

실제 위치 데이터를 기반으로 실시간으로 근미래 위치를 예측하는 시스템

송하윤*, 안상우*

*홍익대학교 컴퓨터공학과

E-mail: hayoon@hongik.ac.kr

E-mail: wolf980909@gmail.com

Next location prediction system with history of geodata

Ha Yoon Song*, Sang Woo An*

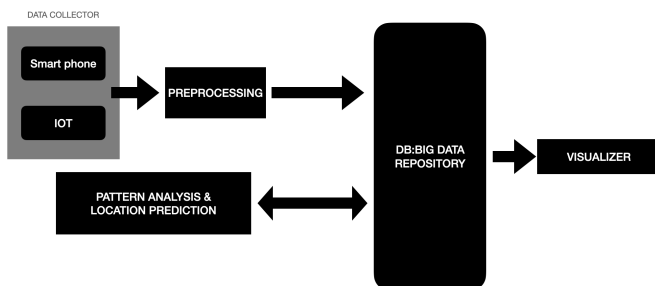
*Dept. of Computer Engineering, Hong-ik University

요 약

소비자의 과거 위치 데이터를 기반으로 다음 경로를 예측하는 것은 마케팅 부분에서 매우 중요한 부분이다. 그러나 전체 데이터를 이용해서 다음 위치를 제공하는 연구는 많았지만 이는 시간이 오래걸리기 때문에 서비스 제공에 이용하기에는 무리가 있다. 그래서 실시간으로 다음 경로를 예측 할 수 있는 서비스를 만들어 보았다. 데이터를 모으는 과정부터 데이터 베이스에 저장하고 활용해 시각화 하는 과정까지 총괄하는 서비스를 만들었다. 이 논문에서는 이동 데이터를 분석해 다음 위치를 예측하는 부분을 다룬다. 이동 데이터를 전처리할때 학습의 편이를 위해 데이터의 형태를 [위도, 경도, 시간] 에서 [라벨값, 시간]으로 바꾸었다. 이 데이터를 CNN을 이용해 학습시킴으로서 실시간으로 예측값을 제공할 수 있다.

1. 서론

본 논문은 사용자들의 과거 이동 경로를 학습해 실시간으로 다음 경로를 예측을 함으로서 다양한 서비스를 제공하는 시스템을 만드는 것을 목표로한다. 전체적인 시스템을 설명해보겠다.



(그림 1)전체 플로우

(그림 1)에서 볼 수 있듯이 우선 스마트폰이나 IoT 기기를 통해 10명의 데이터를 수집하였다. (그림 2)는 데이터 베이스에 넣은 수집한 데이터중 일부이다. 수집한 데이터들은 총 9개의 속성을 가지고 있지만 이 논문에서 필요한 데이터는 [위도, 경도, 시간]이기 때문에 3개의 속성만 가져왔다. 그런 후 모델 학습의 편이를 위해 [라벨값, 시간]의 형태로 전처리 하였다. 그 후 전처리한 데이터를 데이터 베이스에 넣는다.

위치 데이터 패턴 분석을 위해 데이터 베이스에서 전처리한 데이터를 꺼내어 사용한다. 그 후 분석한 내용을 통해 다음 위치를 예측하였다. 이때 사용한 모델은 CNN 이었다.

모델을 통해 예측한 위치 데이터를 다시 데이터 베이스에 저장한 후 이 정보들을 통해 시각화를 하였다.

id	id	created_time	modified_time	name	latitude	longitude	time	year	number_id
1	1280	2021-08-08 16:54:28.090286	2021-08-08 16:54:28.090286	08:20:22	37.546705	126.92331	1581191222	2020/03/03	162
2	1281	2021-08-08 16:54:28.090286	2021-08-08 16:54:28.090286	08:20:30	37.546703	126.92337	1581191230	2020/03/03	162
3	1282	2021-08-08 16:54:28.090286	2021-08-08 16:54:28.090286	08:20:39	37.546620	126.92347	1581191239	2020/03/03	162
4	1283	2021-08-08 16:54:28.090286	2021-08-08 16:54:28.090286	08:20:50	37.546903	126.92345	1581191258	2020/03/03	162
5	1284	2021-08-08 16:54:28.090286	2021-08-08 16:54:28.090286	08:20:59	37.546948	126.92349	1581191269	2020/03/03	162
6	1285	2021-08-08 16:54:28.090286	2021-08-08 16:54:28.090286	08:21:09	37.547048	126.92352	1581191269	2020/03/03	162
7	1286	2021-08-08 16:54:28.090286	2021-08-08 16:54:28.090286	08:21:15	37.547112	126.92357	1581191275	2020/03/03	162
8	1287	2021-08-08 16:54:28.090286	2021-08-08 16:54:28.090286	08:21:22	37.547183	126.92358	1581191282	2020/03/03	162
9	1288	2021-08-08 16:54:28.090286	2021-08-08 16:54:28.090286	08:21:30	37.547242	126.92363	1581191290	2020/03/03	162

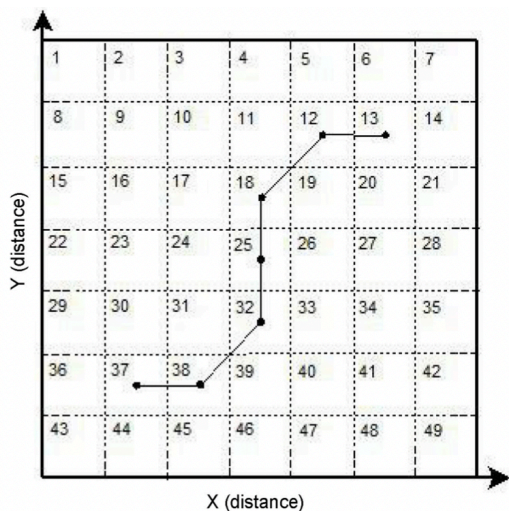
(그림 2)데이터 베이스

제 2절에서는 우리가 사용했던 학습모델인 CNN 알고리즘을 포함해서 이 연구의 프로세스 개요에 대해 논한다.

제 3절에서는 연구에 의한 분석한 최종적인 데이터 셋의 정확도와 세가지 모델 중 더 정확한 모델을 선택을 하였다. 또한 데이터 베이스에 있는 기존 이동 데이터와 학습을 통해 예측한 다음 이동 데이터를 시각화하는 작업을 하였다. 제 4절에서는 결론을 도출하였다.

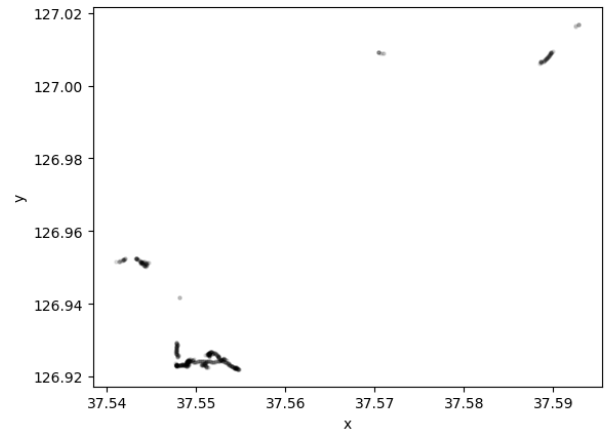
2. 프로세스 개요

첫 번째 프로세스는 10명의 위치 데이터를 수집하는 것이다. 초기 데이터 셋은 [위도, 경도, 시간]로 이루어져 있다. 즉, 위치 데이터는 3차원 데이터라고 할 수 있다. 그러나 분류 알고리즘인 CNN 에서 3차원 데이터를 다룰 때는 2차원보다 계산량이 월등히 많아 진다. 계산의 편이를 위해 (그림 3)과 같이 ‘Area Partitioning and Labeling’을 통해 [라벨값, 시간] 형태로 만들었다. 지역을 분할한 후 지역마다 라벨링을 하여 [위도, 경도]로 알 수 있었던 위치의 좌표를 라벨값으로 표현할 수 있게 되었다. 따라서 경로는 라벨값의 변화로 나타낼 수 있게 되었다. 지역을 나눌 때는 데이터 값의 위도, 경도 값의 최대, 최소값을 구한 후 각자 100등분을 했다.



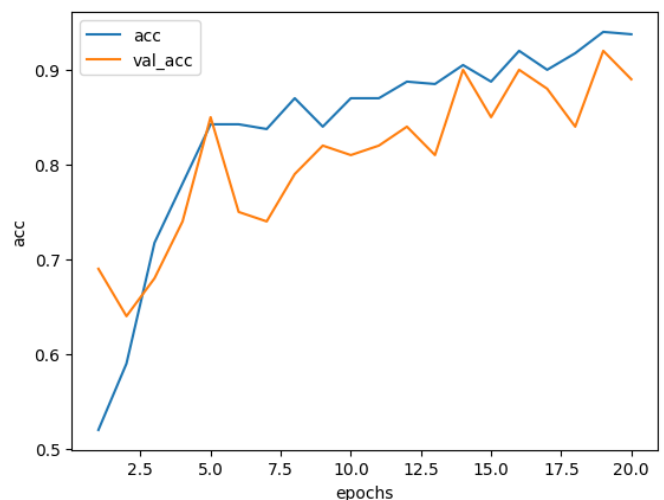
(그림 3)Area Partitioning and Labeling

두 번째 프로세스는 각 사람의 이동경로에 대해서 상세한 실제 경로 데이터를 추출한다. (그림 4)는 사람의 이동 경로 데이터를 위경도를 이용하여 위치 데이터로 출력한 그림이다. 점의 색상이 진할수록, 중복 방문 횟수가 많음을 의미한다.

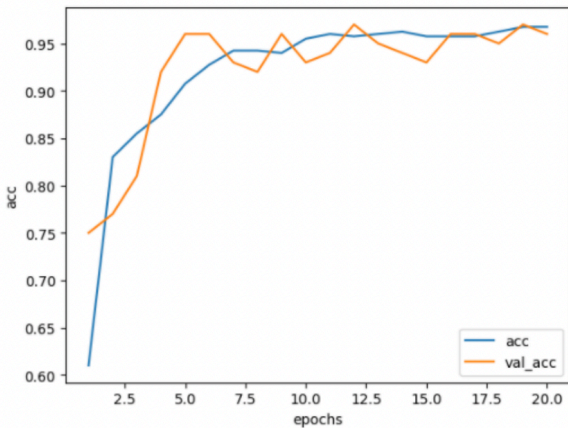


(그림4)위도 경도 이동 데이터 경로

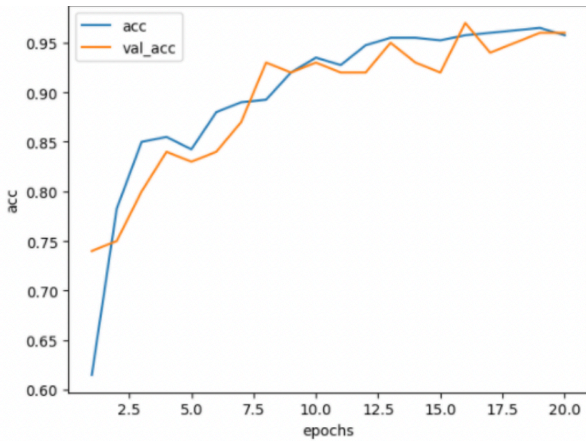
세 번째 프로세스는, 수집한 이동경로에 대해 CNN 알고리즘을 적용한다. CNN 이란, 이미지를 분석하기 위해 패턴을 찾는 데 유용한 알고리즘이고 데이터에서 이미지를 직접 학습하고 패턴을 사용해 이미지를 분류한다. 이를 발전시켜 위치 데이터를 하나의 그림으로 취급하여 학습시켰다. 이때 데이터들은 랜덤으로 8:2로 훈련 데이터, 테스트 데이터로 나뉘어서 전처리 하였다. 총 3가지 CNN 모델을 사용하여 훈련을 시켰다. <표 1>에서는 각 CNN 모델의 구성과 테스트의 정확도를 표현하였다. (그림 5)에서는 훈련의 정확도를 나타내었다. 전반적으로 훈련이 진행될수록 정확도가 높아진다는 것을 알 수 있다.



(그림5-1) 모델 1의 훈련 정확도



(그림 5-2) 모델 2의 학습 정확도



(그림 5-3) 모델 3의 학습 정확도

네 번째 프로세스는, 훈련된 모델을 통해 실시간으로 다음 위치의 라벨값을 추출해내는 과정이다. (그림 6)는 모델 1의 주어진 경로의 다음 값의 실제 값과 예측값을 표현한 것이다. 겹쳐져 있는 곳은 정확하게 예측한 것이고 그렇지 않은 곳은 예측에 실패한 것이다. 테스트를 하기 위해 100개의 경로를 테스트 해 보았다. 모델1은 89%, 모델 2는 97%, 모델 3은 96%의 정확도를 보였다.

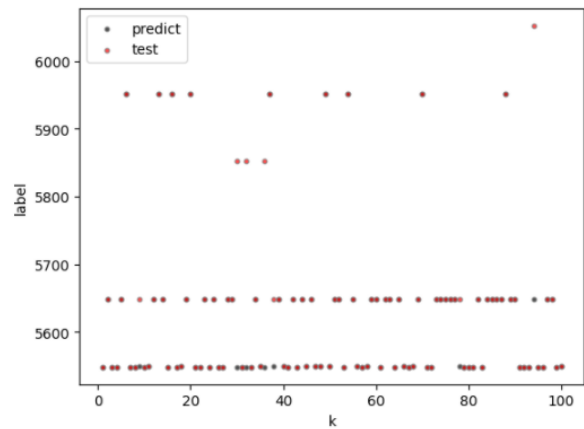
model	architecture	accuracy
CNN1	Cov(5x5)-MaxPool(2v2)-Cov(5x5)-MaxPool(2v2)-dropout(0.5)-FFN	0.89
CNN2	Cov(5x5)-Cov(5x5)-MaxPool(2v2)-dropout(0.5)-FFN	0.97
CNN3	Cov(5x5)-MaxPool(2v2)-MaxPool(2v2)-dropout(0.5)-FFN	0.96

<표1>모델의 구성

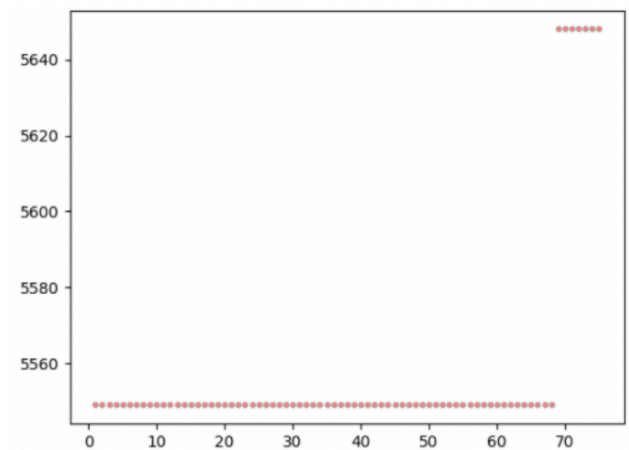
다섯 번째 프로세스는, 모델을 통해 추출된 예상 위치 데이터를 다시 데이터 베이스에 넣는 과정이다. 이는 나중에 시각화 할때 쓰인다.

여섯 번째 프로세스는, 데이터 베이스에 있는 입력 경로와 예측 위치를 꺼내어 시각화를 하는 과정이다. (그림

7)은 테스트 데이터 셋에 있던 데이터를 시각화 한 것이다. 가장 오른쪽에 있는 점이 모델이 예측한 위치이다. 그 앞에 있는 점들은 입력 경로이다.



(그림 6) 모델 1의 예측값과 테스트값



(그림 7) 기존경로 학습을 통한 예상 경로

3. 실험 결과

이 논문에서는 분류 모델을 통해 사람의 위치 예측 시스템을 만들었다. 딥러닝에서 분류 모델인 CNN을 통해 대상의 이동 패턴을 분석하고 다음 위치를 예측하였다.

세 가지 모델에서 Epoch 를 증가할 수록 학습의 정확도가 높아지는 것을 알 수 있었다. 모델 2는 모델 1에서 MaxPool 층을 하나 제거한 모델이고, 모델 3은 모델 1에서 Cov 층을 하나 제거한 모델이다.

(그림 5-1,2,3)은 각 모델들의 훈련 정확도를 나타낸 그래프이다. 20번째 Epoch의 정확도로 따질때 3가지 모델 중 2번째 모델의 정확도가 가장 높은 것을 알 수 있었다.

(그림 6)는 모델 1의 예측값과 테스트 값을 비교해 놓은 점포도이다. 총 100개의 테스트 데이터 세트를 사용하였다. 각 점들은 훈련된 모델에 100개의 경로를 넣었을 때 다음 경로의 라벨값을 예측해 산출해 낸 것이다. 표의

x축은 경로 데이터의 순서를 나타낸 것이고 y축은 라벨값을 나타낸 것이다. 검정색 점은 모델을 통해 예측한 라벨값이고 빨간색 점은 실제 라벨값이다. 검정과 빨간점이 겹친 경우 정확하게 예측한 것이고 그렇지 않을 경우 잘못 예측한 것이다. 예측 값의 정확도는 <표1>에서 볼 수 있듯이 모델 2의 경우가 97%로 가장 높았다.

CNN 모델을 통해 다음 위치를 예측 할때 상당히 정확도가 높다는 것을 알 수 있다. 또한 이미 훈련되어있는 모델에 실시간으로 데이터를 넣으면 바로 예상 위치를 추출해 낼 수 있다.

(그림 7)은 테스트 데이터를 학습된 모델에 넣었을 때 얻어낸 예상 위치를 데이터베이스에서 꺼내서 경로를 표현한 것이다. 그래프의 y축은 라벨값을 나타내고 x축은 사용자의 경로의 길이이다. 기존 경로의 길이는 75 이고 모델을 통한 라벨값을 붙여서 총 경로의 길이는 76이 되었다. (그림7)을 보면 0~69까지는 5550 라벨 지역에 있었다는 것을 알 수 있고 70~75까지는 5650 라벨 지역에 있었다는 것을 알 수 있다. 모델의 예측값은 76 번째에는 5650 라벨 지역을 예측했다는 것을 알 수 있다.

4. 결론

이로써 위치 데이터를 입력 받는 부분부터 예측 경로를 시각화하는 것까지 전체적인 시스템을 만들었다. IoT와 스마트 폰을 이용해서 실제 위치 데이터를 수집을 하였고 이를 전처리 하여 데이터 베이스에 저장하였다. 전처리 된 데이터는 모델 훈련용 데이터와 테스트 용 데이터로 나뉘었다. 훈련된 모델에서 추출한 예상 위치 데이터를 다시 데이터 베이스에 저장한 후 기존 경로와 합하여 시각화를 하였다. 이러한 시스템을 통해 실시간으로 다음 위치의 예측값을 얻을 수 있게 되었다. 이러한 시스템을 통해 실시간 예상 위치 제공 서비스를 할 수 있다. 이러한 서비스는 스포츠 부분이나 배달, 영업 부분에서 많은 도움이 될 것이다. 사용자의 과거 이동 경로를 통해 높은 정도의 확률로 미래 위치를 알 수 있다면 불필요한 비용들을 상당히 줄일 수 있을 것이다.

Acknowledgement

이 연구는 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 (NRF-2019R1F1A1056123)

참고문헌

- [1] Dabin You, Ha Yoon Song, "Trajectory Pattern Construction and Next Location Prediction of Individual Human Mobility with Deep Learning Models." *Journal of Computing Science and Engineering* 14.2(2020):52-65.
- [2] Scarselli, F., Gori, M., Tsoi, A., Hagenbuchner, M. & Monfardini, G. 2009, 'The graph neural network model', *IEEE Transactions on Neural Networks*, vol. 20, no. 1, pp. 61-80.
- [3] W. Shi, G. Pallis and Z. Xu, "Edge Computing [Scanning the Issue]," in *Proceedings of the IEEE*, vol. 107, no. 8, pp. 1474-1481, Aug. 2019, doi: 10.1109/JPROC.2019.2928287.
- [4] Dong Yup Kim, Ha Yoon Song, Method of predicting human mobility patterns using deep learning, *ScienceDirect ,Neurocomputing*, Pages 56-64, 2017