

신변보호자 경로이탈 감지를 위한 GPS 기반 LSTM 학습 연구

김지형¹, 유재현²¹성신여자대학교 AI 융합학부 학부생²성신여자대학교 AI 융합학부 교수

20211320@sungshin.ac.kr, jhyoo@sungshin.ac.kr

A Study on LSTM Learning for Detecting Anomalous Trajectories of Protected Individuals by using GPS

Jiyoung Kim¹, Jaehyun Yoo²¹School of AI Convergence, Sungshin Women's University²School of AI Convergence, Sungshin Women's University

요 약

본 연구는 LSTM 모델이 수용 가능한 익명 보행자의 GPS 경로 범위와 훈련 데이터 셋의 크기에 대한 양상 분석을 목적으로 한다. 시계열 데이터인 GPS 경로 그리고 순환 신경망 LSTM 과 입력 구조를 이해하고, 두 가지 실험을 설계하여 LSTM 의 훈련 데이터 셋 수용을 파악한다. 실험에서는 장거리 데이터 셋을 학습한 모델과 그렇지 않은 모델을 비교하고, 훈련 데이터 셋 크기에 따른 학습 모델의 예측 값을 비교한다. 두 실험을 통해 GPS 경로 범위와 학습 가능한 경로의 가짓수에 대한 비교 분석 결과를 제시한다.

1. 서론

본 연구는 익명 보행자의 GPS 경로에 대한 오차 값을 비교하여 LSTM 모델이 수용할 수 있는 경로의 범위와 경로의 수효에 대한 양상을 분석한다.

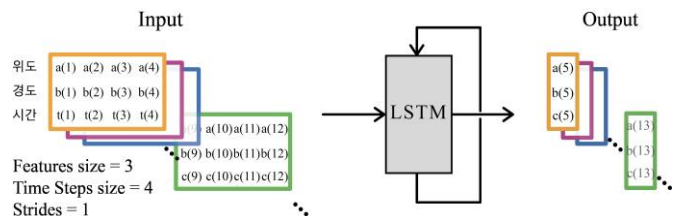
GPS 경로는 시계열 데이터(Time-series data) 즉 시간순으로 관측된 데이터로, 이전 위치와 현재 위치가 다음 위치에 영향을 미친다[1]. 이러한 특성을 가진 GPS 데이터를 고려하여 순환 신경망의 일종인 LSTM 을 통해 경로 예측에 대한 양상을 검토하고자 한다.

본 논문은 2 장에서 LSTM 및 GPS 입력 데이터 구조에 관해 서술하고, 3 장에서는 실험 결과를 분석한다. 마지막 4 장에서는 결론 및 향후 연구를 제시한다.

2. LSTM 및 GPS 입력 데이터

LSTM(Long Short-Term Memory, 장단기 메모리)은 기존의 RNN(Recurrent Neural Network, 순환 신경망)이 가진 장기 의존성 문제를 보완한 모델이다[2]. 특정 경로에 대한 시작 위치를 오래 기억해야 하는 데이터의 특성상, GPS 경로 데이터는 LSTM 과 같은 모델을 통해 분석하는 것이 적합하다.

Raw 데이터는 지도 API 를 이용해 수집하였으며, 출발지와 목적지 그리고 출발 시간을 입력하면 특정 시간 간격으로 GPS 경로와 관측 시간이 저장된다. 예를 들어, A 경로는 서울시 내 구청 사이의 경로 7 개를, B 경로는 서울시 내 지하철역 1 번 출구 사이의 경로 7 개를 포함한다. Raw 데이터는 (위도, 경도, 시간) 3 개의 컬럼을 가지고 있으며, 1 분 간격으로 GPS 값이 저장된 2 차원 배열이다.



(그림 1) LSTM 및 입력 데이터 구조.

기본적으로 LSTM 은 3 차원 형식의 입력 구조를 요구한다. 따라서 입력 데이터는 [Samples size, Time steps size, Features size]와 같이 3 차원 배열로 구성되어야 한다. Samples size 는 데이터의 전체 길이이며, Time steps size 는 LSTM 층에 들어가는 데이터의 개수이다

[1]. 여기서 Features size 는 데이터 셋의 컬럼 수와 같다.

3. 실험 결과

먼저 LSTM 모델이 학습할 수 있는 GPS 경로의 범위를 알기 위해 앞서 수집한 경로를 단거리, 중거리, 장거리로 분류한다. 분류한 경로 중 이동 범위가 반경 35km 이상인 7 개의 경로를 통틀어 Giant 경로라고 정의한다.

Giant 경로를 제외한 훈련 데이터 셋과 해당 경로를 포함한 데이터 셋을 통해 두 개의 모델을 만들어 오차 값을 비교한다. 두 데이터 셋은 150 가지의 동일한 경로 수의 데이터를 가지며, Samples size 는 서로 동일하지 않다.

<표 1> Giant Dataset 학습 모델의 테스트 오차 거리 비교

Test Dataset	Giant 경로 포함	
	분산 (km ²)	Giant 경로 제외
C 경로	31.8	22.0
K 경로	9.3	3.7
M 경로	48.8	35.8
P 경로	28.9	19.5

테스트 데이터 셋인 C, K, P, M 경로는 단거리 및 중거리 경로를 가진 데이터로, 각 데이터 셋은 7 가지의 경로를 가지고 있으며, 총 28 가지의 경로를 테스트한다. <표 1>는 각 경로에 대한 테스트 결과이다. 이를 통해 장거리 경로를 제외한 학습 모델은 단거리 및 중거리에 대해 오차 값이 비교적 낮은 결과를 보여준다.

두 번째, LSTM 모델의 수용 가능한 훈련 데이터 셋의 크기를 분석하기 위해 훈련 데이터 셋 크기 별 학습 모델의 오차 값을 비교한다. 데이터 셋의 크기는 25 가지, 50 가지, 100 가지, 150 가지, 200 가지 경로까지 총 5 개의 크기로 나눈다.

테스트 데이터 셋 C, K, P, M 경로에 대해서 5 개의 학습 모델의 오차 거리를 비교한다.

<표 2> 훈련 데이터 셋 크기 별 테스트 예측 결과

Test Dataset	Train Dataset size				
	25	50	100	150	200
	분산 (km ²)				
C 경로	7.9	5.69	1.416	0.12	1.46
K 경로	39.1	14.1	3.66	0.39	1.46
M 경로	63.8	81.7	44.3	17.1	11.13
P 경로	12.4	41.4	7.0	0.16	1.51

<표 2>에서 훈련 데이터 셋 크기 별 예측 결과를 보여준다. 25 경로를 학습한 모델은 다른 학습 모델에 비해 오차 값이 가장 크다. 즉 LSTM 은 25 가지 보다 큰 훈련 데이터 셋을 필요로 한다. 하지만 200 경로 학습 모델은 150 경로 학습 모델보다 성능이 오히려 떨어진다는 것을 알 수 있다. 이는 LSTM 모델이 150 가지보다 많은 데이터 셋을 수용하지 못함을 의미한다.

훈련 데이터 셋 최대 크기는 모델의 구조와 Optimizer, Unit size 에 따라 학습 데이터 셋의 크기가 달라질 수 있다.

4. 결론

본 연구에서는 LSTM 모델을 통해 익명 보행자의 최대 GPS 경로 범위와 훈련 데이터 셋의 최대 크기에 대해 분석하였다. 이동 범위가 넓은 Giant 데이터를 훈련 데이터 셋에 포함한 모델과 포함하지 않은 모델의 예측 오차 값을 비교한 결과, Giant 를 포함한 모델은 단거리 및 중거리에 대하여 오차 값이 더 크다는 것을 알아냈다. 또한 훈련 데이터 셋 크기 별로 다르게 학습한 모델의 예측 결과를 비교하여 훈련 데이터 셋 크기의 한계를 확인하였다.

향후에는 특정 범위 내 익명 보행자의 경로에 대해 오차 거리를 줄이고, 더 나아가 보행자의 이상 경로를 탐지하는 연구가 진행될 것이다.

ACKNOWLEDGEMENT

이 논문은 경찰청이 지원한 '사회적 약자 보호 강화 기술 개발(www.kipot.or.kr)'의 지원을 받아 수행된 연구결과입니다. [과제명: 저전력 복합측위, 근접탐색 기술 기반 범죄피해 안전조치 대상자 위치추적 통합관제 플랫폼 개발 / 과제번호: RS-2023-00236101]

참고문헌

- [1] Yoon, Seungwon, et al. "Pedestrian GPS Trajectory Prediction Deep Learning Model and Method" *Journal of The Korea Society of Computer and Information* Vol. 27 No. 8, pp. 61-68, 2022.
- [2] Kim, Taehoon, et al. "Flight data prediction method using LSTM based-deep learning model" *Fall Conference of KSAS*, 2021, pp. 968-969.
- [3] Hwang, Chulhyun and Shin, Kwangwook, "CNN-LSTM Combination Method for Improving Particular Matter Contamination (PM2.5) Prediction Accuracy" *Journal of the Korea Institute of Information and Communication Engineering* Vol. 24, No. 1, pp. 57-64, 2020.