
Global Report

모빌리티 데이터 시뮬레이션을 위한 인구 이동성 모델링

1. 개요
2. 관련 연구
3. 모델링
4. 결과검토_spatio-측량 분포의 유사도
5. 결론

1. 개요

모빌리티(mobility : 이동성) 데이터는 집단 및 개별 수준에서 인간의 모빌리티와 관련된 많은 연구에 성공적으로 적용되어 왔다. 집단수준은 인구흐름의 분석을 포함한 각 개인의 여행(운송)에 초점을 맞추어 개인 간 상호작용을 통해 지역 환경에 다양한 영향<인구분포 매핑(2014년 Deville 등), 토지이용 분류(2017년 Ros & Muñoz), 동적 교통흐름 분석(2014년 Calabrese 등), 이동성 패턴 분석(2016년 Sifa-Nowicka 등), 지역사회 탐지(2010년 Ratti 등)>을 미친다. 개별수준의 모빌리티 연구는 행동 프로파일링(2012년 Furlotti/Gabrielli/Rinzivillo/Renso) 및 상호작용 분석(2011년 Calabrese/Smoreda/Blondel/Ratti)을 포함한다. 그러나 여기에서 특히 개별 이동궤적(경로)에 대한 접근은 심각한 개인 정보보호문제를 야기할 수 있다고 경고하고 있다[1]. 따라서 데이터 가용성(특히 휴대전화에 대한 접근)은 유럽연합, 미국 및 중국과 같은 법률에 의해 많은 국가에서 규제되고 있다[2][3][4][5]. 특히 엄격한 규정이 시행되고 있는 유럽연합에서는 개인정보가 포함되어 있지 않더라도 위치데이터는 개인정보로 간주한다. 이러한 처리의 복잡성으로 인해 일관된 모델링 프레임워크 개발 필요성이 대두되었다. 이 연구에서는 두 가지 주요 기능을 가진 새로운 3W (WHO-WHERE-WHERE) 모빌리티 모델링 방법을 제시한다.

- 실제 교통흐름과 모빌리티 통계를 반영하여 개인의 프라이버시를 보호할 수 있는 교통수요자 도출 방안
- 학습영역(인구조사, 택시 픽업 및 하차)에 대해 자유롭게 이용할 수 있는 정보 모델링을 통해 이동궤적의 형태로 개인의 이동성 도출방안

이러한 두 개의 데이터셋을 이용하여 3W(WHO-WHERE-WHERE) 모빌리티 모델을 평가하였다. 이를 위해 대규모 합성 데이터에 알고리즘을 적용하여 GPS(Global Positioning System) 트래커를 통한 실제 개별 이동궤적에 대한 데이터셋을 이용하여 실제 인구흐름을 파악하는 알고리즘의 성능을 검증하였다.

2. 관련 연구

모빌리티 데이터 프라이버시에 대한 현재 연구의 목적은 PPDP(Privacy-Preserving Data Publishing)의 원칙을 충족하는 것이다. 그간의 주요 관련 연구를 간단히 요약하면 다음과 같다:

- 2010년 Fung 등은 완전한 유용성을 유지하면서 개인의 프라이버시를 동시에 보호하는 방식으로 이동궤적을 처리할 수 있는 방식을 제시하였다[6].
- 2019년 Fiore 등은 개인 모빌리티 데이터의 개인정보보호에 대한 많은 접근법이 이루어지지 않았으며, PPDP 기준을 충족하는 솔루션을 제공하지 못하고 있다고 주장한 바 있다[7].

데이터의 공간적 및 시간적 해상도를 줄이는 것과 같은 간단한 정보처리 기법은 개인정보를 저장하지 않으면서 데이터 유틸리티에도 크게 영향을 미친다는 것은 분명하다. 가장 유망한 접근법은 2012년 Rui Chen 등이 제시한 실제 개별 운동궤적을 이용하여 인공궤적을 생성하는 것이다[8]. 합성 이동궤적 생성기법은 인공궤적을 생성하는 데 필요한 원시데이터의 일부를 생성하는 것이다. 이 방법의 주요장점은 표현에 대한 잡음(오류)과 실제 개별 운동궤적을 거의 구별할 수 없는 출력데이터의 형태를 이용하여 가장 엄격한 개인정보보호 기준을 채택할 수 있다는 것이다. 아울러 이러한 방법의 개인정보보호 수준은 특정 개인정보가 출력 단에서 유추될 수 없기 때문에 충분한 것으로 평가받고 있다. 하지만 합성된 데이터는 제한된 속성 세트만 유지하고 운동궤적에 대한 개별 특성은 복제되지 않는다. 나아가 교통수요자의 개인정보보호를 위한 모빌리티 모델인 WHERE(Work Home Extracted REgions)를 제안한다. 이 기법은 원시데이터의 확률분포를 기반으로 운동궤적을 생성하는 것이다. 즉, 집의 위치 및 통근거리와

같은 이동성을 기반으로 하는 에이전트 기반 모델이다. 이는 개인의 이동성 모델을 시뮬레이션 하여 입력데이터 특성을 복제할 수 있어 다른 개인정보보호 방법과 차별화된 기법으로 평가받고 있다. 특히 WHERE는 시간당 인구이동 분포 및 일일 이동범위와 같은 제한된 집단 이동성 통계집합만을 포함한 휴대전화 데이터만 복제하도록 설계되었다.

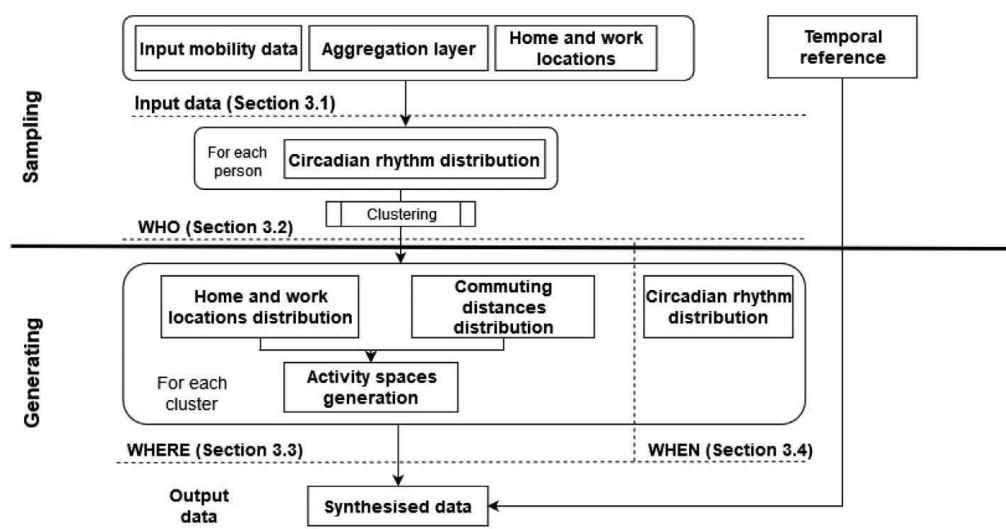
3. 모델링

3.1. 3W 생성 프로세스

개인(교통수요자)의 이동성(모빌리티)의 세 가지 측면(WHO-WHERE-WHEN)을 알고리즘의 기초로 이용되었다. 이를 모델 구성요소는 다음과 같다.

- WHO(Working HOurs) 쉬프트그룹은 유사한 시간적 이동성을 가진 사람들의 그룹을 추출한다.
- WHERE(Work Home Extracted REgions)는 이동성의 공간적 측면을 제어한다.
- WHEN(Work-HomE circadian) 리듬은 이동성의 시간적 측면을 제어한다.

3W(WHO-WHERE-WHEN) 생성 프로세스를 [그림 1]에 나타낸다. 이는 데이터 전처리단계를 포함하여 모든 모델 입력데이터 및 파라미터에 대한 설명이다. 이 모델은 두 단계로 나누어 모든 분포함수를 계산하여 이동궤적을 합성한다. WHO 구성요소는 샘플링을 담당하며, WHERE 및 WHEN은 궤적을 생성한다.



* 자료 : Kamil Smolak et al., “Population mobility modelling for mobility data simulation”, Computers, Environment and Urban Systems 84 (2020) 101526, p.3.

[그림 1] 3W(WHO-WHERE-WHEN) 생성 프로세스

3.2. WHO 구성요소

인간의 이동패턴은 사회경제적 지위, 사회관계 및 여행목적 등과 같은 많은 변수와 관련이 있다. WHO 구성요소의 아이디어는 유사한 모빌리티 패턴을 갖는 행동그룹과 전체 인구에서 자신의 몫을 캡처하는 것이다. 이 연구에서는 유사한 이동궤적과 일주기(circadian) 리듬을 갖는 교통수요자들을 대상으로 이동패턴을 도출하였다. 일주기(circadian) 리듬은 경험적 분포를 기반으로 세 가지 범주(HWO : Home/Work/Other)로 나뉜다. 이 세 가지 범주는 하루 중 일정시간에 교통수요자에게 할당된 확률분포를 의미한다. 입력 이동성 데이터를 분석하여 HWO 분포함수를 도출하는 과정은 다음과 같다.

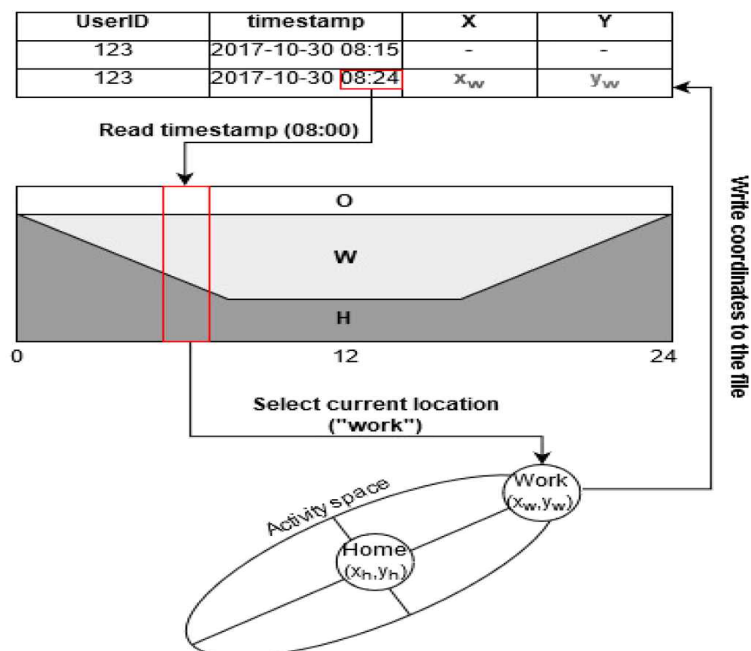
- 가정/직장/기타장소에 대한 세 개의 벡터를 만든다. 이때 각 벡터의 길이는 미리 설정된 해상도에 의해 결정된다.
- 가정 및 직장 위치를 제공된 정보에서 읽어낸 다음 교통수요자의 이동패턴을 통해 세 벡터 중 하나의 시간슬롯에 대한 시간과 장소를 확인한다.
- 이동패턴의 끝에 도달하게 되면 공통 벡터의 시간슬롯을 정규화 하여 각 집계기간에 대한 HWO 분포함수(확률분포함수)를 도출한다.

3.3. WHERE 구성요소

WHERE 구성요소 생성 프로세스는 우선적으로 합성될 총 인원수가 집계계층 공간에 수요자를 배치하는 단계에서 결정되어야 한다. 각 클러스터에 생성될 수요자 수는 상이한 이동성 그룹의 점유율을 재 확장하는 데 사용된다. WHERE 모듈은 생성된 각 수요자에 대해 가정 및 직장의 위치에 따른 활동공간을 할당한다. 이때 특별한 의미가 있는 장소는 합성된 그룹에 할당되기 이전에 생성된 분포함수를 이용하여 설정된다. 먼저 home location을 이용하여 가정의 위치를 선택한다. 이어서 선택한 위치에서 평균 통근거리(d)를 계산한다. 직장 위치는 가정 위치의 반경(d) 내에서 선택되며, 이때 반경의 폭은 평균 셀 크기에 의해 결정된다. 최종적으로 선택된 가정 및 직장 위치를 수요자의 활동공간을 구성하는 데 이용한다.

3.4. WHEN 구성요소

교통수요자의 활동공간은 해당 수요자가 이용할 수 있는 모든 가능한 위치를 결정하는 단서가 된다. 이 단계에서는 이러한 위치 간 이동성에 대한 시뮬레이션을 위해 수요자의 이동성 루틴에 대한 정보가 필요하다. 3W 모빌리티 시뮬레이션 알고리즘 구성을 [그림2]에 나타낸다.



- ※ 1) 시간은 합성된 클러스터에 대한 HWO의 시간슬롯을 선택하는 데 이용됨
- 2) 확률적으로 현재위치가 선택되며, 좌표는 출력파일에 기록됨

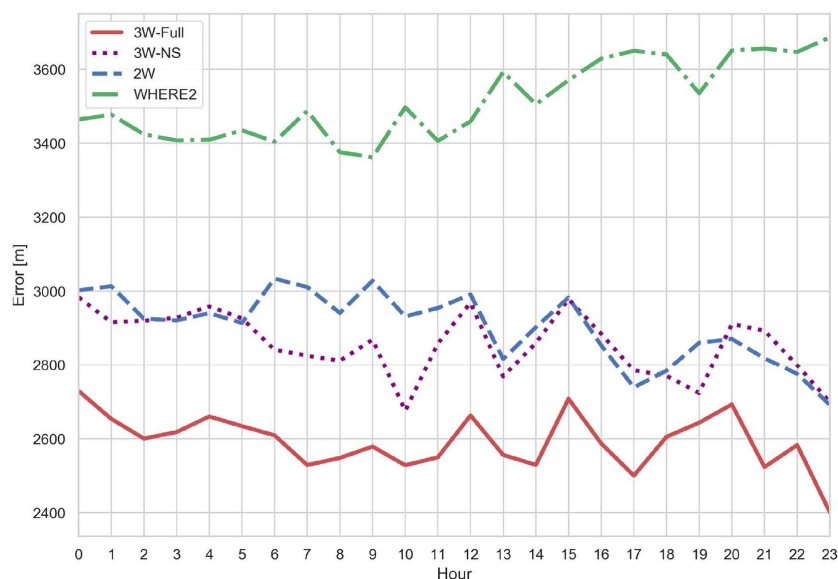
* 자료 : Kamil Smolak et al., "Population mobility modelling for mobility data simulation", Computers, Environment and Urban Systems 84 (2020) 101526, p.5.

[그림2] 3W 모빌리티 시뮬레이션 알고리즘 구성

3W 모빌리티 시뮬레이션 알고리즘은 생성된 클러스터의 HWO(가운데 부분)에서 적절한 시간슬롯을 가진 날짜와 시간을 일치시키며, 현재 위치는 가정, 직장 또는 다른 지역이 될 수도 있다. 다른 지역을 선택하면 수요자의 활동공간의 위치가 무작위로 선택된다. 임의의 위치 선택은 수요자에 대한 개인정보보호를 수용한다. 마지막으로 선택한 위치는 현재 타임스탬프와 함께 기록된다. 한 수요자에 대한 위치 참조가 끝에 도달하면 알고리즘은 다음 수용자로 이동하여 프로세스를 반복한다. 프로세스의 종점에서 출력파일에는 각 수요자의 이동궤적, 타임스탬프 및 할당된 좌표의 고유 식별자가 포함된 합성된 개별 이동궤적이 도출된다.

4. 결과검토_spatio-측량 분포의 유사도

입력 모빌리티 데이터의 유동인구 분포와 합성된 데이터셋을 기반으로 EMD(Earth Mover's Distance)를 계산한다. 이를 토대로 합성 모빌리티 데이터집합에 대한 EMD 거리 비교결과를 [그림 3]에 나타낸다.



* 자료 : Kamil Smolak et al., “Population mobility modelling for mobility data simulation”, Computers, Environment and Urban Systems 84 (2020) 101526, p.7.

[그림 3] 합성 모빌리티 데이터집합에 대한 EMD 거리 비교

주어진 실험환경에서 3W 모델은 WHERE 방법을 능가하여 평균정확도가 35%이상 향상되었음을 알 수 있다. 3W-Full 데이터집합의 평균 위치오차는 3W-Full 데이터집합의 경우 2,593m, WHERE 데이터셋의 경우 3,515m로 나타났다. 3W-NS 및 2W 변형으로 생성된 데이터집합에 대한 3W-전체 데이터집합의 이득은 모빌리티 그룹을 추출하는 데 매우 중요한 인자이다. 3W-Full을 기반으로 한 전체인구는 인공궤적을 생성하는 데 이용되는 입력데이터와 동일함을 알 수 있다. 2W 및 3W-NS 변형으로 생성된 데이터집합은 모든 시간슬롯에 걸쳐 유사한 EMD를 가지며, 3W-NS 데이터집합이 약간 더 나은 성능을 제공하는 것을 알 수 있다.(평균거리 2W: 2904m, 3W-NS: 2857m)

5. 결론

이 연구에서는 향상된 개인정보보호 규정을 준수하면서 인공궤적 합성기법을 통해 교통수요자의 이동성 모델링을 위한 WHO-WHERE-WHERE 방법을 제안하였다. 제안 모델은 도시규모의 교통수요자 이동성 모델링의 측면에 초점을 맞춘 WHERE 방법의 수정판이다. 이 연구에서는 유연성을 높이고 정확도를 향

상시킬 수 있는 다른 접근방식을 제안하였다. 아울러 개별수요자 운동궤적을 사용하여 합성 모빌리티 데이터집합에 대한 EMD 거리를 비교하여 과거의 테스트사례 대비 이 방법이 우수함을 검증하였다. 모빌리티 모델은 도시의 이동성(사람, 자동차 등) 관련 분야에 광범위하게 응용되고 있다. 그러나 모빌리티 데이터공유는 개인정보보호 문제로 인해 접근성이 제한된다. 이 논문에서는 향상된 프라이버시 보호 이동성인 WHO-WHERE-WHERE(3W) 모델을 제시하고 있다. 이 연구결과는 모빌리티 모델링을 통해 국내 대도시의 도심에서 이동성(사람, 자동차 등) 관련 정책에 벤치마킹 자료로 활용할 수 있을 것이다.

[참고문헌]

- [1] Rein Ahas, Siiri Silm, Olle Järv, Erki Saluveer & Margus Tiru, “Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones”, Journal of Urban Technology, Vol.17, 2010-Issue 1, 30 Mar. 2010, pp.3~27.
<<https://doi.org/10.1080/10630731003597306>>
- [2] European Commision (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/EC (general Da).
- [3] Snape, J. (2016). California penal code 2016 book 1 of 2. Lulu Press.
- [4] State of California (2015). An act to add Chapter 3.6 (commencing with Section 1546) to Title 12 of Part 2 of the Penal Code, relating to privacy.
- [5] Greenleaf, G., & Livingston, S. (2016). China’s new cyber security law {-} also a data privacy law? Technical report 144. University of New South Wales.
- [6] B. C. M., Wang, K., Chen, R., & Yu, P. S. (2010). Privacy-preserving data publishing: A survey of recent developments. ACM Computing Surveys (Csur), 42(4), 1-53.
- [7] Fiore, M., Katsikouli, P., Zavou, E., Cunche, M., Fessant, F., Le Hello, D., ... Stanica, R. (2019). Privacy of trajectory micro-data: A survey. arXiv preprint arXiv:1903.12211.
- [8] Rui Chen, Gergely Acs, and Claude Castelluccia (2012). Differentially private sequential data publication via variable-length n-grams. Proceedings of the 2012 ACM conference on computer and communications security (pp. 638-649).

<원문제목> Population mobility modelling for mobility data simulation

<원문출처> Computers, Environment and Urban Systems 84 (2020) 10152