딥러닝 기반의 도시 지역 차량궤적 예측 알고리즘 개발 연구

최성진¹ · 김지원² · 유화평¹ · 가동호¹ · 여화수^{3*}

¹한국과학기술원 건설 및 환경공학과 박사과정, ²퀸즈랜드대학교 토목공학과 교수,

³한국과학기술원 건설 및 환경공학과 교수

Deep-learning Based Urban Vehicle Trajectory Prediction

CHOI, Seongjin¹ · KIM, Jiwon² · YU, Hwapyeong¹ · KA, Dongho¹ · YEO, Hwasoo^{3*}

¹Ph.D Course, Department of Civil and Environmental Engineering, KAIST, Daejeon 34141, Korea

²Professor, School of Civil Engineering, The University of Queensland, St. Lucia 4072, Australia

³Professor, Department of Civil and Environmental Engineering, KAIST, Daejeon 34141, Korea

*Corresponding author: hwasoo@kaist.ac.kr

Abstract

Recently, as a variety of position sensors are developed, a large amount of urban position data is collected in the urban traffic networks. Based on the data collected through such location sensors, high-resolution urban mobility data of individual users using urban road networks is generated and collected in the transportation systems. Urban mobility data generated by these sensors provide a novel spatio-temporal insights into the mobility patterns of traffic network users and can be used to develop models and strategies to predict traffic flows in urban areas and improve traffic efficiency. This study proposes an algorithm for predicting urban mobility patterns. Deep learning based algorithm is used to train mobility patterns in urban areas and predict mobility. The proposed algorithm is trained and tested using Bluetooth data collected in Brisbane for one year. As a result of evaluating the performance of the algorithm with the test dataset, the proposed algorithm shows an average prediction accuracy of 70% or more.

Keywords: artificial neural network, deep learning, urban mobility prediction, vehicle trajectory

초록

최근 다양한 위치추적센서를 통하여 수집된 데이터를 기반으로 교통 분야에서는 도시 도로망을 이용하는 개별 사용자의 고해상도 이동성 데이터가 생성 및 수집되고 있다. 해당 센서에서 생성된 도시 지역 이동성 데이터는 교통 네트워크 이용자들의 이동 패턴에 대한 시공간적인 새로운 통찰력을 제공하며, 이는 도시 지역 교통 흐름을 예측하고 교통 효율을 향상시키는 모델 및 전략을 개발하는데 사용될 수 있다. 따라서 이 연구는 도시 지역 이동성 패턴을 예측하고자 도시 지역 차량 궤적을 예측하는 알고리즘을 제안한다. 도시 지역을 구역으로 나누어 거시적 이동성 패턴을 분석한 선행 연구와는 달리, 본 연구에서는 교차로 단위의 차량 궤적 데이터를 생성하여 보다 미시적인 이동성 패턴을 분석하려고 한다. 본 연구에서는 딥러닝 기반의 모델을 사용하여 차량 궤적을 예측하였다. 한 차량이 앞서 진행한 교차로 시퀀스를 입력하여 다음에 이 차량이 진행할 교차로를 예측한다. 제안된 알고리즘은 브리즈번에서 1년간 수집된 블루투스 데

J. Korean Soc. Transp.
Vol.37, No.5, pp.422-429, October 2019
<https://doi.org/10.7470/jkst.2019.37.5.422>

pISSN : 1229-1366
eISSN : 2234-4217

ARTICLE HISTORY

Received: 28 June 2019

Revised: 30 July 2019

Accepted: 3 September 2019

Copyright ©
Korean Society of Transportation

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

이터를 이용하여 학습하고 시험한다. 시험 데이터 세트로 알고리즘의 성능을 평가한 결과 제안된 알고리즘이 평균 70% 이상의 예측 정확도를 보였다.

주요어: 도시 지역 이동성 패턴, 딥러닝, 빅 데이터, 차량 궤적

서론

최근 다양한 위치 추적 센서와 통신기술이 발달함에 따라 도시 공간에서 다양한 객체의 대량의 위치 데이터를 수집하고 있다. 이러한 위치추적센서를 통하여 수집된 데이터를 기반으로 교통 분야에서는 도시 도로망을 이용하는 개별 사용자의 고해상도 Urban Mobility 데이터가 생성 및 수집되고 있다. 개별 사용자의 Urban Mobility 데이터를 수집하는 방법은 크게 두 가지로 분류할 수 있다. 첫 번째 방법은 스마트폰과 같이 GPS (Global Positioning System)가 장착된 위치 인식 장치를 사용하는 것이다. 이러한 위치 인식 장치들은 사용자의 위치를 계속해서 추적하여 기록한다. 두 번째 방법은 도로 변에 설치된 센서를 이용하는 방법이다. 예를 들어, 도로 변에 설치된 블루투스 매체 접근 제어 스캐너(Bluetooth Media Access Control Scanner, BMS)와 같은 센서를 통해서 해당 센서 주변을 지나가는 Bluetooth 장치들의 식별번호인 MAC ID (Media Access Control)와 통과 시간을 기록한다. 이렇게 수집된 데이터 중 같은 이용자 ID 혹은 MAC ID를 가진 데이터들의 위치와 시간 값을 연결함으로써 도시 교통 네트워크에서의 개별 차량 궤적 데이터를 생성할 수 있다.

각각의 개별 차량 궤적 데이터는 개별 차량의 기·종점, 출·도착 시간 그리고 경로 선택 정보를 포함한다. 이외에도 각 지점을 통과한 시간이 기록되기 때문에 통행시간과 평균 통행 속도 등의 정보도 도출할 수 있다. 이와 같은 정보는 도시 교통 네트워크의 이동성 분석을 위한 다양한 정보를 포함하고 있다. 더욱이 일정 기간 동안 해당 도시 교통 네트워크에서 수집된 개별 차량 궤적 데이터를 분석할 하나의 데이터 세트로 결합한다면, 이 데이터 세트는 O-D 행렬, 경로 선택 패턴, 지역 간 교통류 패턴과 같은 해당 도시 교통 네트워크 전반의 이동성 패턴을 나타낼 수 있다(Kim and Mahmassani, 2015).

차량 궤적을 통하여 도시 교통 네트워크의 이동성 패턴을 분석하는 다양한 방법들이 제시되고 있다(Mazimpaka and Timpf, 2016). 도시 교통 네트워크에서 이동하는 차량의 방대한 궤적 데이터를 분석하여 개별 차량의 다음 위치(Noulas et al., 2012; Gambs et al., 2012; Mathew et al., 2012), 목적지(Krumm and Horvitz, 2006; Ziebart et al., 2008; Horvitz and Krumm, 2012; Xue et al., 2015), 그리고 교통체증 및 사고와 같은 교통 사건·사고 발생 여부를(Wang and Kim, 2016) 예측한다. 그 중, 이 연구에서 초점을 둘 방법은 차량 궤적을 기반으로 개별차량의 다음 이동 위치를 예측하는 다음위치 예측문제(next location prediction problem)이다.

선행 연구(Choi et al., 2018)에서 딥러닝 기반의 다음위치 예측문제 방법론이 제시되었다. 이 선행 연구에서는 호주 브리즈번 시의 도시 교통 네트워크를 일정한 반경(300m, 500m, 1,000m)을 기반으로 구역을 분할했고, 차량 궤적데이터를 구역 시퀀스(cell sequence, region sequence) 데이터로 변환했고 딥러닝 방법론의 일종인 회귀신경망(Recurrent Neural Network)을 기반으로 한 모델을 이용하여 차량 궤적을 예측하였다. 원시 차량 궤적데이터를 구역 시퀀스 데이터로 변환하여 원시 차량 궤적 데이터에 존재할 수 있는 미시적 소음(microscopic noise)의 부정적 효과를 제거하고 딥러닝 기반의 모델로 구역 간 O-D 정보와 거시적 교통 흐름을 포함한 구역 단위의 거시적 움직임을 분석하였다.

본 연구에서는 선행 연구와는 달리 미시적 움직임을 분석하고자 한다. 선행 연구에서는 도시 네트워크를 구역으로 나뉘 구역 단위의 차량 궤적을 예측하였는데, 이를 통해서 데이터 기반으로 거시적 교통 흐름을 분석할 수 있었다. 본 연구에서는 선행 연구와는 다르게 도시 네트워크의 교차로 단위의 차량 궤적을 예측하는 알고리즘을 개발하고자 한다. 이를 통해서 교차로 단위의 교통 수요 및 이동 경로를 예측할 수 있고, 각 교차로에서의 개별 차량의 회전 의도(turning intention)를 예측할 수 있다. 또한 교차로에 있는 차량들의 회전 의도 예측값을 종합하여 각 교차로에

서의 실시간 회전 비율(split ratio)을 추정할 수 있다.

결론적으로, 본 연구에서는 도시 네트워크의 교차로 수준의 미시적인 이동성 패턴을 포착하기 위해 차량 궤적 예측 모델을 개발하고자 한다. 기존 연구(Choi et al., 2018)에서와 같이 도시 교통 네트워크를 구역으로 분할하여 차량 궤적을 구역 시퀀스로 나타내는 대신, 본 연구에서는 각 교차로를 관심지점(POI, Point of Interest)으로 설정하여 차량 궤적을 POI 시퀀스로 표현하여 딥러닝 기반의 알고리즘을 이용하여 주어진 POI 시퀀스를 기반으로 다음 POI를 예측하는 알고리즘을 개발한다.

연구방법

1. 도시 지역 차량 궤적 데이터

본 연구에서 사용한 도시 지역 이동성 데이터는 호주 브리즈번 시의 교차로에 설치되어 있는 블루투스 MAC 스캐너(Bluetooth Media Access Control Scanner)를 기반으로 수집된 데이터이다. 이 데이터는 퀸즈랜드 교통부(Queensland Transport and Main Roads, TMR)와 브리즈번 시의회(Brisbane City Council, BCC)에서 제공하고, 브리즈번 시내의 주요 교차로에 설치된 블루투스 MAC 스캐너를 통하여 데이터가 수집된다. 각각의 블루투스 MAC 스캐너는 약 10-15초 간격으로 탐지 범위 안에 있는 블루투스 기기를 감지하고 이를 기록한다(Bhaskar and Edward, 2013). 이와 같이 기록된 데이터 중 같은 MAC ID를 가진 데이터를 시간 순서대로 연결하면 해당 블루투스 장치가 장착된 개별 차량의 이동 궤적을 구성할 수 있다. 이와 같이 수집된 개별 차량의 이동 궤적 데이터는 해당 차량이 실제로 이동한 정확한 기점과 종점은 알 수 없지만, 블루투스 MAC 스캐너가 설치되어 있는 주요 교차로 단위의 기·종점은 알 수 있다. 이러한 문제를 보완하기 위하여 본 연구에서는 비교적 다른 지역보다 블루투스 MAC 스캐너가 많이 설치되어 있는 브리즈번 도심지역을 연구 대상으로 설정하였다.

Figure 1은 호주 브리즈번 시의 블루투스 MAC 스캐너의 공간적 분포를 보여준다. 브리즈번 도심 주변의 149개의 주요 교차로를 관심 지역(Point of Interest, POI)으로 설정하여 POI 시퀀스 데이터를 생성했다. 이 데이터는

Figure 1. Spatial distribution of Bluetooth MAC sensors in Brisbane city

2015년 7월부터 2016년 6월까지 수집되었다. 데이터 수집 과정에서 포함된 데이터 노이즈를 제거하기 위하여 전처리 과정을 거쳤다. 예를 들어, 차량 궤적 상에서 차량의 속도가 너무 높게 책정되는 비현실적인 이동경로나 차량이 이동 중 일부 센서에 감지되지 않아 데이터 포인트가 누락된 경우 데이터 세트에서 제외하여, 총 6만개의 차량 궤적 데이터를 사용하였다.

2. 딥러닝 기반 차량 궤적 예측 알고리즘

본 연구에서는 차량 궤적데이터를 기반으로 생성된 POI 시퀀스를 예측함으로써 도시 지역 차량 궤적을 예측한다. POI 시퀀스를 예측하는 모델은 차량 궤적의 기점(origin), 경로(route), 그리고 종점(destination) 모두를 예측할 수 있는 기능을 가지고 있다.

Figure 2는 딥러닝 기반 차량 궤적 예측 알고리즘의 구조를 보여준다. 본 연구에서는 딥러닝 모델 중 하나인 피드포워드 신경망(Feed-Forward Neural Networks)을 기반으로 하는 차량 궤적 예측 알고리즘을 사용하였다. 먼저 이 알고리즘의 입력층에서는 최대 입력 시퀀스 길이(l) 초매개변수(hyperparameter)로 설정하여 몇 개의 POI가 예측 알고리즘으로 입력될 것인지를 결정한다. Figure 2의 경우에는 l 이 3인 경우이다. Figure 2에서 대상 차량이 A, B, C 순으로 이동하였다고 했을 때, 각각의 POI는 one-hot vector의 형태로 차량 궤적 예측 알고리즘에 입력된다. 이 one-hot vector는 POI 임베딩층(POI embedding layer)를 통하여 특성 벡터(feature vector)로 변환된다. 이 특성 벡터는 각 POI의 위치적 특성들이 반영되어 있다. 이렇게 변환된 특성 벡터들은 피드포워드 신경망 층으로 입력되어 계산된다. 피드포워드 신경망 층은 은닉신경뉴론개수(h)를 초매개변수로 가지고 있다. 피드포워드 신경망 층은 총 3개의 층으로 이루어져 있으며, 각각의 층의 계산이 끝날 때마다 Relu라는 활성화함수를 통하여 계산 결과가 활성화된다. 마지막으로 출력층의 결과값은 다중 분류 문제에서 많이 사용되는 소프트맥스(softmax)라는 활성화함수를 통하여 활성화되고 이 값은 전체 POI 중 다음에 해당 차량이 각각의 POI를 지나갈 확률을 나타낸다. 이 중 실제 지나간 POI를 예측할 확률을 예측 정확도($Acc(Y, \hat{Y})$)로 정의하였다.

Figure 2. Structure of deep-learning based urban mobility prediction model

딥러닝 기반 차량 궤적 예측 알고리즘에서 도출한 POI 확률과 실제 결과값을 비교하여 손실값(loss)을 계산하고, 이것을 기반으로 역전파 알고리즘(back propagation algorithm)을 사용하여 각 층에 존재하는 매개변수들(parameter) 학습한다. 본 연구에서는 손실 함수로는 크로스엔트로피 손실 함수(cross-entropy loss function)를 사용하고, 산정식은 Equation 1과 같다.

$$L(Y, \hat{Y}) = -\frac{1}{N} \sum_n (Y_n \log \hat{Y}_n + (1 - Y_n) \log(1 - \hat{Y}_n)) \quad (1)$$

역전파 알고리즘과 매개변수 학습에 필요한 초매개변수 튜닝 기능을 같이 수행해주는 Adam optimizer (Kingma and Ba, 2014)를 사용하여 매개변수들을 학습한다.

딥러닝 기반 차량 궤적 예측 알고리즘 성능 평가

딥러닝 기반 차량 궤적 예측 알고리즘의 성능을 평가하기 위해 앞서 딥러닝 기반 차량 궤적 예측 알고리즘에 들어가는 초매개변수를 결정해야 한다. 알고리즘의 초매개변수는 데이터와 대상지에 따라 달라지기 때문에 다른 데이터와 다른 대상지에 대해서 같은 알고리즘을 사용한다면 초매개변수를 결정하는 과정을 다시 거쳐야 한다. 초매개변수를 결정하기 위하여 본 연구에서는 각 초매개변수의 후보군을 만들어두고 모든 조합을 전부 실험해보는 완전탐색법(exhaustive search)을 사용하였다. 이 알고리즘에 들어가는 초매개변수는 최대 입력 시퀀스 길이(l)와 은닉 신경 뉴론 개수(h)가 있는데, 최대 입력 시퀀스 길이는 1부터 6까지로 설정하였고, 은닉 신경 뉴론 개수는 1부터 1024까지 2의 거듭제곱수들로(총 11개) 구성하였다. 초매개변수의 조합은 총 66가지가 나오고 각 초매개변수의 조합을 서로 다른 난수생성 시드를 사용하여 100번씩 시험하였다.

Figure 3은 다양한 초매개변수 조합에 대한 결과를 나타낸다. Figure 3의 x축은 은닉 신경 뉴론 개수(h)이고 y축은 왼쪽 그래프의 경우 손실값(loss; $L(Y, \hat{Y})$), 그리고 오른쪽 그래프의 경우 예측정확도(accuracy)를 나타낸다. 각각의 그래프는 최대 입력 시퀀스 길이(l)에 따라 6개의 그래프로 나누어져 있다. 그래프 상의 각 점들은 각 초매개변수 조합에 대한 100번의 시험 결과의 평균값을 나타낸다.

Figure 3. Result of the model performance evaluation. Average loss and average accuracy calculated with training and validation dataset

은닉 신경 뉴론 개수(h)가 증가함에 따라 손실값은 줄어들고 예측정확도는 증가하는 추세가 있긴 하지만 일정 은닉 신경 뉴론 개수 이상에서는 증가량이 크지 않고 수렴한 것으로 보인다. 또한, 은닉 신경 뉴론 개수가 증가할수록 학습해야 될 매개변수의 개수가 많아지고 그에 따라 계산 속도도 느려진다. 따라서 이 경우에는 은닉 신경 뉴론 개수가 64개 이상일 때 큰 증가추세가 없기 때문에 계산 속도를 고려하여 적정 h 값을 64로 정하였다.

Table 1은 은닉 신경 뉴론 개수를 64로 설정했을 때 최대 입력 시퀀스 길이에 따른 예측 정확도의 결과를 보여준다. 평균 예측 정확도를 보면 최대 입력 시퀀스 길이가 2 이상이면 비슷한 수준의 성능을 보여준다. 하지만 예측 정

확도의 표준편차를 보면 최대 입력 시퀀스 길이가 2일 때보다 그 이후가 작은 것으로 나타난다. 따라서 본 연구에서는 예측 정확도의 표준 편차를 줄이기 위하여 적정 최대 입력 시퀀스 길이를 3으로 설정하였다.

Table 1. Result of the model performance evaluation. Average and standard deviation calculated with validation dataset when $h=64$

Max. length of input sequence	Average of accuracy (%)	Standard deviation of accuracy (%)
1	68.225	0.149
2	76.397	0.244
3	76.756	0.209
4	76.772	0.205
5	76.872	0.191
6	76.809	0.195

Figure 4는 시험 데이터셋을 이용하여 측정된 예측 정확도의 히스토그램과 경험적 누적 분포 함수를 보여준다. 알고리즘의 평균 예측 정확도는 70.69%이다. Figure 4에서 예측 정확도의 분포를 보면 예측 정확도가 50% 이상인 케이스는 전체의 약 75% 정도이고, 예측 정확도가 75% 이상인 케이스는 전체의 약 50% 이다.

Figure 4. Prediction result based on test dataset

결론

본 연구는 도시 지역의 이동성 패턴을 분석하여 개별 차량의 이동 궤적을 예측하는 것을 목표로 하고 있다. 해당 목표를 이루고자 브리즈번 지역의 블루투스 MAC 스캐너로 수집된 도시 지역 차량 궤적데이터를 분석하고 딥러닝 기반의 차량 궤적예측 알고리즘을 제안한다. 본 연구에서 제안된 알고리즘은 브리즈번 블루투스 데이터를 기반으로 성능이 평가되었고 평균 70.69%의 예측 정확도를 보여주었다. 본 연구에서는 차량이 이동한 경로 데이터를 기반으로 다음 POI를 예측하려고 하였지만 차량의 이동 경로에 영향을 주는 요인은 다른 것들이 있을 수 있다. 따라서 추후 연구에는 이러한 요인들도 반영하여 예측하는 알고리즘을 개발하여야 한다.

본 연구에서 제안된 알고리즘은 다양한 분야에 적용 가능하다. 예를 들어, 링크 단위의 차량경로 선택을 예측하거나 운전자의 회전 의도를 예측하는데 적용될 수 있다. 교차로에 접근하는 운전자의 회전 의도를 예측하는 것은 교차로 주변에서 차량 충돌 회피 시스템이나 차량 충돌 경보 시스템 등을 포함한 교차로 지원 시스템(Intersection

Assistance System, IAS)의 필수적인 요소이다. 본 연구에서 제안된 알고리즘을 사용하면 어떤 차량이 이전 교차로를 통과하는 시점에 다음 교차로에서 어떤 움직임을 보일지 예측할 수 있다. 운전자 회전 의도를 예측하는 기존 연구들은(Rodemerk et al., 2015; Phillips et al., 2017; Zyner, 2018) 회전 의도의 예측 시점이 매우 짧다. 예를 들면 Zyner의 논문에서는 회전의도를 2초전에, Rodemark의 논문에서는 회전의도를 10m전에 예측하는데 본 연구의 방법론을 사용하면 이전 교차로를 지났을 때 예측이 가능하다. 그렇기 때문에 기존 연구들의 회전의도 예측 시점이 너무 짧다라는 한계점을 해결할 수 있다.

Funding

This research was supported by the Ministry of Land, Infrastructure, and Transport (MOLIT, KOREA) Project ID: 19TLRP-B146733-02, Project Name: Connected and Automated Public Transport Innovation (National R&D Project).

ORCID

CHOI, Seongjin <http://orcid.org/0000-0001-7140-537X>

KIM, Jiwon <http://orcid.org/0000-0001-6380-3001>

YU, Hwapyeong <http://orcid.org/0000-0002-7007-4887>

KA, Dongho <http://orcid.org/0000-0001-9963-1416>

YEO, Hwasoo <http://orcid.org/0000-0002-2684-0978>

References

- Bhaskar A., Edward C. (2013) Fundamental understanding on the use of Bluetooth scanner as a complementary transport data, *Transportation Research Part C, Emerging Technologies* 37, 42-72.
- Choi S., Yeo H., Kim J. (2018) Network-Wide Vehicle Trajectory Prediction in Urban Traffic Networks using Deep Learning, *Transportation Research Record*, 2672, 173-184.
- Gambs M., Killijian O., del Prado Cortez M. N. (2012), Next place prediction using mobility markov chains, in *Proceedings of the First Workshop on Measurement, Privacy, and Mobility*, ACM, 3.
- Horvitz E., Krumm J. (2012), Some Help on the Way: Opportunistic routing under uncertainty, In *Proceedings of the 2012 ACM conference on Ubiquitous Computing* (371-380), ACM.
- Kim J., Mahmassani H. S. (2015), Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories, *Transportation Research Procedia*, 9, 164-184.
- Kingma D. P., Ba J. (2014), Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*.
- Krumm J., Horvitz E. (2006), Predestination: Inferring destinations from partial trajectories, In *International Conference on Ubiquitous Computing* (243-260), Springer, Berlin, Heidelberg.
- Mathew W., Raposo R., Martins B. (2012), Predicting future locations with hidden Markov models, In *Proceedings of the 2012 ACM conference on ubiquitous computing* (911-918), ACM.
- Mazimpaka J. D., Timpf S. (2016), Trajectory data mining: A review of methods and applications, *Journal of Spatial Information Science*, 2016(13), 61-99.
- Noulas A., Scellato S., Lathia N., Mascolo C. (2012), Mining user mobility features for next place prediction in

- location-based services, In 2012 IEEE 12th international conference on data mining (1038-1043), IEEE.
- Phillips D. J., Wheeler T. A., Kochenderfer M. J. (2017), Generalizable intention prediction of human drivers at intersections, In 2017 IEEE Intelligent Vehicles Symposium (IV)(1665-1670), IEEE.
- Rodemerik C., Winner H., Kastner R. (2015), Predicting the driver's turn intentions at urban intersections using context-based indicators, In 2015 IEEE Intelligent Vehicles Symposium (IV) (964-969), IEEE.
- Wang, G., Kim, J. (2016), The prediction of traffic congestion and incident on urban road networks using naive bayes classifier, In Australasian transport research forum 2016 proceedings, Melbourne, Australia.
- Xue A. Y., Qi J., Xie X., Zhang R., Huang J., Li Y. (2015), Solving the data sparsity problem in destination prediction, The VLDB Journal—The International Journal on Very Large Data Bases, 24(2), 219-243.
- Ziebart B. D., Maas A. L., Dey A. K., Bagnell J. A. (2008), Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior, In Proceedings of the 10th international conference on Ubiquitous computing (322-331), ACM.
- Zyner A. G. (2018), Naturalistic Driver Intention and Path Prediction using Machine Learning.