

딥러닝 분석을 이용한 지하철역 혼잡도 변화의 동적 분석 및 예측 모델에 관한 연구

김규진, 박성환, 최성훈, 김양중*

한국공학대학교 컴퓨터전자공학과

{qjink126, park0219ok, tnqkr9289, *zeroplus}@tukorea.ac.kr

A Study on Dynamic Analysis and Prediction Modeling of Subway Station Congestion Changes Using Deep Learning Analysis

Kyujin Kim, Sunghwan Park, Sunghun Choi, Yangjung Kim*
Tech University of Korea

요약

최근 행정안전부는 민원이 폭주한 서울지하철 9 호선과 김포골드라인에 실시간 지하철역 혼잡도를 제공하는 모델을 제시하였고, 시범운영을 통해 전국 4 개 광역도시권으로 확대될 예정이다. 이를 위해서, 지하철역 혼잡도 변화를 예측하기 위한 딥러닝 기반 모델의 연구와 분석이 필요하며, 시간대별 및 기상 조건을 고려하여 지하철역 혼잡도 변화를 정확하게 예측함으로써, 서울시와 같은 대도시의 유동인구의 실시간 정보제공, 안전관리 및 편의성을 제공하고자 한다. 분석을 위해 대규모 공공 데이터셋을 활용하였으며, TensorFlow 기반의 다층 퍼셉트론(Multi-Layer Perceptron, MLP)과 같은 심층 신경망 구조를 포함하는 모델을 구현해 분석한 결과를 제시한다.

I. 서론

지하철역의 혼잡도는 지하철을 이용하는 도시 유동인구 변화를 위치 및 시간 대역 별로 확인하는 수치로, 대도시의 유동인구의 패턴을 이해함으로써, 공공 안전, 교통 트래픽 관리 및 도시계획에 활용될 뿐 아니라, 지하철 이용자에게 유의한 정보서비스가 될 수 있다. 본 연구는, 시간대와 기상 조건을 고려한 지하철역 혼잡도 변화의 동적 분석 및 예측 모델 개발을 통해 서울시의 유동인구 변화를 보다 정밀하게 이해하고 예측하고자 하며, 이를 위해 딥러닝 기술을 이용해 더욱 정확도를 높이고자 한다. 이를 위해서, 공공데이터포털, 서울 열린데이터광장, 기상자료개방포털 등에서 수집된 대량의 관련 데이터를 기반해 분석하였고, 이 수집된 데이터는 지하철 역사별 승하차 인원, 시간대별 이용객 수, 기상 조건 등이 포함되며, 이를 통해 지하철역 혼잡도 변화뿐만 아니라, 이러한 변화가 서울시 전체의 유동인구 패턴에 미치는 영향을 종합적으로 분석한다. 이를 위해, TensorFlow 와 같은 고급 딥러닝 프레임워크를 사용하여 다층 퍼셉트론(Multi-Layer Perceptron, MLP)을 포함한 복잡한 심층 신경망 모델을 설계하고 학습한 결과를 바탕으로 혼잡도의 변화를 예측 및 분석한다[1].

[표 1] 지하철역별 시간대 및 기상 조건에 따른 승하차 인원 데이터

날짜	시간	역명	호선	승차	하차	기온(°C)	강수량(mm)	풍속(ms)	습도(%)	적설(cm)
2021-01-01	5:00:00	가락시장	3,8	51	31	-9.7		1.9	72	
2021-01-01	5:00:00	가산디지털단지	7	119	48	-9.7		1.9	72	
2021-01-01	5:00:00	강남	2	144	88	-9.7		1.9	72	
2021-01-01	5:00:00	강남구청	7	21	16	-9.7		1.9	72	
2021-01-01	5:00:00	강동	5	129	19	-9.7		1.9	72	
...
2023-10-31	23:00:00	홍제	3	113	591	17.4		1.8	79	
2023-10-31	23:00:00	화곡	5	169	903	17.4		1.8	79	
2023-10-31	23:00:00	화랑대(서울여대입구)	6	58	401	17.4		1.8	79	

본 연구를 통해 제시되는 모델은 기존의 방법들과 차별화된 알고리즘을 통해 정확도를 높임으로써, 도시 정책, 교통혼잡관리 및 행정과 이를 이용하는 이용자들에게 신뢰도가 높은 융합서비스를 제공할 것으로 기대되며, 분석결과를 위해 제안하는 알고리즘을 기반해 구현된 결과를 확인함으로써 연구목적의 정당성을 보이고 이를 통해 응용 서비스 모델을 제안해 향후 연구방향을 제시하고자 한다.

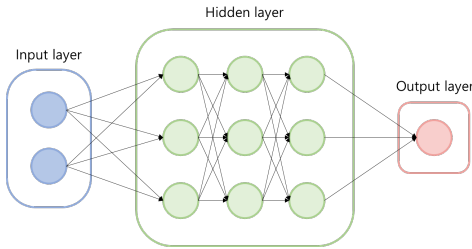
II. 본론

1. 데이터 수집 및 전처리

본 연구에서 사용되는 데이터는 지하철 혼잡도 예측을 위해 공공데이터 포털, 서울 열린데이터 광장, 기상자료개방 포털[2][3][4] 등에서 수집된 대량의 데이터를 기반해, 역사별 승하차 데이터, 시간대별 이용객 수, 날씨 (기온, 강수량, 풍속, 습도, 일조량, 적설)등을 사용한다. 수집된 데이터는 딥러닝 알고리즘에 적합한 형태로 전처리되며, 이 과정에서 불필요한 정보의 제거, 결측치 처리, 특성 추출 및 변환 등이 수행된다. 이는 복잡한 혼잡도 패턴을 더욱 효과적으로 모델링할 수 있는 견고한 데이터셋을 구축하는 데 필수적이다 [5].

2. 딥러닝 모델 구축

본 연구의 목적은 시간대별 지하철 혼잡도를 예측하는 딥러닝 기반 모델을 구축하는 것이다. 이를 위해서, TensorFlow 와 같은 고급 딥러닝 프레임워크를 활용하여, 다층 퍼셉트론(Multi-Layer Perceptron, MLP)을 포함한 복잡한 심층 신경망을 설계한다. 모델은 여러 개의 은닉층과 다양한 형태의 뉴런으로 구성되며, 비선형 활성화 함수, 드롭아웃, 배치 정규화 등의 기법을 사용하여 모델의 성능을 극대화한다. 학습 과정에서는 크로스 엔트로피 손실 함수와 아담 최적화 알고리즘을 사용하여 최적화를 수행하며, 과적합을 방지하기 위한 다양한 전략이 적용된다[6][7].



[그림. 1] DNN(Deep Neural Network)

3. 모델 학습

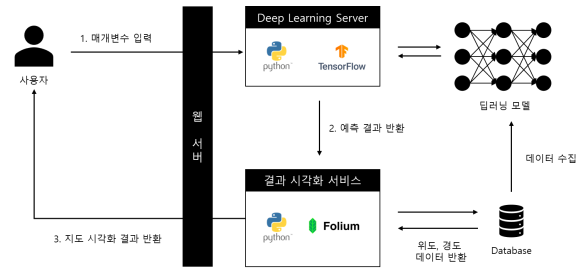
수집된 데이터를 기반으로 모델 학습이 진행되며, 학습 데이터셋과 테스트 데이터셋을 나누어 교차 검증 방법을 통해 모델의 일반화를 도출 및 평가한다. 학습 과정에서는 MSE 와 같은 손실 메트릭을 모니터링하여 모델의 학습 진행 상황을 분석한다. 평가 단계에서는 R², MSE 및 RMSE 와 같은 회귀 성능 지표를 사용하여 모델의 예측 정확도를 판별한다.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - f(x_i; w))^2}$$

이때, y_i 는 i 번째 관측치의 실제 '승하차_인원' 값을 나타내며, $f(x_i; w)$ 는 모델이 해당 관측치에 대해 예측 값이다. RMSE 는 모델의 예측값과 실제값 사이의 오차의 제곱근을 평균값으로, 예측 오차의 크기를 나타내는 수치이며, RMSE 값이 낮을수록 모델의 예측이 실제값에 근접함을 의미하므로, 이로써 모델의 예측력이 우수함을 판별할 수 있다. 본 연구에서는 RMSE 를 통해 모델의 예측 성능을 정량적으로 분석 및 평가하였고, 학습 과정에서 RMSE 값을 지속적으로 모니터링함으로써, 모델의 학습 진행 상황을 실시간으로 평가하고, 더욱 정확한 예측값을 도출하기 위해 도출된 알고리즘을 통해 모델 구조나 학습 매개변수를 추출하여, 모델의 성능을 높였다.

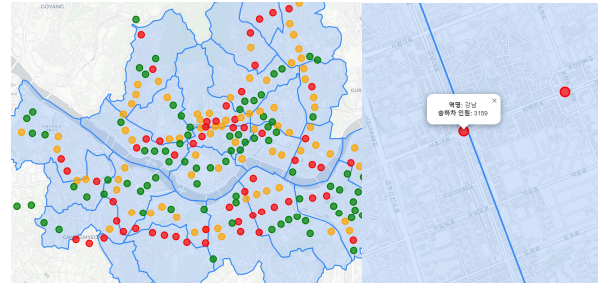
4. 실험 및 분석

본 연구에서 개발된 딥러닝 기반 모델은 서울시 지하철역 중에서도 강남역을 중심으로 한 혼잡도 변화 예측을 실험하고자 한다. 이 모델은 실제 지하철 시스템과의 통합 가능성을 가지며, 지하철 이용자들에게 실시간 혼잡도 정보를 제공하고, 교통 관리자에게는 효과적인 운영 계획 수립에 필요한 예측 정보를 제공할 수 있다.



[그림. 2] 혼잡도 예측 서비스 흐름도

일정기간 내의 수도권 1~8 호선 지하철역 인접지역의 혼잡도를 예측해보는 실험을 진행하였다. 이 실험은 사용자가 입력한 특정 날짜와 시간에 대한 혼잡도 예측을 수행하며, 추가적으로, 해당 시점의 기상 데이터도 함께 고려하여 예측의 정확도를 높이고자 관련 팩터를 예측 알고리즘에 응용하였다. 이는 본 모델이 기존의 다른 딥러닝 모델과 차별화되는 주요 특징 중 하나로, 기상 조건 변화가 혼잡도에 미치는 영향을 모델링함으로써 예측값에 영향도를 분석할 수 있다. 일정기간, 즉, 2021 년부터 2023 년 10 월 31 일까지 수집된 수도권 지하철 1~8 호선 승객의 승·하차 데이터를 기반으로 딥러닝 모델을 수행하고, 데이터셋은 8:2 의 비율로, 80%는 모델 학습에 사용되었고, 나머지 20%는 실험용 데이터로 성능 평가와 예측 결과의 검증하는데 사용된다. 실제 실험은 사용자가 특정 날짜와 시간을 입력하면, 해당 정보와 해당 시간대의 기상 조건 데이터가 딥러닝 서버로 전송되어 처리되는 프로세스를 수행하게 되며, 바로 서버에서는 딥러닝 모델이 이 데이터를 분석하여 혼잡도를 예측하고, 이 예측 결과는 다시 결과 시각화 서비스로 전달되어 사용자가 쉽게 이해할 수 있는 형태로 지도에 표시된다. 최종적으로, 이 시각화된 결과는 웹 페이지 형태로 사용자에게 제공되어 실시간 혼잡도 정보를 확인할 수 있도록 개발되었다.



[그림. 3] 지하철역 인근 혼잡도 예측 시각화

본 연구를 통해 결과적으로, 평균 약 84%의 정확도를 보였으며, 오차율은 약 16% 정도로 보여, 기상 조건의 정확도, 공휴일, 특별한 이벤트 등의 다양한 외부 요인에 의해 발생한 것으로 분석된다.

III. 결론

본 논문은 시간대별 특성 및 기상 조건을 고려한 지하철역 혼잡도 변화를 예측하는 딥러닝 기반 시스템을 개발하였으며 이에 실험한 결과를 제시하였다. 기존 방법과는 차별적인, 다층 퍼셉트론(Multi-Layer Perceptron, MLP) 네트워크와 같은 심화된 신경망 구조를 활용하여 혼잡도 변화를 보다 정밀하게 예측하는 모델을 도출한 것으로써, 유동인구가 많은 대도시의 대중교통을 이용하는 이용자들의 안전과 교통관리 및 편의서비스를 제공할 수 있을 것으로 보이며, 향후 시스템을 안정화 및 고도화를 통해 정확도를 높이고 향상시키는 보강 연구를 진행할 것이다.

ACKNOWLEDGMENT

본 연구는 고용노동부 및 한국산업인력공단의 '2023 년
일학습병행(4 년 대학과정)지원 사업'의 지원을 받음

참 고 문 헌

- [1] TensorFlow Documentation, "Deep Learning with TensorFlow," (<https://www.tensorflow.org/learn>).
- [2] 공공데이터포털(<https://www.data.go.kr/>)
- [3] 서울 열린데이터광장(<https://data.seoul.go.kr/>)
- [4] 기상자료개방포털(<https://data.kma.go.kr/cmmn/main.do>)
- [5] 서울시 지하철 승·하차 데이터 및 시간대별 이용객 데이터 관련 정책 보고서 및 연구 자료.
- [6] Dozat, T. (2016). Incorporating Nesterov Momentum into Adam. ICLR 2016 workshop, San Juan, Puerto Rico.
- [7] Anqi Mao, Mehryar Mohri, Yutao Zhong. (2023). Cross-Entropy Loss Functions: Theoretical Analysis and Applications, v2.