

통신 신호 기반 이동 메타데이터를 활용한 인간 활동 추론

권나현, 여예린, 조예지, 윤보라, 김정현

세종대학교

nhkwon9803@sju.ac.kr, yerinee0949@sju.ac.kr, yezy0390@sju.ac.kr, 22012142@sju.ac.kr, j.kim@sejong.ac.kr

Human Activity Inference from Communication Signal-based Mobility Metadata

Nahyeon Kwon, Yerin Yeo, Yeji Cho, Bora Yoon, Junghyun Kim

Sejong Univ.

요약

본 논문은 GPS 외에도 이동통신 기지국 신호, 센서 신호 등 다양한 통신 기반 메타데이터를 활용하여 인간 활동을 추론하는 모델을 제안한다. 제안 모델은 AI Hub의 교통수단 판별 데이터를 기반으로, 시계열 및 이미지 형태의 입력 데이터를 구성하고, 이를 병렬 구조의 신경망으로 처리하여 이동 목적을 예측하였다. 10명의 사용자 데이터를 활용하여 실험한 결과, 85.50%의 정확도와 0.56의 macro F1-score, 0.86의 micro F1-score를 기록하며, 제안 모델이 인간 활동의 복합적 맥락을 효과적으로 반영함을 입증하였다.

I. 서론

인간의 일상 활동은 대체로 일정한 패턴을 따르며, 정기적인 이동은 집, 직장, 쇼핑 등과 같은 주요 목적지를 중심으로 이루어진다[1]. 이러한 이동 및 활동 정보를 분석하는 것은 혼잡 지역 식별, 이동 분포 예측, 교통 계획 수립 등 다양한 분야에 활용될 수 있다[2]. 특히, 모바일 기기를 통해 수집되는 다양한 센서 및 통신 데이터를 기반으로 인간의 활동을 추정하는 연구는 통신 기반 서비스의 고도화에 기여하고 있으며[3], 이는 인공지능 기반 통신 기법들의 발달로 더욱 가속화되고 있다[4].

기존 연구들은 인간 행동의 목적을 파악하기 위해 다양한 접근을 시도해 왔으며, 주로 GPS와 같은 단일 위치 정보에 의존했다[1, 2]. 그러나 이러한 정보들만으로는 고차원의 맥락 정보를 포착하는 데 한계가 있다[5]. 이러한 한계를 극복하기 위해서는 모바일 센서 데이터나 사용자 정보와 같은 다양한 데이터를 활용하는 것이 필요하다. 따라서 본 논문에서는 GPS 정보뿐만 아니라 연령대, 성별 등의 사용자 특성과 이동통신 신호, 교통수단 정보 등 다양한 메타데이터를 통합적으로 활용하여 인간 활동의 목적을 추론하는 모델을 제안한다. 이는 기존의 위치 기반 방법론을 확장하여, 더욱 풍부한 행동 맥락을 반영함으로써, 인간 활동의 정확한 활동 목적 추론과 패턴 인식이 가능함을 보이는 데에 목적이 있다.

II. 본론

통신 신호 기반 이동 메타데이터를 활용하여 인간의 활동을 추론하는 모델을 설계하기 위해, 본 연구에서는 AI Hub[6]에 공개된 ‘교통수단 판별 데이터’를 활용했다. 해당 데이터는 2022년에 모바일 앱 ‘Ctracker’를 통해 수집되었으며, 교통수단별로 구분된 global positioning system (GPS), 이동통신(base transceiver station, BTS), inertial measurement unit (IMU), 와이파이가 접속 및 중단 간(origin-destination, OD) 궤적 데이터로 구성되어 있다. 이 중 OD 궤적 데이터에는 ‘이동 목적’ 정보가 포함되어 있으며, 본 연구에서는 이를 정답 레이블로 활용하여 사용자의 활동 목적을 추론하였다. 데이터는 사용자의 출발지부터 도착지까지의 경로와 그에 대응되는 목적으로 구성되어 있으며, 이동 목적은 출퇴근, 등하교, 업무, 여가, 쇼핑, 학원, 기타의 총 7가지 범주로 분류된다. 이때, 우리는 출퇴근과 같이 주기적인 활동이 많은 30대 사용자 중 데이터 수가 가장 많은 10

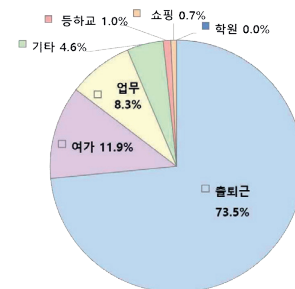


그림 1. 이동 목적별 데이터 분포 비율

명을 선정하여 실험에 활용했으며, 총 데이터 샘플 수는 175,860개이다.

먼저, 수집된 데이터에 인간 활동 예측에 활용 가능한 의미 있는 패턴이 존재하는지 확인하기 위해 다양한 시각화 분석을 수행하였다. 이동 목적별 데이터 분포는 그림 1에 나타나 있으며, 전체 중 ‘출퇴근’이 73.5%로 가장 큰 비중을 차지하고, ‘등하교’가 0.7%로 가장 적은 비중을 보인다.

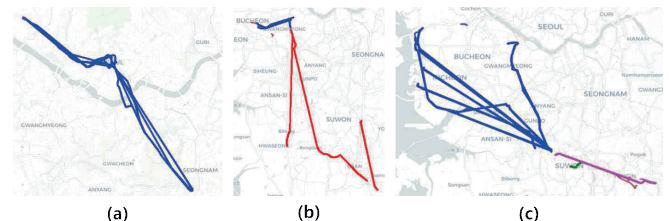


그림 2. 사용자들의 활동 패턴 시각화

이어서, 이동 목적에 따른 경로 패턴의 차이를 분석하기 위해, 일부 사용자들의 목적별 이동 경로를 그림 2에 시각화하였다. 그림에서 파란색은 ‘출퇴근’, 빨간색은 ‘업무’, 분홍색은 ‘여가’의 경로를 나타낸다. 그림 2-(a)의 사용자는 반복적으로 출퇴근하는 패턴이 명확하며 다른 활동은 나타나지 않는다. 그림 2-(b)의 사용자는 ‘출퇴근’의 패턴은 일정한 데에 반해, ‘업무’의 경로는 상대적으로 불규칙하다. 한편, 그림 2-(c)의 사용자는 ‘출퇴근’ 경로의 공간적 범위는 넓지만, 경로의 시작 지점이 특정 위치에서 반복적으로 나타나는 특징을 보인다. 이는 사용자마다 이동 목적에 따라 경로 패턴의 규칙성과 불규칙성이 혼재되어 있음을 시사한다.

우리는 예측 모델을 설계하기에 앞서, 다양한 센서에서 수집된 데이터를 동기화하고 통합하여 하나의 학습 가능한 데이터셋으로 구성하였다. 먼저 교통수단별로 나누어있는 데이터를 사용자 단위로 병합한 뒤, 센서마다 상이한 샘플링 간격을 1초 단위로 샘플링하여 시간 축을 기준으로 동기화하였다. 또한, 불필요하게 반복되는 정보를 줄이기 위해 OD 궤적 데이터의 타임스탬프에 해당하는 데이터 중 첫 번째 값만을 사용하였다. 사용자의 이동 패턴을 충분히 반영하기 위해 데이터를 10초 단위로 재샘플링하였고, 이후 OD 궤적을 고정된 길이로 자르는 윈도우 기반 시계열 분할 기법을 적용하였다. 윈도우 크기는 180, 스트라이드는 1로 설정하여 예측 모델이 최소 30분 이상의 이동 행적을 학습에 활용할 수 있도록 하였다. 또한, 궤적 내 시퀀스 길이가 180 미만인 경우, 부족한 부분은 0으로 패딩하여 하나의 시퀀스를 구성하였다. GPS의 위도, 경도, 고도 및 BTS의 신호 세기와 같은 궤도에 따라 변화하는 데이터에 결측이 존재할 경우, 인접 시점의 값을 이용하여 선형 보간법으로 보완하였다. 또한, 각 사용자의 주요 활동 반경을 기준으로 GPS 데이터를 정규화하여 상대적인 위치 정보를 제공하였으며, BTS 데이터도 함께 정규화하여 사용자의 활동 지역에 따른 상대적인 신호 세기를 나타냈다. 이와 함께, 타임스탬프로부터 시간대 및 주말 여부 등의 파생 변수를 생성하였다. 한편, 이동 경로의 시각적 패턴을 추론하기 위해, 정규화되지 않은 GPS 위도와 경도 데이터를 이용해 시계열 기반 이미지로 변환하였다. 마지막으로, 하나의 이동 목적을 가지는 OD 궤적을 데이터를 학습/검증/테스트로 무작위로 분할했으며, 각 데이터셋 간에는 궤적 내 경로가 겹치지 않도록 구성하였다. 그림 3은 분할된 데이터셋의 이동 목적별 분포를 나타내며, 그래프를 통해 각 데이터셋의 레이블 분포가 비슷하게 나뉜 것을 볼 수 있다.

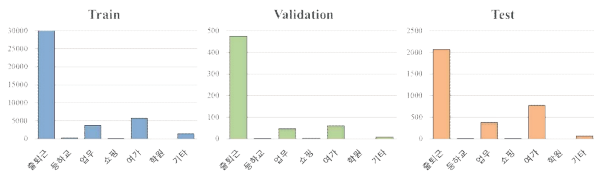


그림 3. 학습/검증/테스트 데이터의 이동 목적 분포

제안 모델의 전체 구조는 그림 4와 같다. 모델은 입력 데이터의 형태에 따라 세 가지 분기의 병렬 뉴럴 네트워크를 통해 각기 다른 특성을 효과적으로 추출하도록 설계되었다. 먼저 연령대, 주말 여부 등 OD 궤적 내에서 변하지 않는 특성은 두 개의 선형 레이어로 특성을 추출한다. 다음으로, 이미지로 변환한 GPS 이동 경로 데이터는 2차원의 합성곱 레이어를 이용해 공간적 이동 패턴을 학습한다. 마지막으로 사용자가 이동함에 따라 변화하는 BTS 신호 세기와 정규화된 GPS 위도, 경도, 고도와 같은 시계열 특성은 장단기 의존성을 모두 반영할 수 있는 LSTM 레이어를 통해 특성이 추출된다. 추출된 특성들은 모두 연결하여 두 개의 선형 레이어를 거친 후, 이동 목적을 예측하는 데 사용된다. 모델 학습에는 Adam 옵티마이저를 사용했으며, 학습률은 0.0001, L2 정규화 계수 0.03으로 설정하였다. 총 에포크는 40이며, 배치 크기는 256으로 설정하였다. 또한, 편향을 방지하기 위해 학습 데이터는 무작위로 순서가 섞이도록 했다. 모델 훈련에는 교차 엔트로피(cross entropy) 손실 함수를 사용하였다.

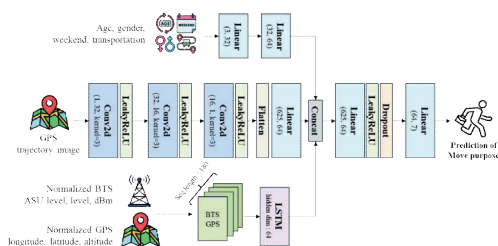


그림 4. 제안하는 이동 목적 예측 모델

모델의 성능을 검증하기 위해 정확도와 F1-score를 주요 평가지표로 사용하였다. 실험 결과, 제안 모델은 85.50%의 예측 정확도를 기록하였으며, macro F1-score는 0.56, micro F1-score는 0.86으로 나타났다. 이와 같이 macro F1-score가 micro F1-score에 비해 성능이 낮게 나타난 것은 레이블 간 성능 편차가 크다는 것을 의미한다. 레이블별 예측 성능은 그림 5과 같다. 그래프에 따르면, 대부분의 ‘출퇴근’은 높은 정확도로 잘 예측되는 반면, ‘여가’ 및 ‘기타’ 등의 목적은 오답이 상대적으로 많았다. 이는 ‘출퇴근’ 패턴에 비해 ‘여가’와 ‘쇼핑’ 등의 패턴이 상대적으로 드물게 나타나는 비정형적 패턴이기 때문으로 해석된다.

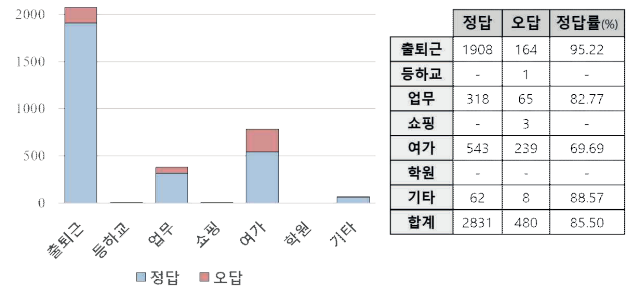


그림 5. 이동 목적별 정답 및 오답 분포

III. 결론

본 논문에서는 GPS 기반의 위치 정보뿐만 아니라 통신 신호 데이터와 사용자 정보 등의 메타데이터를 활용하여 인간의 활동 패턴을 인식하는 모델을 제안한다. 우리의 제안 모델은 실험 결과 정확도 85.50%, macro F1-score 0.56, micro F1-score 0.86으로, 사용자의 다양한 정보를 활용하여 인간의 패턴을 파악할 수 있음을 보였다. 향후 연구에서는 사용자 수를 확대하고, IMU와 같은 센서 데이터를 추가한다면 예측 성능이 향상될 것이라고 기대된다. 또한, SHAP, LIME 등의 XAI 기법을 통해 모델의 해석 가능성을 높일 계획이다.

참 고 문 헌

- [1] Zong, F., Tian, Y., He, Y., Tang, J., and Lv, J., "Trip destination prediction based on multi-day GPS data," *Physica A: Statistical Mechanics and its Applications*, vol. 515, no. 1, pp. 258-269, Feb. 2019.
- [2] Chen, Y., Jin, Z. and Li C., "Trip Purpose Prediction Based on Hidden Markov Model with GPS and Land Use Data," in *Proc. International Conference on Intelligent Transportation Engineering (ICITE)*, Beijing, China, pp. 55-59, Sep. 11-13, 2020.
- [3] Yoshidomi, S., Shirakawa, H. and Kasamatsu, D., "A Prediction Method of User Destination Using Historical Crowd Movement," in *Proc. Global Conference on Consumer Electronics (GCCE)*, Nara, Japan, pp. 58-59, Oct. 10-13, 2023.
- [4] Lee, H., Lee, B., Yang, H., Kim, J., Kim, S., Shin, W., Shim, B., and Poor, H. V., "Towards 6G Hyper-Connectivity: Vision, Challenges, and Key Enabling Technologies," *Journal of Communications and Networks*, vol. 25, no. 3, pp. 344-354, Jun. 2023.
- [5] Wei, H., Lei, G., Qin, L. and Tian, L., "Cluster Analysis of Trip Purpose Based on Residents' Travel Characteristic," in *Proc. International Conference on Intelligent Transportation Engineering (ICITE)*, Beijing, China, pp. 574-579, May. 21-22, 2022.
- [6] AI Hub, "교통수단 판별 데이터", AI Hub, 2024. [Online] Available: <https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&toPMenu=100&aihubDataSe=data&dataSetSn=71545>