

# GPS 기반 이동수단 분류 방법 및 수집 주기 최적화 연구

정치윤<sup>1</sup>, 김무섭<sup>1</sup>, 정현태<sup>1</sup>, 정승은<sup>2,\*</sup>

## GPS-based Mobility Mode Classification and Sampling Rate Optimization

Chi Yoon Jeong<sup>1</sup>, Mooseop Kim<sup>1</sup>, Hyun Tae Jeong<sup>1</sup>, and Seungeun Chung<sup>2,\*</sup>

### 요약

사용자의 이동수단을 분류하여 그 특성을 이해하는 것은 여행 패턴 분석, 목적지 및 이동 경로 예측 등 위치기반 서비스를 제공하기 위한 핵심정보로 활용될 수 있다. 따라서 GPS 데이터를 사용하여 사용자의 이동수단을 분류하는 방법에 관한 관심과 연구들이 증가하고 있다. 사용자의 이동수단을 분류하는 방법은 입력 GPS 데이터의 샘플링 주기에 따라 분류 정확도가 달라지며, GPS 데이터의 샘플링 주기는 수집 기기의 배터리 소모에 영향을 미친다. 따라서 본 논문에서는 GPS 궤적정보를 사용하여 이동수단을 분류하는 새로운 방법을 제안하고, GPS 데이터의 샘플링 주기가 이동수단 분류 방법의 정확도 및 에너지 소비에 미치는 영향을 분석하였다. 이를 위하여 본 논문에서는 GPS 궤적으로부터 속도, 가속도, 저크를 계산하여 다양한 통계적 특징을 추출하였으며, 이동 시간을 고려한 이동 방향의 변화 비율, 정지 비율, 속도 변화율을 특징으로 활용하는 이동수단 분류 방법을 제안하였다. 실험에서는 제안 방법과 기존의 방법을 다양한 기계학습 알고리즘을 사용하여 비교하였으며, 랜덤 포레스트 분류 알고리즘을 사용하는 경우 제안 방법이 기존 방법보다 2.3%~3% 이상 높은 정확도를 가지는 것을 확인하였다. 또한, 제안 방법을 사용하여 GPS 데이터의 샘플링 주기 변화에 따른 이동수단 분류 정확도 및 배터리 소비량을 분석하였다. 실험 결과, 샘플링 주기를 10초로 설정하면 분류 정확도는 최고 정확도 대비 0.6% 감소하지만, 배터리 소비량은 33% 이상 감소하여 이동수단 분류 정확도의 손실 없이 에너지 소모를 최소화하는 것을 확인하였다.

### Abstract

Understanding the characteristic of users' mobility mode is a requisite of Location-based Services (LBS) such as destination estimation, travel pattern and trajectory analysis. The classification accuracy of mobility mode is highly affected by the GPS sampling rate, which has a significant effect on the battery consumption of the device. Therefore, this paper presents a new method to extract mobility features from GPS trajectories for mobility mode classification, and analyze the effect of GPS

<sup>1</sup>한국전자통신연구원 인공지능연구소 휴먼증강연구실 (책임연구원)

<sup>2</sup>한국전자통신연구원 인공지능연구소 휴먼증강연구실 (선임연구원)

\*Corresponding Author : chung@etri.re.kr

접수일자 : 2020. 08. 10.

1차 심사 : 2020. 09. 11.

2차 심사 : 2020. 10. 29.

게재확정 : 2020. 11. 16.

DOI: <http://data.doi.or.kr/10.22733/JITAE.2020.10.02.005>



---

sampling rate on the classification accuracy and battery consumption. Specifically, we propose features including Heading Change Rate (HCR), Stop Rate (SR), and Velocity Change Rate (VCR) based on moving duration, in addition to statistical features such as velocity, acceleration, and jerk measured from the GPS trajectories. The proposed method is compared with previous work using various machine learning algorithms, and the experiment results show that the proposed method shows 2.3%~3% higher accuracy with random forest classification. Also, the results demonstrate that setting the GPS sampling rate to 10 seconds reduces classification accuracy by 0.6% compared to the highest accuracy, but reduces battery consumption by 33%, minimizing battery consumption without loss of classification accuracy.

**Keywords:** Mobility Modes, Trajectory Mining, GPS Sampling Rate, Pattern Recognition

## 1. 서론

일상생활에서 사용자가 항상 휴대하여 사용하는 스마트폰, 스마트워치 등 모바일 기기들의 GPS (Global Positioning System) 탑재가 보편화되면서 GPS 데이터를 활용한 사용자의 위치 및 궤적정보 분석에 관한 관심이 증가하고 있다[1]. 특히 사용자의 이동정보는 사용자 행동의 시공간적 특성, 대중교통의 운행 상태 및 환경 이해를 위한 중요한 요소이며[2], 이동수단(Mobility mode)을 분류하여 그 특성을 이해하는 것은 여행 패턴 분석[3,4], 목적지 및 이동 경로 예측[5] 등 위치기반 서비스에 중요한 정보를 제공해줄 수 있다. 따라서 GPS 데이터를 사용하여 사용자의 이동수단을 분류하는 방법에 관한 관심이 지속적으로 증가하고 있다.

사용자의 이동수단을 분류하는 방법은 전통적인 특징(Feature)기반의 분류 방법과 심층학습 기반의 분류 방법으로 구분될 수 있다. 전통적인 특징기반의 분류 방법[6,7,8]은 GPS 데이터로부터 속도, 가속도, 평균 속도, 평균 가속도 등 전문가가 정의한 특징을 추출한 후, 결정 트리(Decision Tree), 지지 벡터 기계(Support Vector Machines) 등을 사용하여 이동수단을 분류하였다. 이동수단 분류를 위해 사용되는 특징들은 GPS 샘플링 주기의 영향을 많이 받기 때문에 GPS 샘플링 주기에 따라서

이동수단 분류의 정확도가 달라질 수 있다. 예를 들어 샘플링 주기가 길어지면 정류장에 자주 정차하는 버스의 이동 특성이 완화되어 승용차와의 구분이 모호해질 수 있다.

심층학습 기반의 이동수단 분류 방법[9,10,11]은 GPS 데이터에서 위치 정보와 속도 정보를 추출하여 2차원 또는 3차원의 이미지 형태로 변환하고 합성곱 신경망(Convolutional Neural Networks)의 입력으로 사용하여 이동수단을 분류하였다. 심층학습 기반 이동수단 분류 방법도 원본 GPS 데이터를 가공하여 이산화(Bining)하는 과정에서 데이터의 샘플링 주기에 영향을 받기 때문에 특징기반의 분류 방법과 같이 GPS 데이터의 샘플링 주기에 따라서 이동수단의 분류 성능이 달라질 수 있다.

따라서 사용자의 이동수단 분류 방법의 성능은 모델의 학습 방법과 상관없이 GPS 데이터의 샘플링 주기에 영향을 받으며, GPS 데이터의 샘플링 주기는 수집 기기의 배터리 소모에 직접적인 영향을 미친다. GPS 센서는 모바일 기기에 장착된 센서 중 배터리를 가장 많이 소모한다고 알려져 있으나[12], 모바일 플랫폼에서 이동수단 분류 방법의 정확도와 배터리 소모를 고려한 GPS 기반 이동수단 분류 방법의 수집 주기 최적화에 관한 연구는 간과되고 있다. 따라서 본 논문에서는 GPS 데이터의 샘플링 주기가 이동수단 분류 방법의 정확도에 미치는 영향을 분석하였다. 또한, GPS 데이터의

샘플링 주기가 모바일 기기의 배터리 소모에 미치는 영향을 분석하여 이동수단 분류 방법의 최적 GPS 수집 주기를 제안하였다.

본 논문의 구성은 다음과 같다. 2절에서는 이동수단 분류 방법과 관련된 연구 동향에 관하여 살펴보고, 3절에서는 본 논문에서 제안한 이동수단 분류 방법을 소개한다. 4절에서는 GPS 데이터의 샘플링 주기에 따른 이동수단 분류 정확도와 에너지 소비량의 실험 결과를 분석하고, 5절에서는 결론 및 향후 연구 방향을 제시한다.

## 2. 관련 연구

GPS 궤적정보를 사용하여 이동수단을 분류하는 연구는 GPS의 궤적정보에 따라서 단일 모드 궤적(Single-mode-based trajectory)과 다중 모드 궤적(Mixed-mode-based trajectory)으로 구분된다[1]. 단일 모드 궤적은 한 종류의 이동수단으로부터 생성된 것을 가정하여 궤적을 이동수단별로 나누는 과정이 필요하지 않지만, 다중 모드 궤적은 다수의 이동수단에서 생성된 궤적을 포함하여 이동수단의 변경을 탐지하고 궤적을 분할하는 방법이 필요하다. 따라서 단일 모드 궤적이 GPS 데이터의 샘플링 주기에 따른 분류 성능과 에너지 소모를 분석하는데 더 적합하여 본 연구에서는 단일 모드 궤적을 분류하는 방법에 관하여 주로 기술한다.

단일 모드 궤적 기반의 이동수단 분류 분야에서 전통적인 방법은 궤적으로부터 속도 및 가속도 정보를 생성하여 통계적인 특징을 추출하고 기계학습 방법을 사용하여 분류하는 방법이다. 마이크로소프트[6]에서는 GPS 데이터로부터 속도, 거리, 평균 속도 등의 기본적인 특징과 이동 방향의 변화 비율(HCR, Heading Change Rate), 정지 비율(SR, Stop Rate), 속도 변화율(VCR, Velocity Change Rate) 등의 추가적인 특징을 추출한 후 결정 트리를 사용하여 이동수단을 분류하는 방법을 제안하였다. 이동수단 분류 방법의 성능 검증을 위하여 65명의 실험자로부터 10개월 동안의 GPS

로그를 수집하였으며, 2초마다 기록된 GPS 로그의 양은 약 2,000시간 정도 규모이다. 이동수단 분류 정확도 분석을 위하여 수집된 데이터의 70%를 학습 데이터로 사용하고 나머지 데이터를 테스트 데이터로 사용하였다. 4가지 이동수단(걸기, 자전거, 버스, 자동차)을 결정 트리 모델을 사용하여 분류한 결과 71%의 정확도를 보여주었으며, 그래프 기반의 후처리를 적용하는 경우 75%의 정확도를 보여주었다.

GPS 데이터와 가속도계(Accelerometer) 정보에 임베딩 기법을 적용한 방법과 특징기반의 방법의 성능을 비교 분석한 연구도 진행되었다[7]. 임베딩 기법의 경우 GPS 데이터를 윈도우 단위 정보(Movelets)로 변환한 후, 학습 데이터로부터 이동수단을 분류할 수 있는 윈도우 단위 정보 사전을 생성하고 이를 사용하여 이동수단을 분류하였다. 임베딩 기법과의 비교를 위한 특징기반 방법은 GPS 데이터로부터 속도 및 가속도 정보를 계산하고, 각각의 평균값, 중간값, 최댓값, 최솟값 등의 통계적 특징 정보를 추출하여 결정 트리와 k-근접이웃(k-Nearest Neighbors) 방법을 사용하여 이동수단을 분류하였다. 이동수단 분류 성능을 분석하기 위하여 6명의 실험자로부터 1Hz 주기로 GPS 데이터를 수집하고, 5Hz 주기로 가속도계 정보를 수집하였으며, 수집 데이터 중 60%는 학습, 20%는 검증, 나머지는 테스트 데이터로 사용하였다. 5개 범주(걸기, 자전거, 자동차, 버스, 기차)에 대한 이동수단 분류 실험 결과 임베딩 기법 보다 특징기반의 방법이 더 좋은 성능을 보여주었으며, 가속도의 평균 변화 및 80번째 속도 등이 이동수단을 분류하는 중요한 특징임을 확인하였다.

사전에 수집된 GPS 데이터를 분석하는 방법과 달리 실시간으로 입력되는 GPS 데이터를 효과적으로 분류하는 방법에 관한 연구도 진행되었다[8]. 제안 방법은 GPS 신호로부터 속도와 가속도 정보를 생성한 후, 각각의 평균값, 분산, 최댓값, 중간값 등의 정보와 이동 방향의 변화 비율 및 정지 비율을 이동수단을 분류하는 특징으로 사용하였다. 이동수단의 분류 성능을 검증하기 위한 데이터로는 1~5초마다

GPS 로그를 기록한 GeoLife 데이터셋[13, 14,15]을 사용하였으며, 데이터를 순차적으로 전송하면서 이동수단의 분류 성능을 측정하였다. 6개 이동수단 범주(기차, 걷기, 자전거, 버스, 자동차, 지하철)를 가지는 사전 수집된 데이터에 랜덤 포레스트(Random Forests) 방법을 사용하여 이동수단을 분류한 결과 정확도는 약 87.1% 정도를 보여주었다. 실시간 GPS 데이터의 경우 데이터를 분할하는 윈도우 크기가 중요하기 때문에 다양한 윈도우 크기에 대한 분류 정확도를 분석하였으며, 2분 단위로 이동수단을 분류할 때 분류 정확도는 75%로 가장 높은 결과를 보여주었다.

최근 심층학습이 영상분석 및 음향분석 등 다양한 분야에 적용되어 기존의 특징기반 방법들보다 뛰어난 성능을 보여주고 있다[16,17]. 따라서 이동수단 분류 분야에서도 심층학습을 사용한 연구들이 진행되고 있다. 심층학습을 이용한 이동수단 분류의 초기 연구는 심층학습을 특징 추출 단계에 적용하는 연구들이 주를 이루었으며, 최근에는 특징 추출 및 분류 단계에 심층학습을 적용하는 연구들이 진행되고 있다.

초기 연구[9]에서는 GPS 데이터를 사용하여 2차원의 궤적 이미지를 생성한 후 적층잡음제거 오토인코더(Stacked denoising autoencoder)를 사용하여 심층 특징(Deep feature)을 추출하고, 이를 이동수단 분류를 위한 특징으로 사용하였다. 이동수단 분류 단계에서는 심층 특징과 함께 기존 방법에서 제안된 특징[6]을 회귀 분석(Logistic Regression)의 입력 값으로 사용하여 이동수단을 분류하였다. 이동수단 분류 성능 실험을 위해서 GeoLife 데이터셋과 일본 간토 지역에서 30명의 사용자를 대상으로 20일 동안 3초 주기로 수집한 GPS 로그를 사용하였다. 실험 결과, 심층학습으로부터 추출된 특징이 기존 방법의 특징보다 높은 분류 성능을 보여주었으며, 심층 특징과 기존 방법의 특징을 통합하여 사용하는 경우 가장 높은 성능을 보여주었다.

심층학습을 이동수단을 분류하기 위한 심층 특징 추출기로 사용하는 방법과 달리, 심층학습을 사용하여 GPS 궤적정보로부터 특징 추출

및 이동수단 분류의 전 과정을 수행하려는 연구도 진행되었다[10]. 제안 방법에서는 1~5초 단위의 조밀한 GPS 로그로부터 속도, 가속도, 저크(Jerk) 및 이동 방향의 변화율을 합성곱 신경망의 입력으로 사용하여 5가지 이동수단(걷기, 자전거, 버스, 자동차, 기차)을 분류하였다. 합성곱 신경망을 사용한 방법이 기존의 특징기반 방법보다 높은 성능을 보여주었으며, 합성곱 신경망의 앙상블 모델을 사용하는 경우 84.8%의 정확도를 보여주었다.

최근 GPS 데이터의 지리적 위치 정보(위도와 경도)를 이미지의 가로축과 세로축에 매핑하고, 각 영역에서의 이동 방향과 속도를 이미지 채널의 명도(Intensity) 정보로 활용하여 이미지를 생성하고, 이를 합성곱 신경망의 입력으로 사용하여 이동수단을 분류하는 방법[11]이 제안되었다. 제안 방법에서는 입력 이미지의 크기를 100 x 100으로 설정하였으며, 3개의 합성곱 계층과 1개의 전 연결 계층을 사용한 신경망을 사용하여 이동수단을 분류하였다. 이동수단은 버스와 그 외 이동수단으로 구분하였으며, 분류 정확도는 약 96%를 보여주었다.

GPS 궤적정보를 사용하여 이동수단을 분류한 기존의 특징기반 및 심층학습 기반의 연구들은 모두 조밀한 GPS 로그를 가정하고 있으며, GPS 데이터의 샘플링 주기는 수집 기기의 배터리 소모에 영향을 준다. 따라서 GPS 데이터의 샘플링 주기가 이동수단 분류 방법의 정확도 및 배터리 소모에 미치는 영향에 대한 분석이 필요하다.

### 3. GPS 기반 이동수단 분류 방법

GPS 궤적정보를 심층학습을 활용하여 이동수단을 분류하는 방법이 제안되었지만, 심층 신경망의 학습을 위한 데이터 부족 및 신경망의 깊이가 깊지 않아서 특징기반의 이동수단 분류 방법보다 뛰어난 성능을 보여주지 못하였다. 또한, 심층학습의 경우 특징기반의 방법보다 연산량과 에너지 소비량이 많다고 알려져 있다[18, 19]. 따라서, GPS 데이터의 샘플링 주기가 이

동수단의 분류 정확도 및 수집 기기의 에너지 소비에 미치는 영향을 심층적으로 분석하기 위하여 본 논문에서는 특징기반 이동수단 분류 방법을 사용하였다.

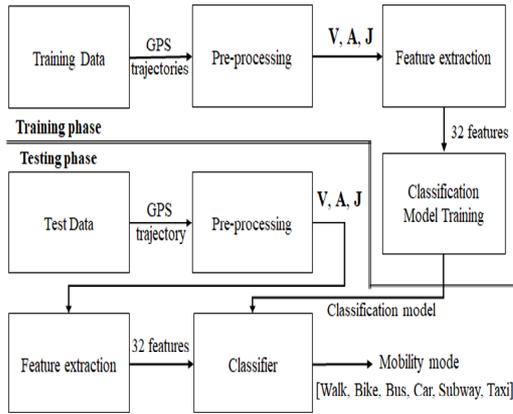


그림 1. 제안 방법의 흐름도.  
Fig. 1. Flowchart of the proposed method.

본 논문에서 제안하는 특징기반 이동수단 분류 방법은 그림 1과 같다. 제안 방법은 GPS 궤적정보를 사용하여 이동수단 분류 모델을 학습하는 단계와 GPS 궤적이 입력될 때 학습된 분류 모델을 사용하여 이동수단을 분류하는 단계로 구성된다. 학습 단계에서는 전처리 과정을 통하여 GPS 궤적 데이터들로부터 속도, 가속도, 저크 등의 정보를 계산하고, 특징 추출과정을 통하여 이동 수단 분류를 위한 32개의 특징을 계산한다. 이동수단 분류 특징은 기계학습 방법의 입력으로 사용되어 분류 모델 학습을 위하여 사용되며, 생성된 분류 모델은 분류 단계에서 사용된다. 이동수단 분류 단계에서는 학습 단계와 동일한 전처리 및 특징 추출과정을 수행하며, 사전에 학습된 분류 모델을 사용하여 이동수단 범주를 판단한다.

전처리 과정에서는 위도와 경도 및 시간 정보를 가지는 GPS 데이터들로 구성된 궤적에서 인접한 GPS 데이터를 사용하여 속도, 가속도, 저크를 계산한다. 연속적인 GPS 데이터 P1과 P2 및 시간 간격  $\Delta t$  가 주어졌을 때, 두 점의 거리는 Vincenty 공식[20]을 통하여 계산되며 P1의 속도, 가속도, 저크는 아래와 같이 정의된다.

$$S_{p1} = \text{Vincenty}(P_1, P_2) / \Delta t$$

$$A_{p1} = (S_{p2} - S_{p1}) / \Delta t$$

$$J_{p1} = (A_{p2} - A_{p1}) / \Delta t$$

전처리 과정에서는 위의 식을 사용하여 GPS 궤적을 구성하는 모든 점의 속도 S, 가속도 A, 저크 J 정보를 계산하였다. 기존의 특징기반 이동수단 분류 방법들은 속도 및 가속도 정보에서만 통계적 특징을 추출하였지만, 제안 방법에서는 기존 연구내용에 추가로 가속도의 변화량을 고려하기 위하여 저크를 계산하여 사용하였다.

특징 추출과정에서는 속도 및 가속도와 저크 정보로부터 기존 방법에서 많이 사용한 평균값, 분산, 최댓값, 최솟값, 중간값, 20분위, 80분위, 사분 범위(Interquartile range) 등의 특징을 추출하였다. 기존 연구의 발표에 따르면 GPS 궤적으로부터 계산되는 이동 방향의 변화 비율, 정지 비율, 속도 변화율 등은 이동수단을 분류하는 데 효과적이라고 알려져 있다[6]. 따라서 본 논문에서도 이동 방향의 변화 비율, 정지 비율, 속도 변화율을 이동수단 분류를 위한 특징을 사용하였으며, 아래와 같이 정의된다[6].

$$HCR_D = |P_c| / \text{Distance}$$

$$SR_D = |P_s| / \text{Distance}$$

$$VCR_D = |P_v| / \text{Distance}$$

위의 식에서  $P_c$ 는 이동 방향의 각도가 임계값 보다 크게 변화한 GPS 점들의 집합을 의미하며,  $|P_c|$  는  $P_c$ 의 원소 수를 의미한다.  $P_s$  는 이동 속도가 임계값 보다 작은 GPS 점들의 집합을 의미하며,  $P_v$ 는 이동 속도의 변화가 임계값 보다 큰 GPS 점들의 집합을 의미한다. 기존 연구에서는 이동 방향의 변화 비율, 정지 비율, 속도 변화율을 단위 거리당 횡수로 계산하였지만, 단위 시간당 비율도 중요한 정보로 활용할 수 있으므로 본 논문에서는 다음과 같이 시간을 고려한 이동 방향의 변화 비율, 정지 비율, 속도 변화율을 계산하였다.

$$HCR_T = |P_c|/Duration$$

$$SR_T = |P_s|/Duration$$

$$VCR_T = |P_v|/Duration$$

이동수단 분류를 위하여 사용된 특징을 정리하면 표 1과 같다. 이동수단 분류를 위하여 전체 이동 거리와 평균 속도, 속도 및 가속도와 저크 정보로부터 추출되는 특징과 본 논문에서 제안한 시간을 고려한 이동 방향의 변화 비율, 정지 비율, 속도 변화율 등 총 32개의 특징을 이동수단 분류를 위하여 사용하였다. 이동수단 분류를 위한 특징은 기계학습 알고리즘의 입력으로 사용하였으며, 본 논문에서는 객관적인 성능 비교를 위하여 기존 연구에서 주로 사용된 결정 트리 및 랜덤 포레스트, k-근접이웃을 사용하여 이동수단 분류 성능을 분석하였다.

표 1. 이동수단 분류를 위해 사용한 특징.  
Table 1. Features used for mobility mode classification.

Category	Features
Basic features from a segment	Distance, Average velocity
Basic features from velocity, acceleration and jerk	Mean, Variance, Maximum, Minimum, Interquartile range, Median, 20 <sup>th</sup> & 80 <sup>th</sup> percentile
Advanced features	HCR <sub>D</sub> , SR <sub>D</sub> , VCR <sub>D</sub> , HCR <sub>T</sub> , SR <sub>T</sub> , VCR <sub>T</sub>

## 4. 실험 결과

### 4.1 이동수단 분류 방법의 정확도 분석

이동수단 분류 방법의 성능 검증을 위하여 공개된 데이터셋 중 가장 큰 규모인 GeoLife 데이터셋을 사용하였다. GeoLife 데이터셋은 마이크로소프트에서 5년 동안 182명의 사용자로부터 위치 정보를 수집하여 구축하였다. 원시 (Raw) 데이터의 경우, 총 17,621개의 궤적으로 구성되며 전체 궤적의 거리는 1,292,961km, 전체 궤적의 시간은 약 5만 시간 정도이다. 수집된 데이터의 91.5%가 조밀하게 기록되었으며, 1~5초 사이 또는 5~10m의 이동이 있을 때 마다 기록되어 있다. 하지만 원시 데이터 중 이

동수단 레이블이 존재하는 사용자는 단 38%의 사용자에게 해당하는 69명으로 실제 성능 검증에 활용 가능한 데이터의 수는 제한적이다. GeoLife 데이터셋은 11개 이상의 이동수단으로 구분되어 있지만, 본 연구에서는 데이터의 양이 적은 비행기, 보트, 모터사이클 등의 범주를 제외하고 6개 범주(걷기, 자전거, 버스, 자동차, 지하철, 택시)의 데이터를 실험에 사용하였다. 또한, 특징 추출을 위한 최소 데이터 수를 확보하기 위하여 GPS 궤적 중 로그가 50개 이상, 120초 이상의 길이를 가지는 궤적을 추출하여 사용하였다. 실험에 사용된 궤적은 전체 궤적 중 44.6%에 해당되는 7,866개이며, 이동수단 범주별 분포는 그림 2와 같다.

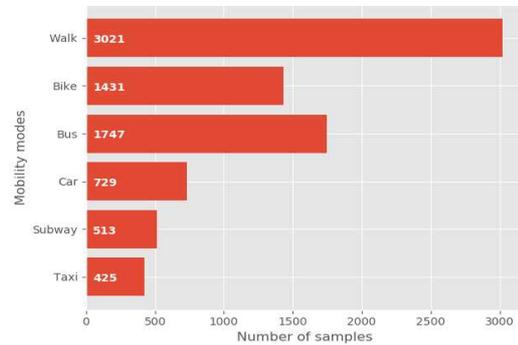


그림 2. GeoLife 데이터셋 중 실험에 사용된 데이터의 구성.

Fig. 2. Details of data used in the experiments in the GeoLife dataset.

본 논문에서 제안한 특징기반의 이동수단 분류 방법의 성능을 살펴보기 위해서 기존의 Zheng 방법(6)과 Erdelic 방법(8)과의 성능을 비교하였다. Zheng 방법은 속도와 가속도 데이터로부터 거리, 평균 속도, 분산, 최고 속도 3개, 최고 가속도 3개 등의 기본 특징과 이동 방향의 변화 비율, 정지 비율, 속도 변화율 등의 특징을 추출한다. Erdelic 방법은 속도와 가속도 데이터로부터 평균, 표준편차, 중간값, 최댓값 등의 기본 특징과 이동 방향의 변화 비율과 정지 비율을 특징으로 추출한다. 제안 방법과 기존 방법 모두 추출된 특징의 분류 알고리즘으로 세 개의 기계학습 방법(결정 트리, 랜덤 포레스트, k-근접이웃)을 사용하였다.

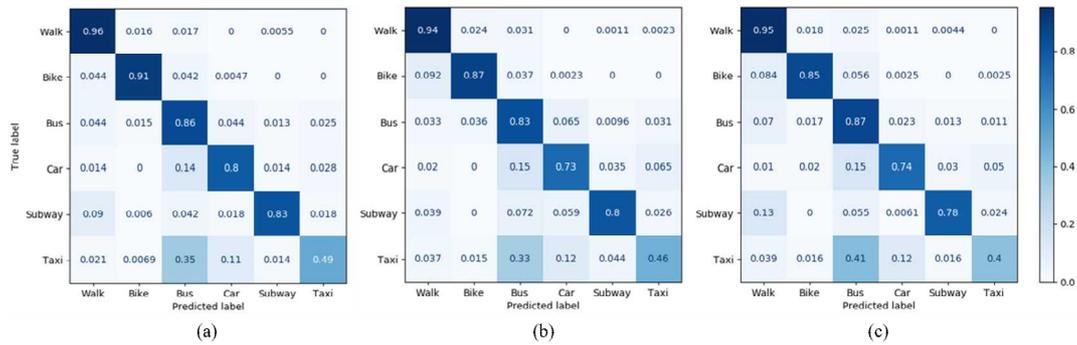


그림 3. 이동수단 분류 혼동 행렬 (a) 제안 방법 (b) Zheng 제안 방법 (c) Erdelic 제안 방법.  
 Fig. 3. Confusion matrix of mobility mode classification using random forest classifier (a) Proposed method (b) Zheng's method (c) Erdelic's method.

본 논문에서 제안한 방법과 기존 방법의 성능을 비교하기 위하여 전체 데이터의 70%를 학습 데이터, 나머지를 테스트 데이터로 랜덤하게 구분한 후, 30번 반복 실험한 정확도의 평균값을 성능 측정 기준으로 사용하였다. 특징 추출과정에서 이동 방향의 변화 비율, 정지 비율, 속도 변화율에 대한 임계값은 기존 연구 [6]에서 설정된 값을 사용하였으며, 각각의 임계값은 15도, 2.5m/s, 0.7로 설정하였다. 기계학습 방법의 설정값은 scikit-learn 라이브러리[21]의 기본값을 사용하였으며, 랜덤 포레스트에서 기본값 트리의 수는 100, k-근접 이웃에서 k의 기본값은 5로 설정되어 있다.

표 2. 이동수단 분류 실험 결과.  
 Table 2. Results of mobility mode classification.

	DT	k-NN	RF
Proposed method	80.21	83.72	87.72
Zheng's method [6]	77.07	77.06	84.72
Erdelic's method [8]	77.08	80.35	85.41

본 논문에서 제안한 방법과 기존 방법에 대하여 30번 반복 실험한 정확도의 평균값은 표 2와 같다. 이동수단 분류 정확도를 살펴보면 본 논문에서 제안한 이동 시간을 고려한 이동 방향의 변화 비율, 정지 비율, 속도 변화율 등의 특징을 포함하는 제안 방법이 모든 기계학습 방법에서 기존 방법들보다 향상된 결과를 보여주었으며, 특히 랜덤 포레스트를 분류 방법으로 사용했을 때 가장 좋은 성능을 보여주었

다. 랜덤 포레스트 분류기를 기준으로 분류 성능을 살펴보면 제안 방법이 기존 방법보다 2.3%~3% 이상 높은 정확도를 보여주었다.

그림 3은 랜덤 포레스트 분류기를 사용했을 때 제안 방법과 Zheng 방법, Erdelic 방법의 이동수단 분류 혼동 행렬(Confusion matrix)을 나타낸다. 이동수단 분류 혼동 행렬을 살펴보면 제안 방법의 경우 기존 방법보다 버스를 제외한 5개 이동수단범주에서 더 높은 분류 정확도를 보여주었으며, 걷기와 자전거 범주의 경우 90% 이상, 버스와 지하철 및 자동차 범주도 모두 80% 이상의 높은 정확도를 보여주었다. 모든 방법에서 택시 범주의 분류 정확도가 가장 낮게 나타났으며, 이러한 결과는 택시, 자동차, 버스 세 범주의 이동 특성이 유사하기 때문으로 판단된다. 택시 범주의 경우 버스로 잘못 분류되는 결과가 가장 많았으며, 기존 연구 [22]와 같이 버스 노선의 정보를 추가로 활용한다면 전체 분류 정확도가 향상될 수 있을 것으로 기대된다.

#### 4.2 GPS 수집 주기 변화에 따른 이동수단 분류 방법의 정확도 및 에너지 소비량 분석

본 절에서는 GeoLife 데이터셋의 샘플링 주기를 변화시키면서 이동수단 분류 방법의 정확도 및 에너지 소비량 변화를 분석하였다. GeoLife 데이터셋의 GPS 데이터를 샘플링하여 샘플링 주기를 변화시켰으며, GeoLife 데이터셋의 샘플링 주기가 1~5초 사이로 가변적이

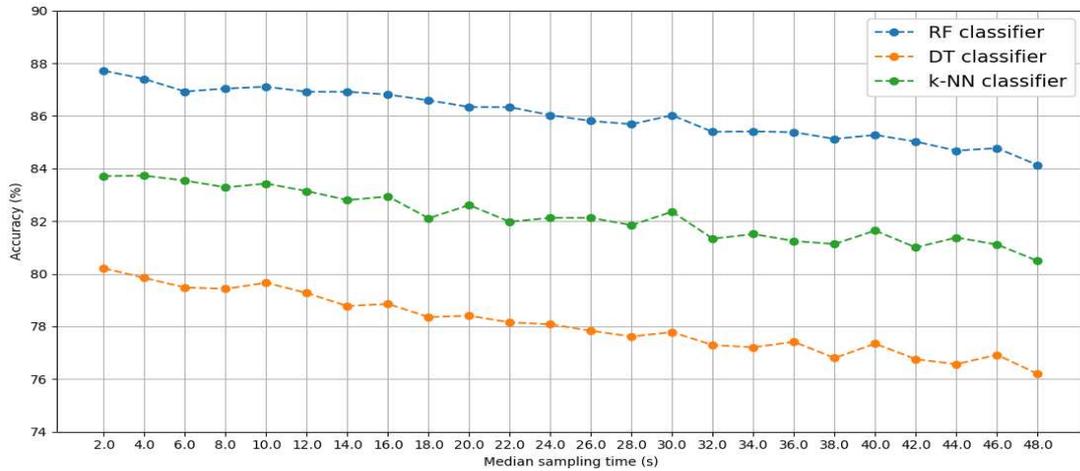


그림 4. GPS 데이터의 샘플링 주기 변화에 따른 이동수단 분류 정확도  
 Fig. 4. Accuracy of mobility mode classification according to the sampling period of the GPS data

어서 샘플링 주기는 전체 데이터의 샘플링 주기 중간값으로 표기하였다.

이동수단의 분류 성능 측정은 이전 실험과 같이 샘플링 주기가 변화된 데이터의 70%를 학습 데이터, 나머지를 테스트 데이터로 랜덤하게 구분한 후, 30번 반복 실험한 정확도의 평균값을 측정하였다. 제안 방법을 사용하여 이동수단을 분류하였으며, 랜덤 포레스트 분류기를 포함한 세 가지 분류 방법에 대한 분류 정확도의 변화를 측정하였다.

GPS 데이터의 샘플링 주기 변화에 따른 이동수단 분류 정확도는 그림 4와 같다. 그림 4를 살펴보면 GPS 데이터의 수집 주기가 길어질수록 이동수단의 분류 정확도가 선형적으로 감소하는 것을 확인할 수 있다. 이는 이동수단 분류에 사용한 알고리즘과 관계없이 동일하게 나타나며, GPS 데이터를 조밀하게 수집할수록 이동수단의 분류 정확도는 향상될 수 있다는 것을 의미한다.

이동수단 분류 방법을 실제 기기에서 활용하기 위해서는 GPS 데이터의 샘플링 주기 변화에 따른 분류 정확도의 변화 분석과 더불어 에너지 소비량의 변화에 대한 분석을 통하여 분류 성능의 감소 없이 최소의 에너지를 소모하는 최적점의 분석이 필요하다. 따라서 본 논문에서는 구글 픽셀 3 스마트폰을 사용하여 GPS 데이터의 샘플링 주기 변화에 따른 에너지 소

비량을 분석하였다. 픽셀 폰의 운영체제는 안드로이드 버전 9.0이며, 배터리 용량은 2915mAh이다. 배터리 소비량은 안드로이드 프레임워크에서 제공하는 BatteryStats[23]을 사용하여 로그를 수집하고, Battery Historian[24]을 사용하여 분석하였다. Battery Historian은 그림 5와 같이 시스템 및 앱의 다양한 리소스 사용량을 시각화해주는 도구로써 시간에 따른 배터리 소비량을 측정할 수 있다.

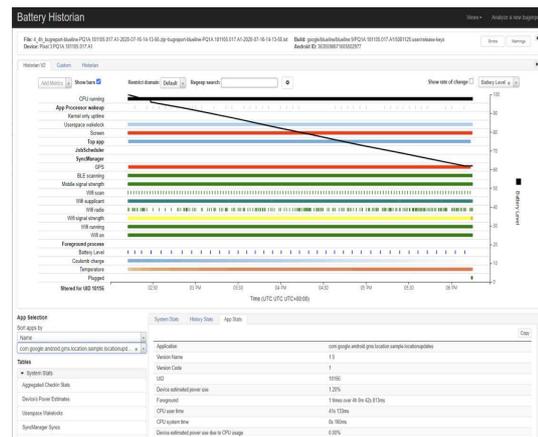


그림 5. Battery Historian을 이용한 에너지 소모 분석 예.  
 Fig. 5. Example of energy consumption analysis using Battery Historian.

표 3. GPS 샘플링 주기에 따른 이동수단 분류 방법의 정확도 및 에너지 소비량.

Table 3. Accuracy and battery consumption of mobility mode classification according to the sampling period of the GPS data.

		Accuracy (%)	Battery capacity (mAh)	Relative battery consumption
Sampling time (s)	2	87.72	99.4	1.0
	4	87.40	95.6	0.96
	6	86.93	82.5	0.83
	8	87.04	78.4	0.79
	10	87.11	76.1	0.77
	12	86.92	75.2	0.76
	24	86.03	42.3	0.43
	36	85.38	41.1	0.41
	48	84.13	39.6	0.40

에너지 소비량 측정을 위한 GPS 데이터 수집 앱은 안드로이드 환경에서 개발되었으며, GPS 데이터의 업데이트 주기는 설정된 샘플링 주기를 사용하여 기록하도록 하였다. 안드로이드는 자체적으로 배터리 사용량 최적화를 위해 설정된 샘플링 주기보다 빠르거나 늦게 GPS 데이터를 업데이트할 수 있어서 가장 빠른 업데이트 간격을 설정된 샘플링 주기의 95%로 설정하여 업데이트 간격을 일정하게 유지하도록 하였다. 에너지 소비량 측정은 같은 위치에서 4시간 동안 GPS 데이터 수집 앱을 실행한 후, 각기 다른 샘플링 주기에 따라 소모되는 배터리 양을 측정하였다.

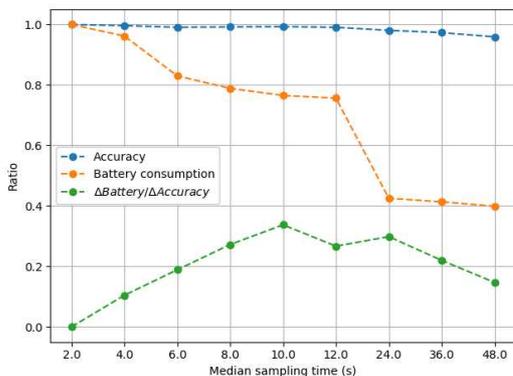


그림 6. GPS 샘플링 주기에 따른 이동수단 분류 방법의 상대적인 정확도 및 에너지 소비량 변화.

Fig. 6. Relative ratio of accuracy and battery consumption of mobility mode classification according to the sampling period of the GPS data.

GPS 수집 주기에 따른 에너지 소비량 및 랜덤 포레스트 기반의 이동수단 분류 정확도는 표 3과 같다. 표 3을 살펴보면 GPS 데이터 수집 주기에 따른 배터리 소비량의 차이는 크게 나타났으며, 수집 주기가 48초 일 때 수집 주기가 2초인 경우보다 60% 정도 배터리 소비량이 감소하였다. 수집 주기가 2초인 경우를 기준으로 GPS 데이터 수집 주기별 이동수단 분류 정확도의 상대적인 감소량과 배터리 소비의 상대적인 감소량을 분석하면 그림 6과 같다. 그림 6을 살펴보면 샘플링 주기가 변화할 때 이동수단 분류 정확도의 상대적인 변화는 크지 않지만, 배터리 소비량의 상대적인 변화는 매우 크게 나타났다. 그림 6에서 이동수단 분류 정확도의 변화율에 따른 배터리 소비 변화율의 비율은 정확도의 변화가 작고 배터리 소비량 감소가 클수록 큰 값을 가지며, 샘플링 주기가 10초일 때 가장 높게 나타났다. GPS 데이터의 샘플링 주기를 10초로 설정하는 경우 이동수단 분류 정확도는 최고 정확도 대비 0.6% 감소하였으며, 에너지 소비량은 33% 이상 감소하였다. 따라서 GPS 데이터의 샘플링 주기를 10초로 설정하면 이동수단 분류 정확도의 손실 없이 에너지 소모를 효과적으로 줄일 수 있는 것을 확인하였다.

## 5. 결론

본 논문에서는 GPS 궤적정보를 사용하여 이동수단을 분류하는 방법을 제안하고, GPS 데이터의 샘플링 주기가 이동수단 분류 방법의 정확도 및 에너지 소비에 미치는 영향을 분석하였다. 본 논문에서 제안한 특징기반 이동수단 분류 방법은 속도 및 가속도 정보에서 통계적 특징을 추출한 기존 방법과 달리 GPS 궤적으로부터 속도, 가속도, 저크 정보를 계산하여 다양한 통계적 특징을 추출하였다. 또한, 본 논문에서는 이동 시간을 고려한 이동 방향의 변화 비율, 정지 비율, 속도 변화율을 이동수단 분류를 위한 특징으로 제안하였다. 본 논문에서 제안한 특징을 다양한 기계학습 알고리즘을 사용

하여 기존의 방법과 비교하였으며, 랜덤 포레스트 분류기를 기준으로 제안 방법이 기존 방법보다 2.3%~3% 이상 높은 정확도를 보여주었다. 또한, GPS 데이터의 샘플링 주기를 변경하여 이동수단 분류 방법의 정확도 및 에너지 소비량 변화를 분석하였다. GPS 데이터의 샘플링 주기가 길어질수록 이동수단 분류 정확도의 변화에 비해 배터리 소비량의 감소가 큰 것으로 나타났다. 실험 결과 GPS 데이터의 샘플링 주기를 10초로 설정하는 경우 이동수단의 분류 정확도는 샘플링 주기를 2초로 설정한 경우보다 0.6% 감소하는 반면 에너지 소모량은 33% 이상 감소하여 이동수단 분류 정확도의 손실 없이 에너지 소모를 효과적으로 줄일 수 있다는 것을 확인하였다.

본 논문에서는 특징기반의 이동수단 분류 방법으로 한정하여 GPS 샘플링 주기에 따른 분류 정확도 및 에너지 소비량을 분석하였다. 향후 연구에서는 심층학습 기반의 이동수단 분류 방법에서의 GPS 샘플링 주기에 따른 분류 정확도 및 에너지 소비량 분석을 진행할 예정이다.

## 감사의 글

본 연구는 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음[20ZS1100, 자율성장형 복합인공지능 원천기술 연구].

## 참고 문헌

[1] X. Yang, K. Stewart, L. Tang, Z. Xie, and Q. Li, "A Review of GPS Trajectories Classification Based on Transportation Mode", *Sensors*, Vol. 18, No. 1, 2018.

[2] Z. Xiao, Y. Wang, K. Fu, and F. Wu, "Identifying Different Transportation Modes from Trajectory Data Using Tree-Based Ensemble Classifiers", *ISPRS Int. J. Geo-Inf*, Vol. 6, No. 57, 2017.

[3] M. Lin and W.-J. Hsu, "Mining GPS

data for mobility patterns: A survey", *Pervasive Mob. Comput*, Vol. 12, pp. 1-16, 2014.

[4] J. Wiest, M. Hoffken, U. Kreßel, and K. Dietmayer, "Probabilistic trajectory prediction with Gaussian mixture models", *Proc. of the IEEE Intelligent Vehicles Symposium*, pp. 141-146, 2012.

[5] L. Ding, H. Fan, and L. Meng, "Understanding taxi driving behaviors from movement data", In *Lecture Notes in Geo-information and Cartography*, pp. 219-234, 2015.

[6] Y. Zheng, Y. Chen, Q. Li, X. Xie, and W.-Y. Ma, "Understanding transportation modes based on GPS data for web applications", *ACM Trans. Web*, Vol. 4, No. 1, 2010.

[7] B. D. Martin, V. Addona, J. Wolfson, G. Adomavicius, and Y. Fan, "Methods for Real-Time Prediction of the Mode of Travel Using Smartphone-Based GPS and Accelerometer Data", *Sensors*, Vol. 17, No. 9, 2017.

[8] M. Erdelic, T. Caric, E. Ivanjko, and N. Jelusic, "Classification of Travel Modes Using Streaming GNSS Data", *Transportation Research Procedia*, Vol. 40, pp. 209-216, 2019.

[9] Y. Endo, H. Toda, K. Nishida, and J. Ikedo, "Classifying spatial trajectories using representation learning", *Int. J. Data Sci. Anal.*, Vol. 2, pp. 107-117, 2016.

[10] S. Dabiri and K. Kevin Heaslip, "Inferring transportation modes from GPS trajectories using a convolutional neural network", *Transportation Research Part C: Emerging Technologies*, Vol. 86, pp. 360-371, 2018.

[11] G. Lee, J. Maeng, and S. Song, "Mobility Mode Classification Method

- 
- for Trajectory Data Using CNN”, *J. of KIIT.*, Vol. 17, No. 12, pp. 13-30, 2019.
- [12] I. Khan, S. Khusro, S. Ali, and J. Ahmad, “Sensors are Power Hungry: An Investigation of Smartphone Sensors Impact on Battery Power from Lifelogging Perspective”, *Bahria University Journal of ICT*, Vol. 9, No. 2, pp. 8-19, 2016.
- [13] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, “Mining interesting locations and travel sequences from GPS trajectories”, *Proc. of Intl. conf. on World Wild Web (WWW 2009)*, pp. 791-800, 2009.
- [14] Y. Yu Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, “Understanding Mobility Based on GPS Data”, *Proc. of ACM conf. on Ubiquitous Computing (UbiComp 2008)*, pp. 312-321, 2008.
- [15] Y. Zheng, X. Xie, and W.-Y. Ma, “GeoLife: A Collaborative Social Networking Service among User, location and trajectory”, *IEEE Data Engineering Bulletin*, Vol. 33, No. 2, pp. 32-40, 2010.
- [16] K. J. Piczak, “Environmental sound classification with convolutional neural networks”, *Proc. of the IEEE 25th International Workshop on Machine Learning for Signal Processing*, pp. 1-6, 2015.
- [17] C. Y. Jeong, H. S. Yang, and K. D. Moon, “A Novel Approach for Detecting the Horizon using a Convolutional Neural Network and Multi-Scale Edge Detection”, *Multidimensional Systems and Signal Processing*, Vol. 30, No. 3, pp. 1187-1203, 2019.
- [18] A. Suleiman, Y. Chen, J. Emer, and V. Sze, “Towards closing the energy gap between HOG and CNN features for embedded vision”, *Proc. 2017 IEEE Intl. Symposium on Circuits and Systems (ISCAS)*, pp. 1-4, 2017.
- [19] C. Y. Jeong and M. Kim, “An Energy-Efficient Method for Human Activity Recognition with Segment-Level Change Detection and Deep Learning”, *Sensors*, Vol. 19, No. 17, 2019.
- [20] T. Vincenty, “Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations”, *Survey Rev.*, Vol. 23, pp. 88-93, 1975.
- [21] F. Pedregosa et. al, “Scikit-learn: Machine Learning in Python”, *J. of Machine Learning Research*, Vol. 12, pp. 2825-2830, 2011.
- [22] D.J. Patterson, L. Liao, D. Fox, and H. Kautz, “Inferring high-level behavior from low-level sensors”, *Proc. of ACM Conf. on Ubiquitous Computing*, pp. 73-89, 2003.
- [23] Batterystats, <https://developer.android.com/topic/performance/power/setup-battery-historian#gather-data>
- [24] Battery Historian, <https://developer.android.com/topic/performance/power/battery-historian>