로 입력받아 GPU로 위치 정보의 위도, 경도, 군집의 위도, 경도 정보를 전달한다. GPU 내 수많은 처리 유닛에서 전달 받은 정보를 이용해서 각각 거리를 계산하고 다시 host로 계산된 결과를 전달받는다.

3명 이상의 집단 이동성 모델을 도출하는 알고리즘은 알고리즘 1을 여러 번 반복하는 형태이다. 먼저 처음 두 명의 집단 모델을 도출하고 결과로 나온 집단 모델과 다음 인원의 개인 모델을 비교하여 다시 집단 모델을 생성하고 이 과정을 반복한다.

Algorithm 2. Modeling Group Mobility

Require: cluster[]: *Position Data Set* for N individuals and *Individual Mobility Model* for N individuals

```

1 // input how many individual (= count)
2 for 2 <= i <= count do
3   if i = 2 then
4     result is GroupAnalysis(clusters[1], clusters[2])
5   else
6     result is GroupAnalysis(clusters[i], result)
7   end if
8 end for
9 // delete overlapped clusters
10 return result

```

Algorithm 2. A Algorithm of Mobility Model of Group of More than Three

입력 값은 모든 인원의 위치 정보들과 개인 이동성 모델이다. 1 줄처럼 프로그램이 시작되면 분석에 사용될 인원수를 입력 받고, 분석에 포함될 인원을 선택한다. 3-8 줄처럼 처음 입력된 두 명을 이용해서 집단 군집을 먼저 생성하고, 집단 군집과 다음 인원의 개인 이동성 모델을 비교해서 새로운 집단 군집을 생성한다. 9 줄에서는 정확히 같은 위치에 군집이 생성되는 경우에 군집 내에 포함되는 위치 정보의 수가 적은 군집을 제거한다.

단, 분석 순서는 개인 모델에서 군집의 수가 많은 것을 우선으로 한다. 이유는 군집의 수가 적은 인원부터 분석이 시작될 경우 나중에 추가되는 인원의 군집 수가 아무리 많더라도 앞서 진행된 분석에 의해 대부분의 군집이 분석에 영향을 끼치지 못할 수 있기 때문이다.

다음 알고리즘은 집단 모델을 만들 때 사용되는 Group Analysis 알고리즘이다.

Algorithm 3. Group Analysis using GPU

Input : cluster1, cluster2,
position data1, position data2

```

1 for cluster1 do
2   for cluster2 do
3     if Distance(cluster1, 2) ≤ radius1 + 2 then
4       // set memory using GPU specific functions
5       SetGPU(blocks, threads)
6       GPUmemallocation(array's length)
7       GPUmemcpy(host to device)
8       GPUcompute(cluster1, cluster2, pos1, pos2)
9       GPUmemcpy(device to host)
10      GPUmemfree()
11      micro cluster's center is Average(cluster1, 2)
12      micro cluster's radius is Max(micro distance)
13      macro cluster's center is Average(cluster1, 2)
14      macro cluster's radius is Max(macro distance)
15    end if
16  end for
17 end for
18 return micro cluster, macro cluster

```

Algorithm 3. A Algorithm of Group Model Using GPU

위 알고리즘은 두 명의 집단 이동성 모델 도출 알고리즘과 비슷한 형태를 갖는다. 입력 값은 군집의 두 set과 각 위치 정보들이다. 1-3 줄처럼 각 set의 군집들의 거리와 반지름을 이용해서 서로 겹치는 부분이 있는지 확인한다. 만약 서로 겹친다면 4-13 줄처럼 Micro Cluster와 Macro Cluster에 포함될 위치 정보들을 추출하는데 여기서 GPGPU를 활용한다. 4-7 줄에서 메모리 할당 및 각 data를 GPU로 전송하고, 8 줄에서 GPU에서 동작하는 함수를 호출한다. 각각의 GPU device에서 거리를 계산하고 계산된 거리 값과 반지름을 비교해 양쪽 군집에 모두 포함되는 경우 Micro Cluster의 위치 정보로, 한쪽 군집에라도 포함되는 경우 Macro Cluster의 위치 정보를 결과로 반환한다. 9-10 줄에서 각각의 장치에서 계산된 거리 값과 위치 값을 다시 host 컴퓨터로 전달받고 이렇게 포함된 위치 정보들을 가지고 11-14 줄에서 새로운 Micro Cluster와 Macro Cluster를 생성한다. 다음 Fig. 3은 실험 참가자 9 명이 1년간 수집한 실제 위치 정보를 가지고 분석한 집단 모델의 시각화 결과이다.



Fig. 3. A Visualization of Mobility Model with a Group of Nine

위의 그림에서 나타난 위치는 홍익대학교 부근으로 실험 참가자 9 명이 대부분 학생이기 때문에 학교 근처에서 집단 군집이 생성된 것이다.

5. Markov 모델

집단 이동성 모델에서 나타난 군집간의 이동 패턴을 표현하기 위해서 Markov 모델로 나타낼 것이다[15].

$$P[X_{n+1} = j | X_n = i, \dots, X_0 = i_0] = P[X_{n+1} = j | X_n = i] \quad (1)$$

위 수식과 같이 Markov 모델은 현재의 상태가 오직 바로 이전의 상태에 의해서 결정되는 확률 모델이다. Markov 모델은 상태와 상태 전이 확률로 구성된다. 우리의 연구에서는 집단 이동성 모델에 나타나는 군집을 Markov 모델의

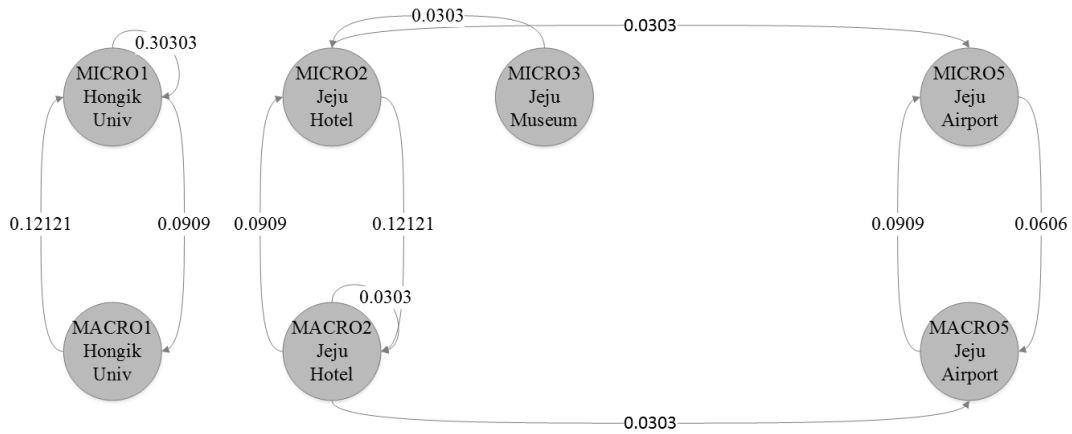


Fig. 4. Markov Model of Two from 8 Hour to 9 Hour

상태라고 보았을 때, 위치 정보의 이동 궤적을 이용해서 상태 전이 확률을 구한다.

5.1 집단 이동성 모델에서의 Markov 모델

위치 정보의 이동 궤적을 알기 위해서 위치 정보를 1 시간을 주기로 어떤 군집에 속하는지 파악한다. 각 위치 정보가 1 시간 단위로 어떤 군집에 얼마나 머물렀는지 또 다른 군집으로 이동할 경우 어디에서 어디로 이동했는지 각각 계산하여 시간별 전이 확률을 구할 수 있다. Markov 모델을 생성하기 위해서는 먼저 집단 이동성 모델을 만들고 각 군집에 포함된 위치 데이터를 시간별로 구분한다. 위치 정보의 이동 궤적을 구해서 시간대별 전이 확률을 얻는다. 마지막으로 데이터를 시각화하면 집단 이동성 모델에서의 시간대별 Markov 모델을 만들 수 있다.

Table 1은 두 사람의 집단 모델에서 위치 정보를 시간별로 구분하여 8시부터 9시 사이의 이동 궤적을 분류한 표이다. 표에서 Stay at 값은 수집 기간 동안 각 군집에 속하는

위치 정보가 계속 머물러 있었는지 계산한 횟수이고 Transit to 값은 다른 군집으로 이동한 경우, 그 군집과 횟수를 계산한 것이다. 총 이동 궤적의 수는 Stay at의 총합과 Transit to의 총합을 합친 160 개이다. 직업이 학생인 집단의 이동성 모델을 만들었기 때문에 주로 학교 근처에서 이동 궤적이 발생했다. 표에 나타난 값을 이용해서 두 사람의 Markov 모델을 만들 수 있다. 그림 4는 두 사람의 집단 모델로부터 만든 8시부터 9시 사이의 Markov 모델이다. 그림을 통해 알 수 있듯이 8시부터 9시 사이에 이 집단에 속한 두 사람은 공통적으로 가장 높은 확률로 홍익 대학교에서 머무를 것이고 그 다음으로 홍익 대학교의 Micro Cluster에서 Macro Cluster로 이동하거나 또는 제주 호텔의 Micro Cluster에서 Macro Cluster로 이동할 것임을 알 수 있다.

Fig. 5는 9 명의 집단 이동성 모델에 대한 Markov 모델이다. 집단 이동성 모델에서 나타난 군집이 서로 겹치기 때문에 군집 내에 포함된 위치 정보의 수가 가장 많은 첫 번째 군집을 이용해서 Markov 모델로 나타냈다. 이 집단에 속하는 9 명의 구성원들은 공통적으로 가장 높은 확률로 홍익대학교 Micro Cluster에 머물 것이고 가장 낮은 확률로 홍익대학교 Macro Cluster에서 Micro Cluster로 이동할 것이다.

Table 1. Location Sequence of Two from 8 Hour to 9 Hour

	Stay at	Transit to	
Micro Cluster1	10	Macro Cluster1	3
Micro Cluster2	0	Micro Cluster5	1
		Macro Cluster2	4
Micro Cluster3	0	Micro Cluster2	1
Micro Cluster4	0	0	
Micro Cluster5	0	Macro Cluster5	2
Macro Cluster1	0	Micro Cluster1	4
Macro Cluster2	1	Micro Cluster2	3
		Macro Cluster5	1
Macro Cluster3	0	0	
Macro Cluster4	0	0	
Macro Cluster5	0	Micro Cluster5	3
Total Sequences	10	22	

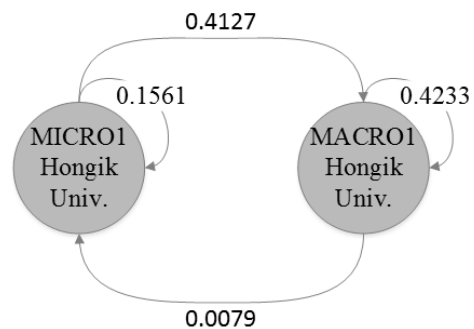


Fig. 5. Markov Model of Group Mobility Model of Nine

6. GPGPU를 이용한 분석 시간 단축

2 명의 집단 이동성 모델과 3명 이상의 집단 이동성 모델로 나타난 군집과 Markov 모델은 앞서 Fig. 1, 2, 3, 4를 통해 보였다. 2 명의 집단 이동성 모델을 도출해 내는데 사용된 작업 환경은 윈도우 서버 2012, 64 GB 램의 서버 컴퓨터이고, R Language를 사용했다 [13]. 실제 두 사람의 집단 이동성 모델을 만들 때의 작업 시간은 약 2 년간 수집한 위치 정보 2,343,043 개를 이용해서 집단 이동성 모델 도출에만 6 일이 소요됐다. 3 명 이상의 인원의 경우 R과 CUDA를 동시에 활용하기 위해서 Linux 운영 체제인 Ubuntu 14.04를 이용하고 CUDA toolkit과 R Language를 사용했다. 집단 모델 알고리즘을 병렬 처리 없이 동작시켰을 때, 9 명의 경우 약 3 주의 처리 시간이 필요한 것을 실험을 통해서 확인했다. 하지만 일명 R+GPU인 R과 CUDA를 활용한 병렬 처리 과정을 통해 처리 시간을 2 분까지 단축할 수 있었다. 다음 표 2는 위치 정보 수에 따른 분석 처리 시간을 표로 나타낸 것이다. 위치 수는 각각 2 명, 9 명을 분석했을 때의 분석에 사용한 위치 정보의 수이다. 이러한 이동성 모

델은 위치 정보를 수집한 자원자들에 의하여 검증되었다.

시간 단축은 위치 정보의 수가 적은 경우보다 많은 경우 큰 감소가 있었음을 볼 수 있다. 이런 변화를 확실하게 보여주기 위해서 참가자 수를 늘려서 실험하였다. 총 23 명의 실험 참가자의 위치 정보를 가지고 군집 모델을 생성하였고 시간 비교를 실시했다. R Language만 가지고 수행한 것과 R과 GPGPU를 활용하여 수행했을 때 각각의 동작 시간을 측정했고, 결과는 Table 3과 같다. 표에서 무한대로 나타낸 것들은 분석 시간이 너무 오래 걸려 확인 할 수 없는 것들이다. 참가자 수가 10 명 이상이 되는 경우 그 분석 시간을 측정하기가 어려웠지만 GPGPU를 활용한 경우 300 초, 즉 5 분 만에 결과를 확인할 수 있었다.

Table 2. Computing Processing Time for Group Mobility Model Establishment (Seconds)

Number of Position	R on CPU	R on GPGPU
2,343,043	518.400	0.739
23,712,540	1,512,000	113.521

Table 3. Computing Processing Time According to Participants

Participants	Number of positions	Execution time(SEC)	
		R	R+GPGPU
2	178,504	1433.414	12.879
3	302,639	2401.594	19.325
4	1,422,472	59675.140	49.340
5	1,691,600	74247.160	55.051
6	1,743,404	91702.480	64.107
7	2,040,674	149470.100	89.188
8	2,546,054	299037.000	153.191
9	3,793,946	646348.200	244.510
10	4,018,261	∞	300.023
11	4,025,324	∞	370.859
12	4,035,005	∞	402.704
13	5,216,166	∞	488.471
14	5,451,029	∞	561.322
15	5,493,722	∞	686.587
16	5,903,093	∞	854.020
17	6,062,186	∞	848.660
18	8,569,728	∞	898.847
19	8,740,887	∞	986.560
20	9,208,781	∞	1082.751
21	9,315,021	∞	1363.556
22	9,379,376	∞	1492.086
23	10,728,701	∞	1743.359

7. 결 론

본 논문에서는 개인이 수집한 위치 정보를 이용한 집단 이동성 모델을 생성하는 방법에 대해서 소개했다. 지금까지 위치를 이용한 연구 중에 다수의 인원을 가지고 집단의 개념을 분석에 이용한 경우는 이번이 처음이다. 위치는 개인이 GPS 수집 장치와 스마트폰을 이용해서 장기간동안 수집한 데이터를 이용한다. 집단 모델은 이전 연구인 개인 이동성 모델과 수집된 위치 정보를 가지고 집단 모델을 생성한다. 집단 모델 내에는 군집들이 서로 가까운 정도에 따라 Micro Cluster, Macro Cluster를 형성하였고, 각 군집의 중심 위치와 반경 등이 포함된 텍스트 파일과 지도 위에 그 모습을 표현한 시각화 파일을 결과로 출력했다. 또한 군집간의 이동 패턴을 분석해서 전이 확률 행렬을 만들고 이를 이용해 집단 모델 내 군집간의 Markov 모델을 생성할 수 있었다. 집단을 이루는 구성원의 수가 많아질 수록 분석해야 하는 위치 정보의 수도 많아질 것이기 때문에 분석 시간을 단축해야할 필요성이 있었다. 이를 위해서 기존 분석 tool인 R에 GPGPU를 활용한 R+GPGPU 기법을 활용해서 분석을 병렬 처리화시키고 분석 시간을 단축할 수 있었다.

앞으로 진행할 연구 방향은 다양한 참가자들을 모집해서 군집 모델을 만들 것이다. 그 후 집단에 포함되는 모든 인원이 공통적으로 방문한 군집이 아닌 사용자의 제어에 따라 많은 인원이 공통적으로 방문한 군집이 나타낼 수 있도록 할 것이다. 그러기 위해서는 기존 척도를 세워야하고 조건에 따른 다양한 방법의 처리 프로세스가 필요할 것이다.

집단 이동성 모델을 생성할 수 있으면 이동성 관련 분야에 많은 도움이 될 것으로 보인다. Location Based Service에서 객체의 속성, 예를 들면 나이, 성별, 성격 등으로 구분을 짓고 각 객체 집단의 이동 성향으로 속성에 맞는 차별화된 서비스를 제공할 수 있다. 또한 Markov 모델을 기반으로 위치 예측 모델에도 활용할 수 있을 것이다.

References

- [1] H. Y. Song, "Probabilistic space-time analysis of human mobility patterns," WSEAS TRANSACTIONS on COMPUTERS, vol.15, pp.222-238, 2016.
- [2] Wolf, P. S. A. and W. J. Jacobs, "GPS technology and human psychological research: a methodological proposal," *Journal of Methods and Measurement in the Social Sciences*, Vol.1, No.1, pp.1-15, 2010.
- [3] C. Song, Q. Zehui, B. Nicholas, and B. Albert-laszio, "Limits of predictability in human mobility," *Science* 19, Vol.327, No.5968, pp.1018-1021, 2010.
- [4] L. Liao, D. Fox, and H. Kautz, "Extracting places and activities from GPS traces using hierarchical conditional random fields," *International Journal of Robotics Research (IJRR)*, Vol.26, No.1, pp.119-134, 2007.

- [5] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo, "Mining user mobility features for next place prediction in location-based services," in *Proc. of 12th IEEE ICDM*, pp.1038-1043, 2012.
- [6] A. P. Dempster, N. M. Laird, and 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.
- [7] D. Y. Choi and H. Y. Song, "Defining Measures for Location Visiting Preference," *Procedia Computer Science*, Vol.63, pp. 142-147, 2015.
- [8] J. S. Lee and H. Y. Song, "Efficient Detection of Positioning Data Error," *Advanced Science Letters*, Vol.21, No.3, pp. 328-331, 2015.
- [9] J. H. Baik and H. Y. Song, "Mobility State Classification with Particle Filter," *New Developments in Computational Intelligence and Computer Science*, Vol.28, pp.75-82, 2015.
- [10] S. Y. Kim and H. Y. Song, "Predicting Human Locations with Big Five Personality and Neural Network," *Journal of Economics, Business and Management*, Vol.2, No.4, pp.273-280, 2014.
- [11] E. B. Lee and H. Y. Song, "An Analysis of the Relationship between Human Personality and Favored Location," *AFIN* 2015, pp.6-10, 2015.
- [12] D. Y. Kim, D. Y. Choi, and H. Y. Song, "Modeling Group Mobility from Individual Mobility Model," *KIPS*, Vol.21, No. 2, pp.376-379, 2014.
- [13] Cyril Zeller, "CUDA C/C++ Basics," Supercomputing 2011 Tutorial.
- [14] Paul Baines, "RCUDA: General programming facilities for GPUs in R," *Journal of Statistical Software*.
- [15] Roy D. Yates and David J. Goodman, "Probability and Stochastic Processes, Second edition," pp.445-500, 2005.



송 하 윤

e-mail : librarymusik@gmail.com

1991년 서울대학교 계산통계학과(학사)

1993년 서울대학교 전산과학(석사)

2001년 Computer Science Department,
University of California at Los
Angeles, USA(Ph.D)

2001년~현 재 홍익대학교 컴퓨터공학과 교수

2009년 Institute of Computer Technology, Vienna University
of Technology, Austria(객원 연구원)

관심분야 : 사물 인터넷, 인간 이동모델, 빅 데이터 분석, 딥러닝



김 동 엽

e-mail : hatddum@gmail.com

2015년 홍익대학교 컴퓨터공학과(학사)

2015년~현재 홍익대학교 컴퓨터공학과
석사과정

관심분야: 인간 이동모델, 병렬처리, 빅
데이터 분석