

딥러닝 모형을 이용한 신호교차로 대기행렬길이 예측

Predicting a Queue Length Using a Deep Learning Model at Signalized Intersections

나 다 혁* · 이 상 수** · 조 근 민*** · 김 호 연****

* 주저자 : 아주대학교 교통연구센터 연구원
 ** 교신저자 : 아주대학교 교통공학과 교수
 *** 공저자 : 아주대학교 교통연구센터 연구원
 **** 공저자 : 아주대학교 교통공학과 석사과정

Da-Hyuk Na* · Sang-Soo Lee** · Keun-Min Cho*** · Ho-Yeon Kim****

* Researcher, Ajou Transp. Research Institute
 ** Professor, Dept. of Transportation Eng., Ajou Univ.
 *** Researcher, Ajou Transp. Research Institute
 **** Graduate Student, Dept. of Transportation Eng., Ajou University

† Corresponding author : Lee Sangsoo, sslee@ajou.ac.kr

Vol.20 No.6(2021)

December, 2021
pp.26~36pISSN 1738-0774
eISSN 2384-1729
<https://doi.org/10.12815/kits.2021.20.6.26>Received 10 November 2021
Revised 26 November 2021
Accepted 3 December 2021© 2021. The Korea Institute of
Intelligent Transport Systems. All
rights reserved.

요 약

본 연구는 영상검지기에서 수집되는 정보를 활용하여 딥러닝 기반으로 대기행렬길이를 예측하는 모형을 개발하였다. 그리고 통계적 기법인 다중회귀 모형을 추정하여 평균절대오차와 평균제곱근오차의 두 지표를 이용하여 비교·평가하였다.

다중회귀분석 결과, 시간, 요일, 점유율, 버스 교통량이 유효한 변수로 도출되었고, 이 중에서 독립변수들의 종속변수에 대한 영향력은 점유율이 가장 큰 것으로 나타났다. 딥러닝 최적 모형은 은닉층이 4겹, Look Back이 6으로 결정되었고, 평균절대오차와 평균제곱근오차가 6.34와 8.99로 나타났다. 그리고 두 모형을 평가한 결과, 다중회귀 모형과 딥러닝 모형의 평균절대오차는 각각 13.65와 6.44, 평균제곱근오차는 각각 19.10과 9.11로 계산되었다. 이는 딥러닝 모형이 다중회귀 모형과 비교하여 평균절대오차가 52.8%, 평균제곱근오차는 52.3% 감소된 결과이다.

핵심어 : 다중회귀, 딥러닝, LSTM, 대기행렬길이, 예측

ABSTRACT

In this study, a deep learning model for predicting the queue length was developed using the information collected from the image detector. Then, a multiple regression analysis model, a statistical technique, was derived and compared using two indices of mean absolute error(MAE) and root mean square error(RMSE).

From the results of multiple regression analysis, time, day of the week, occupancy, and bus traffic were found to be statistically significant variables. Occupancy showed the most strong impact on the queue length among the variables. For the optimal deep learning model, 4 hidden layers and 6 lookback were determined, and MAE and RMSE were 6.34 and 8.99. As a result of evaluating the two models, the MAE of the multiple regression model and the deep learning model were 13.65 and 6.44, respectively, and the RMSE were 19.10 and 9.11, respectively. The deep learning model reduced the MAE by 52.8% and the RMSE by 52.3% compared to the multiple regression model.

Key words : Multiple Regression, Deep Learning, LSTM, Queue Length, Forecasting

I. 서론

1. 개요

국내 도시들은 산업화와 인구 집중 현상으로 일상적인 교통 혼잡이 발생하고 있다. 국토교통부에서 추정된 국내 도로 교통 혼잡비용은 2017년 기준 38.7조 원이며 이는 매년 증가하는 추세를 보이고 있다. 도시부 도로의 교통 혼잡은 일반적으로 교차로의 차량 수요가 교차로의 용량을 넘어서 발생한다. 이러한 교통 혼잡을 대변하는 교통 제어 변수 중 하나는 대기행렬길이(Queue Length)이다. 대기행렬길이는 신호 교차로의 운영효율을 평가하는 지표로 사용되며 신호의 연동 운영과 최적화를 위하여 요구되는 정보 중 하나이다. 특히 과포화 교통 상황 발생 시 대기행렬길이는 신호 최적화의 목적 함수로 사용되어 그 중요성이 더욱 드러나는 변수이다. 그리고 대기행렬길이 정보는 과포화 신호 교차로에서 주기길이 결정이나 현시 분할(Split)을 결정하는 과정에도 사용된다. 또한, 교통 정보 측면에서 도로 이용자들에게 대기행렬길이 정보가 제공된다면 이용자들은 링크별 대기행렬 크기를 통하여 해당 구간의 혼잡 수준을 확인할 수 있고, 이러한 구간을 회피하는 경로를 선택하여 차량의 통행시간을 감소할 수 있다.

실제로, 프랑스의 교통 관리 시스템 SIRIUS는 도로 이용자에게 대기행렬길이 정보를 제공하여 운전자의 경로 전환에 영향을 끼쳤으며, 또한 경로상 지체에 관한 예측 정보가 상세할수록 도로 이용자들이 더 민감하게 반응한다는 연구 결과를 제시하였다. 그러나 국내에는 대기행렬길이를 신호제어 목적으로 사용하고 교통정보로는 제공하지 않고 있다. 국내 정보센터들 중에서 대표적인 서울시 교통 정보 시스템(TOPIS)이나 부산광역시 교통 정보 서비스 센터 등은 도로 이용자들에게 통행속도 등의 정보를 제공하지만 대기행렬길이와 관련된 정보는 미제공하고 있다. 국내의 도로 이용자들이 대기행렬 길이 예측 정보를 받게 된다면 경로 선택을 더욱 효율적으로 하게 될 것이며, 더 나아가 교통 정보 서비스의 질적 향상을 이룰 수 있을 것이다.(Sim et al., 2002; Yim and Ygnace, 1996)

대기행렬길이 정보와 관련된 또 다른 문제점으로는 국내 실시간 신호제어시스템(COSMOS) 등에서 현재 사용되는 대기행렬길이 예측 모형의 오차가 크다는 점이다. 전통적으로 대기행렬길이 예측 모형들은 루프 검지기 자료를 사용하여 추정되었으나 이는 이상적인 상황을 가정한 모형으로 다양한 교통 상황들을 고려하지 못하기 때문이다(Lee and Oh, 2002). 따라서 현재의 대기행렬길이 모형은 결과의 신뢰성과 정확도 측면에서 개선될 필요가 있고, 산출된 값은 정보 제공의 용도로 사용되기 어렵다고 판단된다.

본 연구의 목적은 현재 국내 도로에 많이 설치되고 있는 영상검지기에서 수집된 자료를 기반으로 대기행렬길이를 예측하는 딥러닝 모형을 개발하여 제시하는 것이다. 본 연구에서는 영상검지기 자료를 활용하여 신호제어를 위한 대기행렬길이 정보를 생성하는 것보다는 대기행렬길이 정보를 제공하는 측면에 사용되는 모형을 개발하여 그 적용성을 평가하고자 하였다. 따라서 영상검지기는 5분 단위의 자료를 수집하여 사용하였고, 도출된 모형의 결과를 활용하여 도로 이용자들에게 대기행렬길이 정보를 제공함으로써 궁극적으로 교통 혼잡을 완화하는데 기여할 수 있다고 판단된다.

II. 이론적 배경

1. 선행 연구 고찰

Yim and Ygnace(1996)은 프랑스 파리의 SIRIUS 시스템을 대상으로 운전자들이 가변전광판의 실시간 교통 정보에 반응하는 정도와 이로 인한 구간 교통량 변화를 분석하여 운영 효과를 제시하였다. 분석 결과, VMS

를 통한 실시간 대기행렬길이 제공은 운전자가 대체 경로를 선택하도록 영향을 미치는 것으로 나타났다. 또한 혼잡이 심해짐에 따라 운전자는 가변 메시지 표지판에 반응할 가능성이 커졌으며, 대기행렬길이가 3km를 초과하는 시점이 다수의 운전자가 경로를 전환하는 임계점인 것으로 나타났다.

Han et al.(2000)은 도시부 신호 교차로의 대기행렬을 단기예측하기 위한 인공신경망의 학습자료로 시계열적 패턴을 지닌 검지 자료로 구성된 경우와 시공간적 상관관계를 갖는 검지 자료로 구성된 경우를 비교·평가하였다. 각 모형의 평균제곱근오차를 통해 예측력을 비교한 결과, 시계열적 검지 자료를 학습자료로 구성하는 것보다 시공간적 상관관계를 갖는 교통류 변수를 인공신경망의 학습자료로 구성할 때 더욱 뛰어난 예측력을 나타내는 것으로 나타났다.

Ki et al.(2002)은 대기행렬길이를 예측하는 신경망 모형의 단점인 과도한 학습 시간을 극복하기 위해 신경망 모형에 유전자 알고리즘을 결합한 결합모형을 제시하였다. 기존의 신경망 모형과 비교한 결과, 학습 시간이 단축되었고 평균제곱근오차 값과 변화폭이 더욱 적어 예측력이 좋아진 것이 확인되었다. 해당 연구에서는 영상검지기 및 루프 검지기를 통해 수집한 상류 유입교통량, 링크 통행시간, 대기행렬, 유출 교통량 데이터를 기반으로 모형을 구축하고 하루 4시간, 111개의 Data를 활용하였으나, 향후 더욱 많은 자료를 통해 달별, 요일별, 시간대별 자료를 모형의 학습자료로 활용하는 연구가 필요하다고 제시하였다.

Sim et al.(2002)은 Kalman filter를 활용하여 5분 후의 대기행렬 길이를 예측하는 모델을 제안하였으며 실제 검지기 자료를 활용하여 서울 도심 남산권 네트워크에 적용하였다. 최종 정보 제공지점과 정보의 대상지의 평균통행시간이 5분인 것으로 파악되어 5분 단위로 예측 주기를 설정하고, 평균절대오차, 평균제곱근오차 등을 통해 모델을 평가한 결과 일정 수준 이상의 예측력을 보인다고 판단하였다. 또한 시나리오별 이력 데이터 수집 기간에 차이를 두어 비교한 결과, 이력 데이터의 양이 많을수록 더욱 정확한 대기행렬 길이를 예측할 수 있는 것으로 나타났다. 향후 다른 정체 지역에 적용해봄으로써 지역적 특성을 극복할 수 있는지 검토하고 여타 예측기법을 적용하여 비교하는 과정이 필요할 것이라고 밝혔다.

Lee and Oh(2002)는 기존의 신호제어 시스템 COSMOS의 대기행렬길이 산출 알고리즘의 문제점을 보완하고자 이를 보완한 대기행렬 알고리즘을 개발하였다. 기존 시스템과 달리 링크 내 횡단보도, 신호등과 같은 외부적 요인을 고려하여 알고리즘을 구축한 결과, 기존 알고리즘보다 평균절대오차 등이 작아진 결과를 확인하였고, 해당 연구의 대기행렬길이 산출 알고리즘을 다른 현장에 적용하여 개발된 알고리즘의 타당성을 검토하는 과정이 필요하다고 제시하였다. 그러나 해당 알고리즘은 루프 검지기 데이터를 기반으로 운영되며 현재 신호 교차로에 확대 설치되고 있는 영상검지기 데이터를 기반으로 운영될 수 없고, 이에 대한 타당성도 검증되지 않았다.

Kang and Oh(2005)는 정지선으로부터의 거리와 검지기에서 수집되는 점유율 간의 관계를 대기행렬길이 산출 알고리즘에 적용하여 기존의 COSMOS 시스템과 비교·평가하였다. 그 결과, 현재 사용되는 COSMOS 시스템의 대기행렬길이 산출 결과보다 평균절대비율오차 등이 우수하게 나타났다. 또한 개발된 알고리즘은 점유율을 활용하므로 루프 검지기에서의 가공이 필요하지 않고, 현장 적용 시 운영자 지정 값이 필요하지 않아 기존의 산출 과정보다 축소된다는 점이 확인되었다.

Liu et al.(2009)는 현재 신호의 직전 신호 주기에서의 대기행렬 해소 프로세스를 이용하여 현재 주기의 대기행렬 길이를 추정하는 모델을 제안하였다. 해당 연구에서는 교통 상태 변화 구별 알고리즘을 위해 고해상도 교통 데이터와 LWR 충격파 이론을 적용하였고, 이를 통해 대기행렬길이가 정지선과 검지기 간의 길이보다 길 때도 예측할 수 있게 되었다. 그러나, 향후 상류 교차로 데이터를 추가하여 모델의 견고성을 개선할 필요가 있음을 제시하였다.

Fu et al.(2016)는 LSTM 및 GRU 신경망 딥러닝 기법을 기반으로 실시간 교통류 패턴을 예측하는 모델을 각각 구축하였고, 시계열 분석 모형인 ARIMA와 모델의 성능을 비교하였다. 평균절대오차 등을 통해 예측력

을 비교한 결과, GRU, LSTM, ARIMA 순으로 예측 성능이 높은 것으로 드러나 딥러닝 모델의 예측력이 우수함을 증명하였다. 향후 시계열 데이터 길이의 변경이 가능해진다면 모델이 자동으로 지연시간을 최적화하는 데에 도움이 될 것이며, 더 많은 hidden state를 가지는 순환신경망 모델을 검증할 필요성을 제기하였다.

Lee et al.(2018)는 텐서플로우에 VISSIM을 제어하는 COM Interface를 구축하여 링크 통행시간 및 통과교 통량 데이터를 수집하였으며, 이를 통해 심층신경망(Deep Neural Network, DNN)을 기반으로 하는 대기행렬 길이 예측 모델을 구축하였다. 모델을 통해 예측한 대기행렬길이를 신호 주기를 산정하고, 산정된 신호 주기를 VISSIM 네트워크에 적용시킨 결과 현 TOD 방식에 비해 차량 지체가 약 5% 감소한 것으로 분석되었다. 해당 연구에서는 하나의 교차로를 대상으로 적용하였지만, 향후 연구 과제로 축 혹은 네트워크 단위로 공간적 범위를 확대하여 적용할 것을 제안하였다.

Yin et al.(2018)는 고정된 검지기 데이터가 아닌 개별 차량의 모바일 센서 데이터를 이용하여 실시간 대기행렬길이 예측 모델을 개발하였다. 이는 Kalman Filter와 충격과 이론의 조합을 통해 구축되었으며 대기행렬 길이 및 최대 대기행렬길이를 추정하였다. 기존의 선형 회귀분석 기법과 성능을 비교하기 위해 실제 네트워크 및 시뮬레이션 시나리오에서 평균절대비율오차를 통해 분석한 결과, 모바일 센서 데이터를 활용한 모델이 더욱 우수한 것으로 나타났다.

Luo et al.(2019)는 번호판 인식 데이터를 통해 수집한 각 차량의 통행시간과 각 신호 주기의 대기행렬길이의 관계를 통해 대기행렬길이를 예측하는 회귀분석 모델을 제안하였다. 이러한 회귀분석으로 예측된 대기행렬 길이와 실제 대기행렬길이의 관측값과 평균절대오차를 통해 우수한 예측력을 보이는 것으로 판단하였으며, 이는 교통류 혼잡 상태에서도 높은 예측력을 보이는 것으로 확인되었다.

Gao et al.(2020)는 자율 군집 주행 차량의 정보와 DNN 기법을 이용하여 대기행렬길이를 예측하였다. 이는 이미지 데이터 기반의 예측 방법으로서 실시간 정보 제공이 가능하고, 기존의 충격과 이론 기반의 모델과 달리 차량 유형 및 차량 위치를 정확하게 파악하였다. 평균상대오차를 통해 실제 데이터와 모델 예측 결과를 비교한 결과, 약 10%의 값으로 도출되어 신호제어에 활용될 수 있다고 판단하였다. 그러나, 데이터 전송 지연 및 데이터 손실이 고려되지 않았으므로 향후 이러한 경우를 가정한 추가적인 연구가 필요할 것이라고 하였다.

Cho et al.(2020)는 딥러닝 기법인 RNN LSTM을 통해 공공자전거 대여량을 예측하는 모델을 구축하고 기존의 시계열 분석 기법인 지수평활법, ARIMA와 비교·평가하였다. 각 모델을 통해 한 달 동안의 공공자전거 대여량 예측 결과를 평균절대오차 등을 통해 비교한 결과, 딥러닝 모델의 성능이 다른 기존의 시계열 분석 기법들보다 우수하였다. 해당 연구를 통해 시계열 분석에 RNN LSTM 신경망이 효과적이라는 것을 확인하였으나, 향후 더욱 관련성 높은 데이터를 추가하고 peak와 off-peak로 시간대를 구분하여 모델을 구축한다면 예측력이 더욱 높아질 것으로 판단하였다.

III. 자료 수집 및 분석

1. 대상 지점 선정 및 자료 수집

본 연구를 위한 대상 교차로는 영상검지기가 설치되어 있고 교통량의 변동 패턴이 시간대별로 구분 가능한 흐름을 나타내는 장소로 한정하였다. 이를 위하여 서울시 강남구에 위치한 교차로들을 대상으로 위와 같은 특성을 확인하였고, 이 중 도곡역과 대치역 사거리의 접근로를 대상 지점으로 선정하였다. 도곡역과 대치역 교차로에는 현재 영상검지기가 설치되어 운영되고 있고, 다양한 원시 이력 자료가 수집되어 본 연구에 활용이 가능하였다. 도곡역과 대치역 교차로의 총 8개의 접근로 중 교통량이 가장 많아 혼잡한 대치역 교차로

의 ‘학여울역’ 방면 접근로를 대상지로 선정하였고, 분석에 사용될 자료는 2020년 08월 01일에서 2020년 10월 31일까지 수집하였다.

대상 지점의 원시 이력 자료는 수집일시, 교차로 이름, 카메라 이름, 방향, 차로, 대기행렬길이, 점유율, 보행자, 교통량 합계와 직진, 좌회전, 우회전, 유턴 교통량을 대형, 소형, 버스 교통량으로 구분하여 총 21개의 열로 구성되어 있다. 영상검지기에서 수집되는 대기행렬길이는 차로에 존재하는 차량 수와 차종을 바탕으로 산출되며, 이는 2초 간격으로 수집하여 5분이 되는 시점에서의 평균값을 산출한다. 이는 다양한 실시간 정보와 사고 정보, 예측 정보가 5분 단위로 실행되는 점을 반영하였기 때문이다.

그리고 본 연구에서는 해당 접근로의 직진 전용차로에 대한 대기행렬길이만을 고려하였다. 따라서 수집된 교통량 자료 중 좌회전, 우회전, 유턴 교통량은 분석 과정에서 포함하지 않았다. 또한 교차로 이름, 카메라 이름, 보행자 자료 등도 대기행렬길이와 상관성이 매우 낮은 관계로 분석 과정에서 제외하였다.

2. 최종 데이터 셋 구성

영상검지기에서 추출된 자료를 딥러닝 모형에 입력하기 위하여 일정한 형식으로 자료를 가공하였다. 먼저 데이터의 수집 및 분석 주기는 2020년 08월 1일 00시 00분부터 2020년 10월 31일 23시 55분까지 5분 단위로 설정하였다. 본 연구에서는 접근로의 직진 차로에 대한 분석을 진행하였고, 대치역 교차로의 ‘학여울역’ 방면 접근로의 직진 차로는 2, 3, 4차로에 해당한다. 이때, 직진 차로 이용자의 차로 선택은 무작위하며, 차로 선택에 있어 다른 영향이 없다고 가정하였다. 따라서 해당 시간에 대한 직진 차로의 대기행렬길이, 점유율, 대형, 소형, 버스 교통량은 평균값을 계산하여 적용하였다.

그리고 시간 정보를 단위별로 구분하여 입력하였다. 이 중 연도는 본 연구에서는 2020년도 자료만 사용하였으므로 자료 구성에 제외하였으며 나머지 시간 특성에 대하여는 대기행렬의 시간별 특성이 나타날 것으로 판단되어 구분하여 사용하였다. 또한 평일과 주말 역시 대기행렬길이의 변화에 영향을 줄 것으로 판단하여 요일 정보를 자료에 포함하였다.

위와 같은 과정을 거쳐, 딥러닝 모형에 적용할 최종 데이터 셋을 구축하였고 최종적인 자료 형태는 다음 <Table 1>에 제시되어 있다. 최종 데이터 셋은 11개의 열과 26,497개의 행으로 구성된다. 따라서 딥러닝 모형에 적용되는 데이터의 개수는 약 29만 개로 파악되었다.

<Table 1> Example of Final Data Set

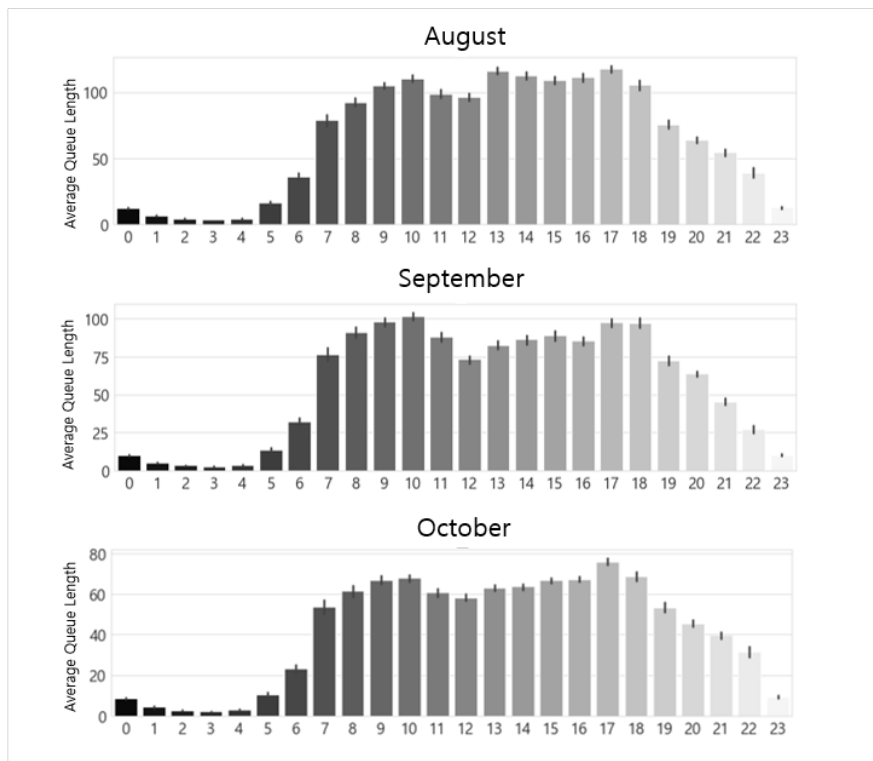
Date & Time	Queue length(m)	month	Date	Hour (h)	Minute (m)	Day	Occupancy (%)	Traffic Volume(TH)(v/5min)		
								Full-Size	Compact car	Bus
2020-08-01 0:00	13	8	1	0	0	7	12	0	15	0
2020-08-01 0:05	11	8	1	0	5	7	11	0	16	0
2020-08-01 0:10	17	8	1	0	10	7	18	0	16	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2020-08-01 8:15	52	8	1	8	15	7	39	0	18	0
2020-08-01 8:20	58	8	1	8	20	7	43	0	20	1
2020-08-01 8:25	64	8	1	8	25	7	48	0	23	0
2020-08-01 8:30	76	8	1	8	30	7	51	0	28	0
2020-08-01 8:35	64	8	1	8	35	7	48	0	24	0
2020-08-01 8:40	91	8	1	8	40	7	58	0	35	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

3. 데이터 기초분석

작성된 최종 데이터 셋에서 월, 요일, 시 등의 시간 정보와 대기행렬길이를 이용하여 자료에 대한 기초적 분석을 수행하였다. 이때 분석을 위하여 Python 3.7과 NumPy, Pandas, Seaborn을 이용하였다.

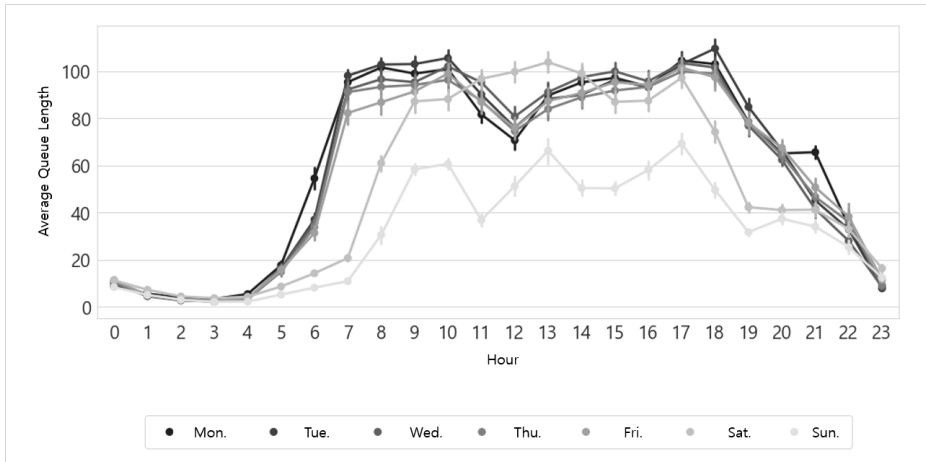
먼저 월별 평균 대기행렬길이를 분석한 결과, 8월의 평균 대기행렬길이는 65.84m이었으며, 9월은 56.50m, 10월은 42.04m이었다. 다음으로 요일별 평균 대기행렬길이를 분석한 결과, 월요일의 평균 대기행렬길이가 61.82m로 가장 길었으며, 화요일이 61.27m로 두 번째로 길었다. 수요일은 60.25m, 목요일은 58.40m, 금요일은 58.69m이었다. 반면, 주말인 토요일과 일요일은 각각 51.28m, 31.98m로 비교적 짧았다. 이는 출퇴근 교통량 수요의 크기에 따른 영향으로 생각할 수 있다. 마지막으로 시간별 평균 대기행렬길이를 분석한 결과, 17시와 10시의 평균 대기행렬길이가 각각 97.06m, 93.26m로 가장 길게 나타났다. 평균 대기행렬길이는 출퇴근 시간대를 중심으로 크기의 변화가 발생하는 것으로 파악되었다.

시간 정보를 복합적으로 구성하여 월-시간별 대기행렬길이, 요일-시간별 대기행렬길이를 분석하였다. 이는 월 특성이나 요일 특성에 따라 시간대별 대기행렬길이의 변화를 파악하기 위함이다. <Fig. 1>은 월-시간별 평균 대기행렬길이를 나타낸다. 약간의 차이는 보이지만, 8월, 9월, 10월 모두 오전에는 10시에, 오후에는 17시에 대기행렬길이가 가장 긴 것으로 나타난다. 또한 7시에 평균 대기행렬길이가 매우 증가하고 18시에 감소하는 추세도 월별로 유사하게 나타났다. 따라서 평균 대기행렬길이는 월별로 큰 차이 없이 출퇴근 교통량에 영향을 받으며 변동 추세도 매우 유사함을 알 수 있다.



<Fig. 1> Average Queue Length by Month-Hour

<Fig. 2>는 요일-시간별 평균 대기행렬길이를 나타낸 결과이다. 평일의 결과 그래프들은 시간대별 대기행렬 길이의 변화 추세가 매우 유사하게 나타났다. 7시에 평균 대기행렬길이가 대폭 증가하고, 18시 이후에는 감소하였다. 반면 주말에 해당하는 그래프들은 앞서 언급했던 시간대에서 평균 대기행렬길이가 감소하였다. 토요일에는 12시에 평균 대기행렬길이가 가장 크게 나타났고, 일요일에는 거의 모든 시간대의 평균 대기행렬길이가 다른 요일들과 비교하여 감소된 것으로 파악되었다. 그리고 날짜별로 분류된 각 요일별 추세를 분석한 결과 날짜별로 변동폭은 매우 크게 나타났지만, 시간대별로 명확한 패턴이 나타나지 않음을 확인하였다.



<Fig. 2> Average Queue Length by Day-Hour

IV. 모형 구축 및 평가

본 연구에서는 교차로에서의 대기행렬길이를 예측하고자 RNN LSTM 신경망을 이용한 딥러닝 모형을 구축하였다. 그리고 통계적 기법인 다중회귀분석 모형을 구축하여 두 모형의 결과를 비교·평가하였다. 예측 모형들의 결과를 비교하기 위하여 다양한 평가지표가 사용된다. 본 연구에서는 오차를 평가하는데 일반적으로 많이 사용되는 평균절대오차(mean absolute error)와 평균제곱근오차(root mean square error)의 두 가지 지표를 적용하였다. 다중회귀분석 모형과 딥러닝 모형을 구축하는 과정에 2020년 08월 01일 00시 00분부터 2020년 10월 30일 23시 55분까지의 5분 단위 자료를 동일하게 적용하였다.

1. 다중회귀분석 모형 구축

다중회귀분석에서 사용되는 독립변수의 개수가 너무 많은 경우 다중공선성으로 인하여 결과의 신뢰성이 낮아질 수 있다. 따라서 독립변수의 개수를 줄이기 위하여, 먼저 신경망 모형에 맞춰 가공된 최종 자료에서 월, 일, 시, 분 데이터를 변환하여 하나의 변수로 처리하였다. 변환된 자료를 활용하여 다중회귀분석 모형을 구축하여 평가한 결과, 대형과 소형 교통량의 유의확률 p값이 0.05보다 큰 값으로 산출되어 이를 독립변수에서 제외된 후 나머지 독립변수들에 대하여 추가로 모형을 구축하였다.

구축된 다중회귀분석 모형의 결과는 다음 <Table 2>에 요약되어 있다. 모형의 유의확률이 0.001 이하로 결

정되어 산출된 회귀식이 통계적으로 유의하다는 결론이 도출되었다. 그리고 모형의 설명력을 나타내어주는 조정(Adjusted)결정계수는 0.960으로 모형의 설명력이 굉장히 높고, VIF값은 모두 10 미만으로 파악되어 다중공선성에 문제가 없는 것으로 나타났다. 또한 모든 독립 변수들의 p값이 0.05보다 작은 값으로 산출되어 종속변수에 대하여 유의한 영향을 미친다고 판단되었다. Beta 값에서 독립변수들의 종속변수에 대한 영향력을 살펴보면, 점유율이 가장 큰 영향을 미치는 것을 확인할 수 있다.

<Table 2> Summary of Multiple Regression Analysis Results

Dependent Variable	Independent Variable	Unstandardized Regression Coefficient		Standardized Regression Coefficient	P-Value	Mulicollinearity Statistic	
		B	Std Error	Beta		Tolerance	VIF
Queue	Time	-0.791	0.014	-0.069	<0.001	0.975	1.024
	Weekday	0.323	0.027	0.015	<0.001	0.992	1.006
	Occupancy	1.672	0.002	0.967	<0.001	0.179	1.054
	bus	-1.302	0.272	-0.006	<0.001	0.951	1.026

이와 같은 결과로부터, 종속변수 Y(대기행렬길이)를 추정하기 위한 다중회귀분석 모형은 다음과 같은 식으로 표현될 수 있다.

$$Y = 18.726 + 0.323X_1 + 1.672X_2 - 1.302X_3 - 0.791X_4$$

여기서, Y = Queue length

X₁ = Weekday

X₂ = Occupancy

X₃ = Bus traffic

X₄ = Time

2. 딥러닝 모형 구축

1) 데이터 구성

대기행렬길이를 예측하기 위한 딥러닝 모형을 구축하기 위하여 <Table 1>에 제시된 자료를 적용하였다. 자료는 5분 단위의 대기행렬길이, 월, 일, 시, 분, 요일, 점유율, 대형, 소형, 버스 교통량 정보를 포함하고, 총 개수는 약 29만 개로 구성되었다.

본 연구에서는, 학습 데이터로 2020년 08월 01일부터 2020년 10월 29일까지의 80일에 해당하는 자료를 이용하였고, 총 259,200개의 데이터로 구성되었다. 검증 데이터는 2020년 10월 30일, 하루의 5분 단위 데이터로 구성되며, 총 2,880개의 데이터로 이루어진다. 모형 평가 자료는 검증 데이터와 마찬가지로 2020년 10월 31일의 5분 단위 데이터로 구성하였다. 그리고 결과 비교를 위하여 모형 평가 자료와 동일한 날에 영상검지기로부터 실제 대기행렬 길이를 수집하여 활용하였다.

2) 모형 구성

먼저 RNN에 LSTM을 적용하여 시계열 자료 분석에 더 알맞게 구성하였다. 또한 과적합을 방지하기 위해 Drop Out을 0.3으로 설정하여 LSTM층과 연결하고 이때 Stateful을 활성화하였다. Stateful은 이전 학습 시의 가중

치를 다음 학습 시의 초기치로 전달하는 방법으로, 시계열 자료 분석에 효과적이라고 알려져 있다. Optimizer는 가장 일반적으로 많이 이용되며, 효율적으로 계산하는 덕분에 수렴 속도가 빠르며 성능이 우수한 adam을 사용하였다. Input Dimension은 데이터 속성의 개수로, 입력 시 10으로 설정하고, 각각은 대기행렬길이, 월, 일, 시, 분, 요일, 점유율, 대형, 소형, 버스 교통량을 의미한다. 그리고 출력층은 Dense로 구성하여 한 개의 예측치를 도출하도록 설정하였다. 학습의 횟수를 의미하는 epoch는 500으로 설정하고, 여기에 Check Point를 함께 입력하였다.

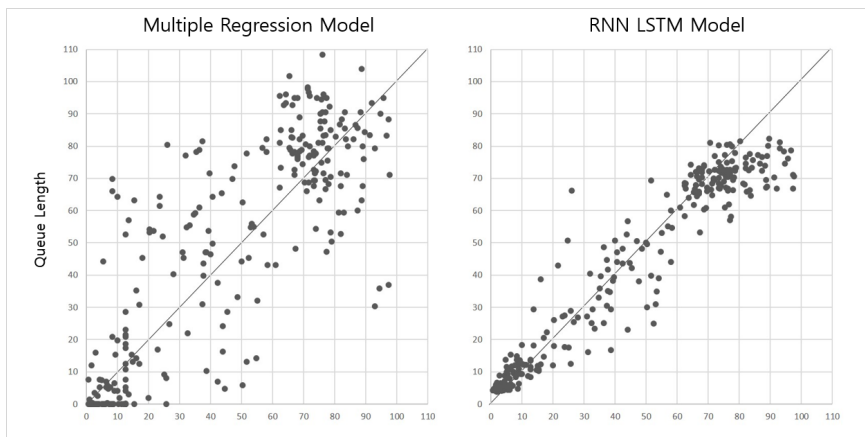
그리고 은닉층의 개수와 Look Back의 길이에 따른 평가 결과를 비교하여 최적 모델을 선정하였다. 우선 LSTM과 Drop Out층으로 구성된 은닉층을 2겹, 3겹, 4겹으로 나누어 평가지표를 비교하였다. Look Back은 입력 시 사용하는 자료를 일정한 길이로 나눌 때 사용하는 것으로, 6, 12, 36으로 나누어 평가지표를 비교하였다. 본 연구에서 사용하는 데이터가 5분 단위이므로 이는 각각 30분, 1시간, 3시간을 의미한다. 각각의 조건이 반영된 개별 모델을 구축하였고, 검증 데이터 기반의 평균절대오차와 평균제곱근오차를 산출하였다. 그 결과, 은닉층이 4겹일 때와 Look Back이 6일 때의 평균절대오차와 평균제곱근오차가 6.34와 8.99로 제일 적은 값으로 산출되었다. 따라서 4겹(DropOut 2겹+LSTM 2겹)의 은닉층으로 구성되고, Look Back이 6인 딥러닝 모델을 최종 모델로 선정하였다.

3. 모델 비교 평가 결과

앞서 제시된 다중회귀분석 모델과 딥러닝 모델에 대하여 각각 2020년 10월 31일, 1일 동안의 5분 단위 대기행렬길이를 예측하였다. 그리고 예측된 대기행렬길이와 실제 관측된 대기행렬길이를 비교하여 평균절대오차와 평균제곱근오차를 산출하였다.

다중회귀분석 모델을 통한 예측 결과는 평균절대오차가 13.65, 평균제곱근오차는 19.10이 산출되었다. 딥러닝 모델의 예측 결과는 평균절대오차가 6.44, 평균제곱근오차는 9.11로 계산되었다. 이러한 결과는 다중회귀분석 모델 대비 딥러닝 모델의 평균절대오차가 약 52.8%, 평균제곱근오차가 약 52.3% 더 감소한 것이다.

다음 <Fig. 3>은 실제 대기행렬길이와 모델별 예측 대기행렬길이를 각각 x축과 y축으로 5분 단위 산점도로 나타낸 것이다. 다중회귀분석 모델의 산점도가 딥러닝 모델과 비교하여 자료의 산포가 크고, 예측 오차의 크기와 분포도 더 큰 것을 쉽게 확인할 수 있다.



<Fig. 3> Scatter plot of actual and forecasted queue length

V. 결 론

도시부 도로의 혼잡 상황을 대표하는 변수 중 하나는 대기행렬길이이다. 대기행렬길이는 신호 교차로 운영 효율을 평가하는 지표로 사용되고, 프랑스의 교통 관리 시스템 SIRIUS의 사례와 같이 도로 이용자들이 해당 구간의 혼잡 수준을 확인하여 경로 선택에 영향을 미치는 교통정보로 사용될 수 있다. 현재 루프 기반의 검지기 자료를 이용하여 대기행렬길이 추정되고 있으나 결과의 신뢰성과 정확성이 부족하여 이에 관한 개선 필요성이 제기되고 있다. 본 연구에서는 영상검지기에서 수집되는 정보를 활용하여 딥러닝 기반으로 대기행렬길이를 예측하는 모형을 개발하였다. 그리고 통계적 기법인 다중회귀분석 모형을 구축하여 두 모형의 예측 결과와 현장 실측 자료를 평균절대오차와 평균제곱근오차의 두 지표를 이용하여 평가하였다.

연구로부터 얻어진 주요 결과는 다음과 같다. 첫째, 다중회귀분석을 통해 대기행렬길이에 대하여 시간, 요일, 점유율, 버스 교통량이 유효한 변수로 도출되었고, 독립변수들의 종속변수에 대한 영향력은 점유율이 가장 큰 것으로 나타났다. 둘째, 딥러닝 모형은 은닉층이 4겹, Look Back이 6일 때 평균절대오차와 평균제곱근오차가 6.34와 8.99로 제일 적은 값으로 산출되었다. 셋째, 두 가지 모형을 평가한 결과, 다중회귀분석 모형과 딥러닝 모형의 평균절대오차는 각각 13.65와 6.44, 평균제곱근오차는 각각 19.10과 9.11로 계산되었다. 이는 다중회귀분석 모형 대비 딥러닝 모형의 평균절대오차가 약 52.8%, 평균제곱근오차는 약 52.3% 감소된 결과이다.

본 연구에서 활용된 영상검지기 자료는 5분 단위이므로 신호제어 용도로는 사용이 어렵지만, 도로 이용자들에게 대기행렬길이 예측 정보를 제공하여 교통 혼잡을 완화하는 용도로 이용 가능할 것으로 평가된다. 그러나 영상 검지시스템의 데이터 수집이 주기내 단위 이내로 짧아질 경우, 본 연구에서 제시된 모형은 향후 신호제어 용도로도 사용될 수 있다. 본 연구에서는 COSMOS 대기행렬길이 모형을 비롯하여 선형연구에서 제안된 대기행렬길이 모형들과의 평가가 자료 수집의 한계로 수행되지 못하였고, 향후 이에 관한 연구가 필요하다.

ACKNOWLEDGEMENTS

이 논문은 2021년도 정부(경찰청)의 재원으로 과학치안진흥센터의 지원을 받아 수행된 연구임.
(No.092021C28S01000, 자율주행 혼재 시 도로교통 통합관제시스템 및 운영기술 개발)

REFERENCES

- Cho G., Lee S. and Nam D.(2020), "Forecasting of Rental Demand for Public Bicycles using a Deep Learning Model," *Journal of the Korean Institute of Intelligent Transport Systems*, vol. 19, no. 3, pp.28-37.
- Do C.(2010), *Principles of Transportation Engineering*, Gyomoon Publishers.
- Fu R., Zhang Z. and Li L.(2016), "Using LSTM and GRU Neural Network Methods for Traffic Flow Prediction," *31st Youth Academic Annual Conference of Chinese Association of Automation*, pp.324-328.
- Gao K., Huang S., Han F., Li S., Wu W. and Du R.(2020), "An Integrated Algorithm for Intersection Queue Length Estimation Based on IoT in a Mixed Traffic Scenario," *Applied Sciences*, vol. 10,

- no. 6, 2078.
- Han J., Kim S. and Choi B.(2000), “Training Sample of Artificial Neural Networks for Predicting Signalized Intersection Queue Length,” *Proceedings of the KOR-KST Conference*, vol. 18, no. 4, p.75.
- Kang J. and Oh Y.(2005), “The Development of Traffic Queue Length Measuring Algorithm Using the Occupancy Rates,” *Proceedings of the KOR-KST Conference*, vol. 48, p.309.
- Ki B., Han J. and Kim S.(2002), “A Study on Prediction of Queue Length Based on the Neural Network and Genetic Algorithm for a Signalized Intersection,” *Journal of the Korean Society of Civil Engineering D*, vol. 22, no. 4, p.595.
- Kim S.(2018), *Coding Chef’s 3-minute Deep Learning, Keras flavor*, Hanbit Publishing Network.
- Kim T.(2017), *Python Deep Learning Keras with Blocks*, Digitalbooks.
- Lee C.(2000), *Introducing ATIS to Seoul Metropolitan*, Seoul Development Institute.
- Lee Y., Son B., Kim H.(2003), “Development of A Simple Method for Determining Queue-End-Location”, *Proceedings of the KOR-KST Conference 44*, pp.1 - 6.
- Lee Y. and Oh Y.(2002), “A Development of Queue Length Estimation in COSMOS,” *Proceedings of the KOR-KST Conference*, vol. 42, p.1.
- Lee Y., Hwang J., Kim S. and Lee C.(2018), “Development of Vehicle Queue Length Estimation Model Using Deep Learning,” *Journal of the Korean Institute of Intelligent Transport Systems*, vol. 17, no. 2, pp.39-57.
- Liu H. X., Wu X., Ma W. and Hu H.(2009), “Real-time Queue Length Estimation for Congested Signalized Intersections”, *Transportation Research Part C*, 17(4), p.412.
- Luo X., Ma D., Jin S., Gong Y. and Wang D.(2019), “Queue Length Estimation for Signalized Intersections Using License Plate Recognition Data,” *IEEE Intelligent Transportation System Magazine*, p.209.
- Nikhil B.(2017), *Fundamentals of Deep Learning* (1st ed.), O’Reilly Media.
- Sim S., Lee C. and Choi K.(2002), “A Queue Length Prediction Algorithm using Kalman Filter,” *Journal of Korean Society of Transportation*, vol. 20, no. 5, p.145.
- Yim Y. and Ygnace J. L.(1996), “Link Flow Evaluation Using Loop Detector Data: Traveler Response to Variable-Message Signs”, *TRR*, Volume 1550 Issue 1, pp.58 - 64.
- Yin J., Sun J. and Tang K.(2018), “A Kalman Filter-Based Queue Length Estimation Method with Low-Penetration Mobile Sensor Data at Signalized Intersections,” *TRR*, vol. 2672, no. 45, pp.253-264.