



코로나19 확산에 따른 대중교통 이용량의 예측 모형 구축*

Developing Prediction Models for Public Transportation Passenger Flow under the Spread of COVID-19

한성혜** · 이경재*** · 추상호**** · 오관교***** · 정준영*****

Han, Seonghae · Lee, Gyeongjae · Choo, Sangho · Oh, Kwankyo · Jung, Junyoung

Abstract

The outbreak of COVID-19 significantly changed peoples' lifestyles, also contributing to a change in their travel behavior. Since public transit usage has decreased due to the fear of infection, improving the passenger flow prediction will help public transportation companies better manage this situation. This study analyzes the prediction of public transportation flow made in South Korea during COVID-19 and proposes a suitable model for prediction. Four models were constructed to predict passenger ridership, and the model with the lowest evaluation metrics value for each city was selected. During the model constructing process, it was found appropriate to predict a day with the previous 14 days, and the prediction results showed that the RNN model performed better than the other models in most of the cities. The prediction results of the RNN model were much better in areas other than the metropolitan areas, while the LSTM model performed better in metropolitan areas. The result shows that the RNN model would perform better at predicting public transportation usage in a short-term pandemic situation such as COVID-19. This research is expected to help companies make better decisions pertaining to public transportation operations, such as scheduling and adjustment of dispatch intervals, during pandemic situations.

주제어 코로나19, 대중교통, 딥러닝, 이용량 예측, 교통카드

Keywords COVID-19, Public Transportation, Deep Learning, Passenger Flow Prediction, Transit Card

1. 연구의 배경 및 목적

1. 연구의 배경

전 세계적으로 코로나19 바이러스의 등장은 마스크를 상시 착용하고 대면 활동이 감소하는 등 우리의 생활양식 변화에 많은 영향을 미쳤다. 2019년 중국 우한에서 최초로 코로나19의 확진자가

발생하였으며, 빠르게 확산하여 2020년 3월 11일 세계보건기구(WHO)에 의해 세계적 대유행인 팬데믹(pandemic)이 선언되었다. 코로나19 바이러스는 SARS-CoV-2 감염에 의해서 발생하는 호흡기 중후군으로 비말을 통해 감염이 이루어지며, 비말 감염의 특성상 집단 모임에 취약하다. 우리나라는 2020년 1월 20일에 확진 사례가 처음으로 보고되었으며, 2월 19일 대구종교집단 집단감염의 발생을 필두로 크게 확산되기 시작하였다. 코로나19

* 이 논문은 2021년 한국교통안전공단의 지원을 받아 진행된 연구이며, 2022년 4월 대한교통학회 춘계학술대회에서 발표한 논문을 수정·보완하였음.

** Master's Candidate, Department of Urban Planning, Hongik University (First Author: lizahan96@gmail.com)

*** Doctorate Candidate, Department of Urban Planning, Hongik University (dl874500@gmail.com)

**** Professor, Department of Urban Design and Planning, Hongik University (Corresponding Author: shchoo@hongik.ac.kr)

***** Researcher, Korea Transportation Safety Authority (fri7979@kotsa.or.kr)

***** Researcher, Korea Transportation Safety Authority (jjyoung@kotsa.or.kr)

의 확산이 본격적으로 일어나면서 신규 확진자 수는 계속해서 늘어나는 양상을 띠었으며, 2021년 11월 5일에는 누적 확진자 수가 375,464명까지 증가하였다(질병관리청, “발생현황”, 2022).

이러한 코로나19 바이러스의 강력한 전염성에 대응하여 중앙 정부는 시설별 운영시간 제한, 사적 모임 인원 제한과 같은 ‘사회적 거리두기’를 시행하였으며, 단계별로 상향됨에 따라 대면 활동이 점차 감소하는 모습을 보였다. 학교에서는 온라인으로 수업을 진행하고 기업에서도 재택근무를 시행하였으며, 온라인 판매와 배송 서비스를 활용하는 등의 언택트(untact) 소비가 증가하였다(배영임·신혜리, 2020). 또한, 대면 모임을 지양함에 따라 출퇴근과 등하교 등의 목적이 아닌 여가활동 등의 불필요한 활동과 외출 역시 감소하였으며, 사람들의 이동에 영향을 미쳤다. 코로나19의 확산에 따라 전국적으로 이동량이 감소하였으며, 폐쇄된 공간에 밀집되어 이동하는 교통수단인 고속·시외버스, 철도, 항공을 중심으로 두드러지게 나타났다.

불특정 다수가 함께 이용하는 버스와 지하철 등의 대중교통은 코로나19에 취약하여 많은 사람의 기피 대상이 되었으며, 이에 따라 사람들은 접촉이 적은 자가용이나 공공자전거, 전동킥보드와 같은 개인형 교통수단의 이용을 선호하게 되었다(신송현·추상호, 2022). 전체적으로 코로나19의 확산에 따른 대중교통 이용량의 감소는 <그림 1>과 같이 대규모 감염에 따라서 크게 변동을 보였으며, 이러한 대중교통 이용률의 감소는 2019년 대비 2020년에 전국이 27.0%, 수도권이 26.9%까지 하락한 것으로 나타났다. 특히 인구가 가장 많은 서울시의 경우에는 광역·도시철도 이용률은 25.0%, 시내버스 이용률은 27.4% 감소하였다(국토교통부, 2021). 또한, 조혜림 외(2020)의 연구결과에 따르면 도로교통 이용자의 감소 폭 대비 대중교통 이용자의 감소 폭이 더 큰 것으로 확인되었다.

2. 연구의 목적

최근 환경문제가 대두되면서 지속 가능한 교통수단으로서 대중교통의 역할은 더욱 중요히 여겨지고 있다. 그러나 코로나19의 확산 이후 감염의 위험성에 따라 사람들의 이동행태에 변화를 초래했으며, 대중교통 이용률은 계속해서 감소하는 추세를 띠었다. 대중교통은 다수의 사람이 이용하는 교통수단으로 삶의 중요한 부분이며, 교통흐름의 정밀한 예측을 통해 대기 시간을 감소시키는 것과 같이 생활의 편의를 극대화할 수 있다. 또한, Jiao et al.(2021)은 대중교통의 승객 변동 추세 예측을 통해 직원의 운영 일정을 최적화하고 비상 대응 능력을 강화하는 등, 대중교통 운송 회사에 많은 도움을 가져다줄 수 있다고 하였다. 이에 따라 코로나19와 같은 감염병의 발생 시 대중교통의 운영 흐름의 판단과 예측이 중요하다고 판단되며, 본 연구에서는 이와 같은 코로나19 상황을 반영한 대중교통 이용량을 예측하고자 하였다.

대중교통 이용량 예측을 위해 시계열 예측에 많이 사용되는 모형들을 선정하였으며, 성능 평가 지표를 이용하여 비교 평가를 진행하였다. 연구의 공간적 범위는 대한민국의 17개 시도를 설정하였으며, 시간적 범위는 코로나19 확산 이후의 약 2년간인 2020년 1월 20일부터 2021년 11월 5일까지로 설정하였다. 자료는 해당 기간 동안 교통카드빅데이터 통합정보시스템, 기상청, 질병관리청에서 구득하였으며, 이를 기반으로 코로나19 확산 이후 대중교통 이용량 예측에 가장 적합한 모형을 구축하고자 하였다.

II. 선행연구 고찰

코로나19의 발생 이후, 확산에 따른 사회 변화와 관련된 다양한 연구가 많이 이루어지고 있다. 교통은 사람들이 살아가는 도시를 이루는 한 부분으로서, 코로나19와 밀접한 관련이 있어 이에 관한 관심 역시 높아지고 있다. 질병의 확산으로 교통에 미치는 영향이 커짐에 따라 관련된 기존 연구를 검토해보았다. 기존 연구는 크게 코로나19와 교통 관련 연구 그리고 교통 이용량 예측 모형 관련 연구로 구분하여 검토하였고 본 연구의 차별성을 제시하였다.

1. 코로나19와 교통 관련 연구

전 세계적으로 코로나19의 확산은 삶의 방식에 변화를 가져왔으며, 교통부문에 많은 영향을 미쳤다. 실제로 OECD(2020)에 의하면 코로나19로 인해 대중교통의 이용과 도로교통량 등의 이동량이 급격히 감소한 것으로 분석되었다. 반면에 도로와 자전거, 공유 모빌리티의 이용이 증가한 것으로 나타나, 사람들의 통행행태에 많은 변화가 일어난 것을 확인할 수 있다. 또한, 재택근무, 온라인 강의 및 온라인 쇼핑 등 비대면 활동이 활성화되면서 외출의 필요성은 더욱 줄어들었다. 이에 대해 코로나19 확산 이후 교통부문의 영향에 관한 연구들이 활발히 수행되고 있으며, 관련된 연구들을 검토하여 동향을 파악하고자 하였다.

먼저, 코로나19가 교통부문 전반에 걸쳐 미치는 영향과 관련한 연구가 확인되었다. 전체적으로 교통량의 감소가 일어난 것으로 나타났으며, 이수진 외(2020)는 코로나19 및 관련 정부 지침이 대전시 통행 수요에 미치는 영향을 다양한 교통수단에 대해 분석하였다. 시간적 범위는 2018년부터 2020년으로 설정하여 대전시 교통량의 변화에 대해 분석을 진행하였다. 분석 결과, 평일보다 주말 통행량의 감소 폭이 더 크게 나타났으며, 이는 평일에는 통근과 같은 필수 통행이 일어나는 것에 비해 주말의 통행은 비교적 선택적인 활동들이 주를 이루는 특성을 갖기 때문으로 해석하였다. 또한, 버스와 지하철의 이용률은 감소하고 자전거의 이용률은 증가한 것으로 확인되었다. 이는 다수의 사람이 함께 이용하는 대중교통의 특성상 코로나19의 감염 위험이 높아 이용객이 감소하였으며, 개인형 교통수단인 자전거의 경우 대중교통에 비해

접촉의 위험성이 낮아 선호되는 것으로 해석하였다. 이를 통해 교통수단의 특성에 따라서 코로나19의 비말 감염이 시민들의 이용에 영향을 미친다고 설명하였다.

장동의 외(2020)는 육상교통부문의 변화를 분석하였으며, 앞선 연구와 같은 맥락으로 모든 부문에서의 교통량이 감소한 것으로 나타났다. 자료의 수집 기간은 2019년 1월부터 2020년 3월까지로, 국내 여객의 통행량을 비교 분석하였으며, 특히 고속버스, 시외버스 그리고 철도 등의 장시간 폐쇄되어 이동하는 교통수단을 중심으로 교통량의 감소가 확인되었다. 또한, 대중교통의 경우, 대규모 확진자가 발생한 지역에 대해서 지역 간 통행 기피 현상이 나타났다. 이를 토대로 코로나19의 여파로 여객운송산업 부문에 피해 규모가 커질 것으로 예상하였으며, 이에 따라 버스 운행 축소, 여객운송 사업자의 지원 확대 등의 산업 지원 대책을 제시하였다.

일부 연구에서는 이처럼 감염에 의한 교통량의 변화가 확인됨에 따라 시민들의 대중교통 이용에도 영향을 미친 것으로 확인되었다. Jenelius and Cebecauer(2020)은 스웨덴의 일일 대중교통 이용객에 대한 코로나19의 영향을 규명하고자 하였다. 스톡홀름, 베스트라예탈란드와 스코네를 대상으로 분석을 진행하였으며, 코로나19의 여파가 가장 강했던 2020년 3월 1일부터 2020년 5월 31일까지 자료를 수집하였다. 분석 결과, 월정기권을 사용하던 승객들은 1회권 등으로 변경하였으며, 관광객들이 주로 사용하는 단기권의 사용량 및 판매량은 거의 0 수준으로 떨어져 이용되지 않는 것으로 나타나는 등, 다른 교통수단에 비해 대중교통 이용객의 감소가 심각한 것으로 확인되었다.

또한, Przybylowski et al.(2021)은 이동에 대한 의향과 감염 안전 기준에 대한 인식 측면에서 코로나19가 모빌리티 이용 특성에 미치는 영향에 대해 분석하였으며, 대중교통 이용자를 대상으로 하였다. 연구는 2020년 5월과 6월에 폴란드 그단스크시에서 설문 조사를 통해 진행되었다. 설문 조사 결과에 따르면 대중교통 이용량은 줄어든 반면에 개인형 이동수단의 이용량은 늘어난 것으로 나타났다. 이처럼 코로나19의 확산 이후로 감염에 대한 두려움으로 인해 대중교통 및 공유 모빌리티와 같은 지속 가능한 통근 수단의 이용이 급격히 감소하고, 이와 반대로 접촉이 적은 자동차와 자전거 같은 개인 이동수단이나 도보를 선호하는 것으로 조사되었다. 종합적으로 이러한 결과는 코로나19 이후 생긴 감염 위험에 대한 인식이 대중교통 이용 의사에 영향을 미치는 것으로 확인되었다.

감염 위험에 대한 인식이 대중교통과 개인 이동수단의 이용량에 영향을 미치는 것으로 나타나며 이 둘의 관계에 관해 분석한 연구도 진행되었다. Teixeira and Lopes(2020)은 코로나19의 발생 이후 공유 자전거와 지하철 운영의 관계를 통해 팬데믹 상황이 도시 교통 시스템에 미치는 영향을 분석하였다. 미국 뉴욕시의 2019년과 2020년 2, 3월의 공유 자전거와 지하철 자료를 사용하여 분석을 진행하였다. 분석 결과에 따르면 코로나19 이후 지하

철과 공유 자전거의 이용량이 모두 감소하는 것으로 나타났지만, 자전거의 경우에 이용자 감소율이 훨씬 낮으며 오히려 평균 이동 시간은 증가하는 모습을 보였다. 이에 따라 공유 자전거의 회복 탄력성이 지하철보다 뛰어난 것으로 확인되었다. 또한, 지하철에서 공유 자전거로의 수단 변경이 이루어지는 것으로 분석하였다. 이러한 결과를 토대로 앞으로 코로나19와 같은 전염력이 강한 질병들에 대해 공유 자전거가 도시 교통 시스템의 회복에 도움이 될 것이라고 주장하였다.

다양한 연구들이 교통의 이용량과 관련하여 코로나19의 영향력에 대해 분석하였다. 설문 조사를 활용하여 그 목적과 연관 짓고, 나아가 감염병에 의한 봉쇄조치와 관련하여 진행된 연구도 확인되었다. Anke et al.(2021)은 설문 조사를 통해 독일에서 코로나19 팬데믹이 모빌리티 이용 특성에 미친 영향을 분석하고, 추가로 봉쇄조치가 일어난 주와 아닌 주의 모빌리티 이용 변화를 비교하고자 하였다. 2020년 3월 21일부터 2020년 4월 19일까지의 설문 조사 결과, 코로나19의 발생 이후 전체적인 통행량은 감소하였으며, 특히 여행과 방문 등의 여가 목적의 통행이 가장 줄어들었다. 이용 수단에 대해서는 대중교통의 이용량은 줄어든 반면 자동차, 도보, 자전거의 이용은 증가하는 것으로 나타났다. 이러한 현상은 도시와 농촌 지역 모두에서 나타났으며, 이때 수단의 변경 이유에 대한 응답으로 자신이나 타인을 감염시킬 위험성이 가장 높게 조사되었다. 추가로, 봉쇄 조치의 영향은 미미한 것으로 확인되었다.

2. 대중교통 이용량 예측 모형 관련 연구

현대사회에서 대중교통은 교통사고와 교통혼잡 등의 교통문제를 해결하며 많은 사람이 이용하는 중요한 이동수단이다. 수단분담률을 확인하면, 인구가 가장 많은 서울특별시의 경우 코로나19 발생 이전인 2019년에 대중교통이 65.5%를 차지하는 것으로 나타났다(서울특별시, 2022), 주요한 교통수단으로 자리매김하고 있는 것을 확인할 수 있다. 한편, 2020년 1월에 국내 첫 코로나19의 확진 사례가 발생하며 코로나19의 발병과 확산으로 인해 사람들의 대중교통에 대한 인식이 부정적으로 변화하였다. 감염 확산에 대한 위험이 존재하여 이에 따라 대중교통 이용률이 <그림 1>과 같이 감소한 것을 확인할 수 있다. Tirachini and Cats(2020)은 이러한 대중교통은 감염 확산의 위험성이 존재하여 발병 단계에 따라 이용에 대한 제한과 규제가 다르게 조정되어야 하며, 이러한 문제에 대한 자세한 분석이 필요하다고 제시하였다. 이처럼 질병으로 인해 이동과 여행이 제한되면서 이용량이 급격히 감소함에 따라 정확한 도시 교통의 수요 예측이 필요할 것으로 판단된다. 따라서 이와 같은 현황을 고려하여 최근 국내외에서 수행된 대중교통 이용량 예측과 관련된 연구를 검토하고 대중교통 이용량에 영향을 미치는 요인과 모형을 확인하고자 하였다. 최근 국

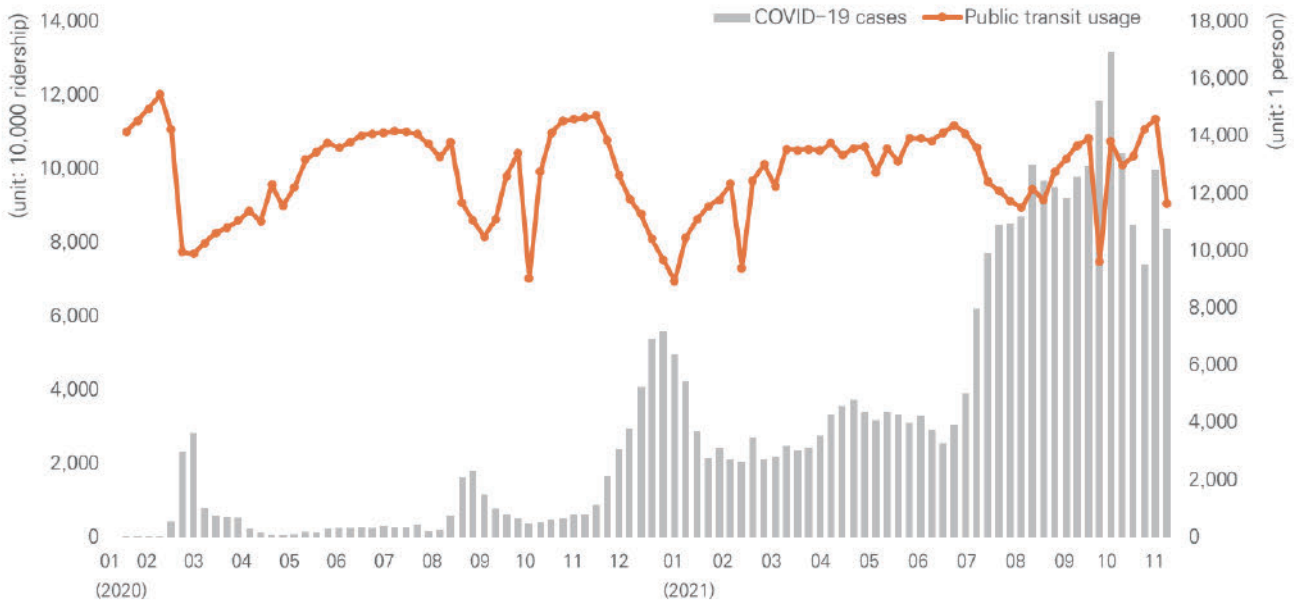


그림 1. 코로나19 확산 및 주별 대중교통 이용량 변화
 Figure 1. Spread of COVID-19 and changes in weekly public transit usage

내외에서 수행된 시계열 자료를 활용한 대중교통 이용량 예측 관련 연구는 다음과 같다.

연구는 크게 코로나19 발생 전후로 나누어 볼 수 있으며, 먼저 발생 이전의 국내 연구로 김동규(2020)는 추세분석법, 지수평활법, ARIMA(Auto-Regressive Integrated Moving Average) 모형을 활용하여 대구시의 도시철도 수송 수요 예측을 수행하였다. 2007년부터 2017년까지 대구 도시철도 1, 2호선의 월별 자료를 사용하여 2018년과 2019년의 수요 예측을 진행하였으며, 시계열 모형을 이용한 각 예측 방법의 정확도를 판단하여 모형의 적용 가능성 및 실효성을 검증하였다. 변수는 도시철도 호선별로 월별 승차 인원 시계열 자료를 이용하였다. 모형별 예측 후 예측치와 실제치를 비교 검증한 결과, ARIMA 모형이 모든 호선에 적용될 수 있으며, 정확성이 높아 가장 최적의 모형으로 판단되었다.

이처럼 대중교통 이용에 관한 자료가 시계열임에 따라 분석 방법은 시계열 분석 방법들을 활용하여 이루어졌으며, 해외 연구로는 Guo et al.(2019)은 중국 광저우 양지역의 승객 수 자료를 활용하여 도시철도 승객 수를 예측하였다. 사용된 변수는 2017년의 승객 유입 및 유출 수와 휴일 여부이다. 예측에 사용된 모형의 종류는 5가지로, SVR(Support Vector Regression), LSTM(Long Short-Term Memory), SVR-LSTM, ARIMA, Fusion-KNN(K-Nearest Neighbor)이 사용되었다. 이들 모형 중 SVR과 LSTM을 융합한 모델인 SVR-LSTM 모형이 코로나19로 인한 불규칙한 승객의 흐름을 가장 잘 예측하여 제일 우수한 것으로 분석되었다.

중국에서 진행된 또 다른 연구로 Sha et al.(2020)은 상하이 지하철역의 교통카드 데이터를 활용하여 미래 승객 수를 예측하였다. 이를 위해서 2016년 3월 한 달 동안 교통카드 데이터와 평일

여부, 풍향, 풍속, 기온, 습도 및 기압의 기상 관측 자료를 수집하였다. 분석은 RNN(Recurrent Neural Network) 기반의 모델인 LSTM(Long Short-Term Memory), GRU(Gated Recurrent Unit)를 사용하여 상하이 지하철 승객 수 예측을 진행하였다. 분석 결과, GRU 모델이 LSTM 모델에 비해 예측력이 우수한 것으로 나타났으며, 이에 따라 RNN 기반의 인공지능망이 자연어뿐만 아니라 승객의 흐름 예측에도 적합한 것으로 분석하였다.

코로나19의 발생 이후로는 감염병이 교통량에 영향을 미침에 따라서 이와 관련된 변수가 추가되어 연구가 진행된 것으로 나타났다. Jiao et al.(2021)의 연구에서는 코로나19의 확산에 따른 베이징의 버스 승객 수를 예측하고 여러 모델 중, 예측도가 가장 정확한 모델을 규명하고자 하였다. 활용된 변수로는 버스 승객 수, 주말 여부, 날씨 및 일일 코로나19 데이터로 2019년 11월부터 2020년 8월까지 수집하였다. 이 중에서 날씨 변수로는 기온, 이슬점, 습도, 풍속, 기압과 강수량이 사용되었으며, 코로나19 변수로는 의심자 수, 완치자 수, 사망자 수와 확진자 수가 고려되었다. 예측에 활용된 모델은 총 7개로 Improved STL-LSTM(ISTL-LSTM), STL-LSTM, LSTM, GRU, LR(Linear Regression), KNR(K-Nearest Neighbor Regression)과 XGBoost(extreme Gradient Boosting)를 학습시켰다. 다양한 변수들을 통해 예측한 결과, 승객 수만 활용한 경우보다 기타 변수들이 있을 때 예측 성능이 더 우수한 것으로 확인되었다. 또한, 단일 모델보다 하이브리드 모델이 더 우수한 것으로 나타났으며, ISTL-LSTM이 가장 정확한 예측도를 나타내는 것으로 분석되었다.

예측의 정확도가 높은 모델을 규명하고자 하였던 앞선 연구와 달리, Fathi-Kazerooni et al.(2020)은 LSTM 모형을 활용하여 뉴욕시 지하철 이용자와 코로나19 사망자 수 및 확진자 수 사이의

상관관계를 파악하고, 나아가 ARIMA 모형으로 코로나19의 사망자 수와 확진자 수가 0에 가까워지는 날짜를 추정하고자 하였다. 또한, 코로나19로 인해 처음으로 사망 및 확진 사례가 발생한 날을 ARIMA를 통해 추정하여 실제 보고된 날짜와 비교하고자 하였다. 이를 위한 변수로는 2020년 3월부터 2020년 5월까지의 코로나19 사망자 및 확진자 수와 뉴욕시 지하철의 개찰구 데이터가 사용되었다. 분석 결과에 따르면 첫째, 코로나19 사망자와 승객 수 사이에 강한 상관관계가 있는 것으로 나타났다. 둘째로는 ARIMA를 통해 추정한 값의 경우, 처음으로 발생한 사망 및 확진 사례에 대하여 신뢰구간 95% 수준에서 실제 날짜와 가까운 값을 구하였다.

3. 연구의 차별성

코로나19의 확산이 대중교통에 미치는 영향과 관련된 연구는 미국과 유럽 등 세계적으로 많이 이루어지고 있다. 앞서 진행된 연구의 검토를 통하여 코로나19가 대중교통 이용량의 감소에 영향을 미치고 있는 것을 확인하였으며, 대부분의 연구는 코로나19 발병 이후의 데이터가 1년 이하의 단기간 자료를 활용한 것으로 나타났다. 또한, 대중교통 이용량 예측에는 LSTM, GRU, ARIMA 및 융합모형 등이 활용되었으며, 많은 연구에서 ARIMA와 LSTM 기반의 예측 모형이 우수한 것으로 분석하였다(표 1). 마지막으로 예측 모형의 변수로 대중교통 승객 수, 휴일

여부, 요일, 날씨, 코로나19 신규 확진자 수 및 사망자 수 등이 영향요인으로 사용된 것을 확인하였다. 따라서 본 연구에서는 예측력을 높이기 위해 코로나19 발생 이후 약 2년 동안의 데이터를 구득하여 시계열 예측에 주로 활용되는 모형을 선정하여 예측 후 비교 분석을 진행하고자 하였다. 더불어, 대중교통 이용량을 예측하는 관련 변수로 선행연구에서 사용된 요인과 코로나19의 확진자 수에 영향을 미치는 변수를 추가로 고려하고자 하였다.

III. 자료 및 연구 방법론

1. 분석 자료

본 연구에서는 코로나19 바이러스의 영향력 속에서 시도별 대중교통 이용량을 확인하기 위해 국내에서 코로나19 확진자가 처음 발생한 2020년 1월 20일부터 연구 시의 가장 최근 일자인 2021년 11월 5일까지 약 2년간 일별 시계열 656일 자료를 활용하였다. 이용된 데이터는 크게 대중교통의 이용량과 관련한 자료와 코로나19와 관련한 자료로 구성되었다. 먼저, 대중교통 이용량 자료를 구성하기 위하여 버스와 지하철을 포함한 대중교통의 목적통행량을 활용하였다. 앞서 검토한 연구결과에 근거하여 대중교통 이용량과 상관성이 높은 자료로 날씨 자료와 기상자료를 추가하였다. 두 번째로, 코로나19 자료를 구성하기 위하여 코로나19 확진자 수와 백신 접종자 수의 자료를 수집하였다. 구득된 변수는 총 5개로 일자별로 656개의 데이터이며, 백신 접종자 수의 자료는 252개로 <표 2>에서 확인할 수 있다.

대중교통 이용량 자료를 구축하기 위하여 교통카드빅데이터 통합정보시스템에서 제공하는 이용량 지표의 목적통행량 데이터를 이용하였다. 목적통행량 데이터는 시도별 일자별 발생량을 수집하였다. 날씨 자료를 통해서는 평일과 주말을 구분하여 주말을 휴일로 처리하였으며, 공휴일도 휴일로 처리하여 휴일 여부를 구분하였다. 기상자료는 기상청의 기상자료 개방포털에서 각 시도청이 있는 지역에서 관측한 기상 데이터 중 일일 강수량을 활용하였다.

코로나19 자료를 구축하기 위하여 질병관리청에서 제공하는 확진환자 발생현황 데이터와 코로나19 백신 일자별 접종 현황 데이터를 이용하였다. 이때, 코로나19와 관련한 자료의 공간적 범위는 전국으로 설정하였다. 확진환자 발생현황 데이터는 외국인을 포함한 국내에서 발생하는 일자별 신규 확진자 수 자료를 이용하였다. 백신 접종 현황 데이터는 접종 횟수에 대해, 2차 접종이 완료된 사람을 기준으로 설정하여 수집하였다. 또한, 백신 접종 현황 데이터의 시간적 범위는 처음 집계된 날의 자료가 한 번에 집계되어 수치가 이상치를 띠는 것을 고려하여 첫째 날의 데이터를 제외하고 2021년 2월 27일부터 2021년 11월 5일까지로 설정하였다.

표 1. 대중교통 이용량 예측 모형

Table 1. Proposed models for public transit usage

Studies	Models	Analyzed dates	COVID-19
Kim (2020)	Moving average method Exponential smoothing method ARIMA	Jan. 2007– Dec. 2019	X
Guo et al. (2019)	LSTM ARIMA SVR SVR-LSTM Fusion-KNN	Jan. 2017– Dec. 2017	X
Sha et al. (2020)	LSTM GRU BPN	Mar. 2016	X
Jiao et al. (2021)	LSTM GRU LR KNR SGBost STL-LSTM ISTL-LSTM	Nov. 2019– Aug. 2020	0
Fathi-Kazerooni et al. (2020)	LSTM ARIMA	Mar. 2020– May. 2020	0

표 2. 자료 설명 Table 2. Data source

Datasets		Data description	Source	Number of records
대중교통 이용량 Public transit usage		버스와 지하철을 포함한 대중교통 목적통행량의 시도별 일별 발생량 활용 Daily generation of public transit purpose traffic by city including bus and subway	교통카드빅데이터 통합정보시스템 (Stcis)	656
휴일 여부 Holiday		평일과 주말을 구분하여 주말과 공휴일을 휴일로 처리 Set holiday through the distinction of weekdays and weekends	-	656
일일 강수량 Daily precipitation		각 시도청이 있는 지역에서 관측한 기상 데이터 중 일일 강수량(mm) 활용 Precipitation data observed in each cities	기상자료개방 포털 (Weather Data Open Portal)	656
코로나19 COVID-19	코로나19 확진자 수 Cases of COVID-19	국내의 일자별 신규 확진자 수 활용 Number of newly confirmed cases of COVID-19 in Korea	질병관리청 (KDCA)	656
	백신 접종자 수 Number of people vaccinated	2차 접종 완료자 수를 기준으로 설정 People who have been fully vaccinated (2 nd inoculation)		252

2. 연구 방법

본 연구는 대중교통 이용량 예측을 하기 위해서 네 가지 모델을 활용하여 예측을 진행하였다. 딥러닝 예측 모형은 순환신경망 모델인 RNN과 LSTM 두 가지로 구축하였으며, 추가로 기존의 연구에서 많이 사용된 시계열 분석 모형인 ARIMA와 기본 통계 모형인 HA를 활용하여 비교하였다.

1) 딥러닝 모형 설계 방향

본 연구에서 이용되는 자료와 같이 시계열로 이루어진 연속적인 데이터를 분석하는 경우 대표적으로 많이 사용되는 딥러닝 예측 모형으로 Rumelhart et al.(1986)이 제안한 순환신경망(RNN, Recurrent Neural Network) 모형이 있다. RNN은 반복적이고 순차적인 데이터 학습에 특화된 인공신경망의 한 종류로 과거의 학습 결과를 현재 학습에 사용하는 딥러닝 네트워크이다. <그림 2>와 같은 순환구조를 띠며 은닉층(Hidden Layer)에 이전 정보를 기억시켜 학습을 진행하므로 시계열 데이터에 적합한 모형이다(Joo and Choi, 2018).

수학적 구조는 <그림 3>과 같으며 입력층의 변수는 대중교통 이용량에 영향을 미치는 변수인 대중교통 이용량, 휴일 여부, 일일 강수량, 코로나19 확진자 수 및 백신 접종자 수로 설정하여 은

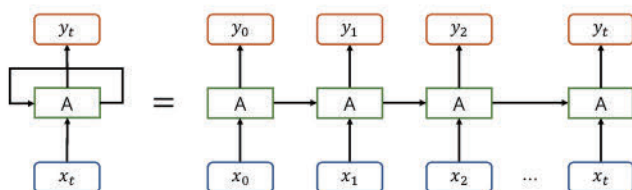


그림 2. RNN 모형 구조
Figure 2. Structure of RNN

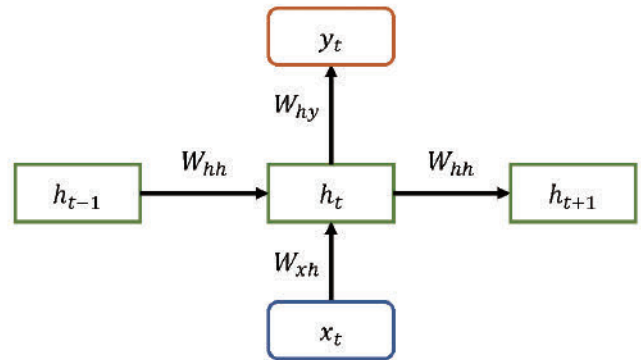


그림 3. RNN의 수학적 구조
Figure 3. Mathematical structure of RNN

닉층으로 전달된다. 이에 따라 은닉층에서 연산 값은 <식 (1)>과 같이, 출력값은 <식 (2)>와 같이 표현되며, 변수는 대중교통 이용량으로 설정하였다. 또한, ϕ 는 은닉층의 활성화 함수를 나타내며 본 연구에서는 ReLU(Rectifier Linear Unit) 함수를 사용하였다. 또한, 일반적으로 LSTM과 GRU 모델이 시계열 예측에 뛰어나지만, 단기 예측에 효과적인 순환신경망 모델은 RNN이다(김경환, 2021). 이에 따라 RNN의 예측 성능이 우수한 경우 단기적 변동성이 높음에 따라 장기적 정보에 대한 부분이 예측 성능 향상에 큰 기여를 하지 않는 것을 확인할 수 있다(김경환, 2021). 다만, 순환구조가 길어지면 학습 과정 중의 기울기 소실 문제로 인한 장기 의존성(Long Term Dependency) 문제가 있다.

$$h_t = \phi_h(W_{hh}h_{t-1} + W_{xh}x_t + b_h) \tag{1}$$

$$y_t = W_{hy}h_t + b_y \tag{2}$$

h_t : t 시점에서의 은닉층

- y_t : t 시점에서의 출력층
- x_t : t 시점에서의 입력층
- ϕ_h : 은닉층의 활성화 함수
- W_{xh} : 입력층에서 은닉층으로의 가중치 행렬
- W_{hy} : 은닉층에서 출력층으로의 가중치 행렬
- b_h, b_y : 은닉층, 출력층의 편향 벡터

LSTM(Long-Short Term Memory)은 RNN의 장기 의존성 문제를 보완하기 위해 RNN을 기반으로 Hochreiter and Schmidhuber(1997)가 제안한 모형이다. <그림 4>와 같이 RNN의 은닉층에 cell-state를 추가한 구조로 시간에 따라 정보를 계속해서 공급할 수 있다. LSTM은 기존의 RNN과 같이 순환 구조를 가지고 있으나, 망각 게이트(Forget Gate), 입력 게이트(Input Gate)와 출력 게이트(Output Gate)를 이용하여 장기 기억 정보를 저장하는 메모리셀 구조를 갖는다. 이때 대중교통 이용량, 휴일 여부, 일일 강수량, 코로나19 확진자 수 및 백신 접종자 수로 구성된 x 변수는 이와 같은 메모리셀 C 로 전달된다.

먼저, 망각 게이트는 과거 정보를 잊기 위한 게이트로 h_{t-1} 과 x_t 에 시그모이드(σ)를 취해준 값이다. 시그모이드(σ) 함수의 출력 범위는 0에서 1 사이이기 때문에 그 값이 0에 가까우면 과거 이전 상태의 정보는 잊고, 1에 가까울수록 이전 상태의 정보를 많이 기억하게 된다. 이러한 망각 게이트 f_t 는 <식 (3)>과 같이 표현된다. 다음으로 입력 게이트는 <식 (4)>와 <식 (5)>와 같이 나타나며 현재 정보를 기억하기 위한 게이트이다. h_{t-1} 과 x_t 에 시그모이드(σ) 함수를 취하고 동일한 값으로 하이퍼볼릭탄젠트(\tanh)를 취해준 다음 <식 (6)>과 같이 각 행렬 인자 간의 연산(*, 곱셈)을 한 값이다. 마지막으로 출력 게이트는 업데이트된 셀 C_t 를 다음 은닉층에 얼마나 전달할지 정하는 게이트로 <식 (7)>, <식 (8)>과 같이 확인할 수 있다.

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \tag{3}$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \tag{4}$$

$$\tilde{C}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{5}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{6}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{7}$$

$$h_t = o_t * \tanh(C_t) \tag{8}$$

2) 통계 기반 모형 설계 방향

ARIMA(Autoregressive Integrated Moving Average, 자동회귀 누적이동평균)는 시계열 예측에 많이 사용되는 통계적 모형인 AR(Auto-Regressive, 자기회귀) 모형과 MA(Moving Average, 이동평균) 모형을 결합시킨 ARMA(Autoregressive Moving Average, 자기회귀 이동평균) 모형을 일반화한 모형이다. 단일 변량 시계열 모형으로, 현재의 시계열 값을 과거 관측값과 예측 오차를 통해 설명한다. 시계열 데이터의 분석 시, 비정상성의 성질을 갖는 경우 잘못된 결과를 나타내게 되어 분석을 진행하기 전에 차분연산기법(differencing, 차분)을 사용하여 정상성 시계열 데이터로 변환해야 한다. 앞선 AR, MA와 ARMA 모형들은 정상성 시계열 데이터에만 적용이 가능하며, 비정상성 데이터를 정상성 데이터로 만들기 위해 차분한 것이 ARIMA 모형으로, ARIMA(p, d, q)로 나타낸다. 이때 p, d와 q는 각각 AR 모형, 차분 그리고 MA 모형의 차수를 의미한다. 전체 식을 도출하기에 앞서, 시계열 데이터 Y_t 를 차분하여 정상성 시계열을 얻었으며, <식 (9)>와 같이 w_t 로 표현하였다. 이에 따라 얻어지는 ARIMA(p, d, q)는 <식 (10)>과 같다.

$$w_t = (1 - B)^d Y_t \tag{9}$$

w_t : 추세가 제거된 정상성 시계열

Y_t : 시계열 데이터

B : 후진연산자

d : 비정상 시계열 차분 차수

t : 시점

$$\phi_p(B)w_t = \theta_0 + \theta_q(B)\epsilon_t \tag{10-1}$$

$$\phi_p(B)(1 - B)^d Y_t = \theta_0 + \theta_q(B)\epsilon_t \tag{10-2}$$

ϕ_p : 자기회귀 모수

θ_0 : 추세 모수

θ_q : 이동평균 모수

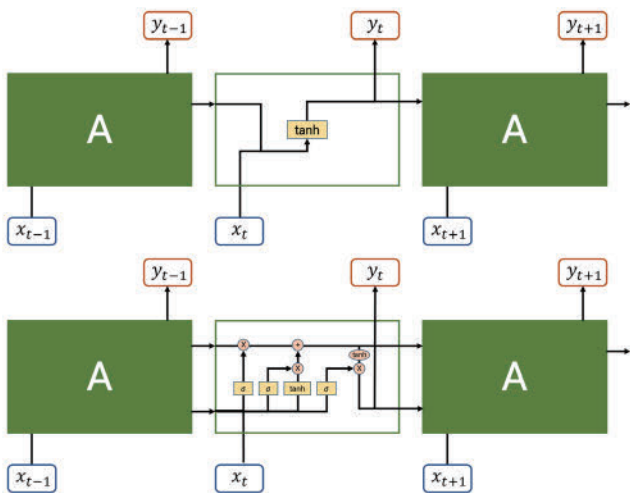


그림 4. RNN과 LSTM의 구조
Figure 4. Structure of RNN and LSTM

p : 자기회귀 시차
 q : 이동평균 시차
 ϵ_t : 오차

HA(Historical Average) 모형은 기본적인 통계 모형으로 <식 (11)>과 같이 설정한 과거 특정 기간 동안의 평균을 통해 현재의 값을 예측하는 모형이다.

$$x_t = \frac{1}{n} \sum_{i=1}^n x_i \quad (11)$$

3) 예측 모형 구축

본 연구는 코로나19 확산 이후의 대중교통 이용량 예측을 위해 5개 변수인 대중교통 이용량, 휴일 여부, 일일 강수량, 코로나19 확진자 수와 백신 접종자 수를 활용하여, 전체 656개의 데이터로 시계열 예측을 진행하였다. 총 데이터 중 2020년 1월 20일부터 2021년 6월 5일까지 503개의 데이터를 학습(training) 세트로 사용하였으며, 예측의 비교 평가를 위해 2021년 6월 6일부터 2021년 11월 5일까지 다섯 달간 데이터를 예측하여 153개를 테스트(test) 세트로 이용하였다. 각각의 모형들은 전국 대중교통 목적통행량을 기준으로 모형을 최적화하였으며, 구축된 모형을 바탕으로 시도별로 예측을 진행하였다.

딥러닝 모형인 RNN과 LSTM은 대중교통 이용량, 휴일 여부, 일일 강수량, 코로나19 확진자 수, 백신 접종자 수를 설명변수로, 대중교통 이용량을 종속변수로 구성하였다. 전체적인 학습변수는 <표 3>과 같이 설정하였다. 더불어, 모형의 구조 변수로는 hidden layer의 unit, hidden layer의 개수와 예측 일자(하루를 예측하기 위해 이전 며칠을 보았는지)를 두어 가장 적합한 모형을 판단하였다. 이에 따라, <표 4>와 같이 각각 128, 1 그리고 14인 경우에 가장 적합한 것으로 나타났으며, 이때 예측 일자가 14일이 적합하게 나온 것은 사회적 거리두기 개편안 발표일이 14일 간격이기 때문일 것으로 사료된다.

통계 기반의 시계열 예측 모델인 ARIMA 모형은 단일 변량 예측 방법을 사용하므로 독립변수와 종속변수가 동일하며, 변수로 대중교통 이용량 데이터를 활용하였다. 분석에 앞서 먼저 활용할

표 3. 모형의 학습 변수 Table 3. Parameters

No.	Parameters	Values
1	Epochs	200
2	Batch size	256
3	Activation function	ReLU
4	Loss function	Mean Absolute Error
5	Optimizer	Adam (Kingma and Ba, 2015)

표 4. 모형의 구조 변수 Table 4. Structural parameters

Units of hidden layer	No. of hidden layers	Predict date	RMSE (training set)
64	1	7	2,302,647
	2		2,073,687
	3		2,509,341
128	1	7	2,202,492
	2		2,246,651
	3		1,977,837
256	1	7	2,072,973
	2		2,300,521
	3		-
64	1	14	1,928,075
	2		2,006,157
	3		2,247,852
128	1	14	1,766,636
	2		2,060,334
	3		1,856,122
256	1	14	1,916,374
	2		1,801,731
	3		-
64	1	21	1,867,829
	2		1,972,060
	3		2,250,868
128	1	21	1,920,272
	2		2,027,335
	3		1,897,058
256	1	21	2,075,083
	2		2,249,535
	3		-
64	1	28	1,827,710
	2		1,950,393
	3		2,183,763
128	1	28	2,359,288
	2		1,987,228
	3		2,563,254
256	1	28	2,184,726
	2		-
	3		-

자료의 정상성을 ADF 검정(Augmented Dickey-Fuller test)을 통해 확인하였다. 자동 탐색을 통해 확인한 모형의 p, d, q 파라미터 값은 <표 5>와 같이 (1, 1, 0)으로 설정되었다.

HA 모형은 ARIMA 모형과 마찬가지로 단변량 예측 방법을

표 5. ARIMA 차수 값 Table 5. Parameters of ARIMA

No.	Parameters	Values
1	(p, d, q)	(1, 1, 0)

활용하므로 독립변수와 종속변수는 동일하게 구축된다. 대중교통 이용량 데이터를 활용하였으며 503개의 학습(train) 데이터 셋으로 전체 평균을 내어 예측 후, 153개의 테스트(test) 데이터 셋과 비교하여 모델의 예측 성능을 평가하였다.

IV. 예측 모형 분석 결과

1. 모델 성능 평가 지표

본 연구에서는 시계열 자료를 이용하여 모형별로 대중교통 이용량 예측을 진행하였다. 예측을 진행하기에 앞서, 예측 모형에는 예측 오차가 발생하며, 이를 통해 모형의 신뢰성 및 안정성을 평가할 수 있다. 이에 따라 예측 모형의 성능을 평가하고 비교하기 위하여 평가 지표를 활용하였으며, RMSE(Root Mean Square Error, 평균 제곱근 오차)와 MAPE(Mean Absolute Percentage Error, 평균 절대 백분율 오차)의 두 가지를 선정하였다.

RMSE는 모델이 예측한 값과 실제 값의 차이를 다루는 예측 모형의 평가 지표 중 하나이다. 관측에서 나타나는 오차인 잔차의 제곱합을 산술평균한 값의 제곱근으로서 관측값 간의 상호 간 편차를 의미하며, <식 (12)>와 같이 표현된다. 다만 RMSE의 경우, 실제 값의 크기에 의존적인 특징이 있어 크기 의존적 에러가 발생하고 이로 인해 예측값이 예측하는 대상의 크기에 영향을 받아 스케일의 차이가 큰 대상 간의 해석이 어려울 수 있다(Willmott and Matsuura, 2005). 이처럼 예측의 대상이 여러 개인 경우 MAPE를 추가로 이용하여 비교하였다. MAPE는 <식 (13)>과 같이 표현되며, 회귀모델의 성능 평가 지표 중 하나이다. 실제 값 대비 예측값을 실제 값으로 나눈 절댓값의 평균에 백분율을 통해 구하며, 퍼센트 값을 갖는다.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \tag{12}$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \tag{13}$$

n: 데이터 수

Y_i : 실제값

\hat{Y}_i : 예측값

RMSE와 MAPE 모두 값이 낮을수록 모형의 예측력이 우수한 것으로 판단할 수 있다.

2. 예측 모형 비교 분석 결과

선정된 네 가지 모형을 활용하여 2020년과 2021년 약 2년 동안의 일별 대중교통 이용량을 예측하고, 실제 이용량과 예측된 이용량으로 RMSE와 MAPE를 계산하여 모형 간 비교 평가를 진행하였다. 대중교통 이용량 및 코로나19와 관련된 변수를 사용하였으며, 먼저 전국의 데이터를 기준으로 하여 모델을 설계하고 적합한 모형을 도출한 후, 채택된 모형을 기준으로 시도별로 대중교통 이용량을 예측하여 비교 분석하였다.

1) 전국

시도별 대중교통 이용량을 예측하기에 앞서 전국 대중교통 목적통행량을 기준으로 최적의 모형을 선정하였다. 딥러닝의 경우, 다양한 파라미터를 시도하여 RMSE가 가장 낮은 값의 모형을 채택하였다. 최적화한 모형에 따른 예측 후 성능 평가 지표로 RMSE를 이용하였으며, 각 모형은 <표 6>과 같은 값을 갖는 것으로 확인되었다. 예측 결과, 네 가지 모형 중에서 RNN이 1,837,403로 가장 낮은 값으로 나타나 예측력이 가장 우수한 것으로 나타났고, <그림 5>와 같이 확인할 수 있다. 한편, 김경환(2021)에 의하면 RNN은 순환신경망 모델 중 단기 예측에 효과적인 모델이며, 이에 따라 전국을 기준으로 예측한 모형 중 RNN

표 6. 전국 모형별 테스트 세트의 RMSE

Table 6. RMSE of test set prediction

Category	RNN	LSTM	ARIMA	HA
전국 Nationwide	1,837,403	2,021,464	5,089,027	3,498,478

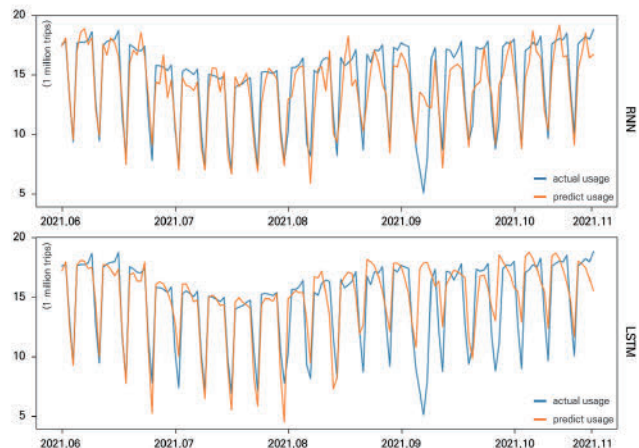


그림 5. 모형별 전국 대중교통 예측 및 실제 이용량
Figure 5. Actual and predict public transportation usage of nationwide

모형이 가장 적합하게 분석된 것은 코로나19와 같은 단기적인 영향력이 큰 상황이 예측력에 영향을 미치는 것으로 판단된다.

2) 시도

전국을 기준으로 최적화된 모형 구조를 바탕으로 동일하게 구축하여 시도별로 예측하였다. 딥러닝인 RNN과 LSTM에 대해서는 같은 파라미터의 모형 구조를 사용하였다. 각 시도별로 예측을 진행하므로 대상이 여러 개이기 때문에 평가 지표로 RMSE와 MAPE를 추가로 확인하였다. 각 모형의 예측력은 <표 7>과 <표 8>처럼 나타났으며, 대부분의 시도에서 RNN, LSTM, ARIMA 및 HA 중 RNN의 예측력이 가장 우수한 것으로 드러났다. 이는 코로나19와 같은 상황이 사람들의 인식에 단기적인 영향을 미치기

때문에 단기 예측에 효과적인 RNN 모형이 적합한 것으로 나타났다. 반면, 17개의 시도 중 14개 시도의 RMSE와 MAPE의 값이 RNN에서 가장 낮은 값을 가지지만, 수도권인 서울특별시, 인천광역시, 경기도의 세 지역에서는 LSTM이 평가 지표에서 각각 790,074, 104,015, 432,261과 9.52, 8.87, 9.15의 값을 가지며 가장 우수한 것으로 나타났다. 이는 수도권 외의 지역들에 비해 코로나19의 영향력이 적기 때문에 나타나는 현상으로 해석할 수 있다. 코로나19의 확산에 대한 영향력이 비교적 적은 이유는 수도권 지역의 경우 비수도권 지역 대비 출퇴근과 같은 필수적인 통행이 많이 이루어져 대중교통 이용량에 대해 코로나19의 영향이 비교적 낮기 때문으로 판단된다. 이와 같이 비수도권과 수도권 지역에서의 RNN과 LSTM의 예측력에 대해 실제

표 7. 시도별 모형별 테스트 세트의 RMSE

Table 7. RMSE test set by cities

Category	RNN	LSTM	ARIMA	HA
서울특별시 Seoul	873,973	790,074	2,334,249	1,546,998
부산광역시 Busan	223,793	242,882	484,943	325,677
대구광역시 Daegu	86,823	112,524	264,449	144,767
인천광역시 Incheon	109,239	104,015	286,545	202,415
광주광역시 Gwangju	36,329	43,517	89,241	62,908
대전광역시 Daejeon	48,217	59,027	106,729	72,188
울산광역시 Ulsan	28,047	32,182	66,080	42,517
세종특별자치시 Sejong	5,291	6,866	16,027	10,518
경기도 Gyeonggido	459,813	432,261	1,109,907	780,451
강원도 Gangwondo	11,793	12,959	29,956	18,879
충청북도 Chungcheongbukdo	19,234	24,958	41,589	30,330
충청남도 Chungcheongnamdo	30,649	37,355	72,063	51,937
전라북도 Jeollabukdo	21,250	28,916	54,655	38,752
전라남도 Jeollanamdo	23,028	28,634	46,902	34,715
경상북도 Gyeongsangbukdo	26,242	28,905	67,982	41,536
경상남도 Gyeongsangnamdo	62,561	75,953	136,819	95,138
제주특별자치도 Jeju	18,526	19,557	37,580	26,483

표 8. 시도별 모형별 테스트 세트의 MAPE

Table 8. MAPE test set by cities

Category	RNN	LSTM	ARIMA	HA
서울특별시 Seoul	11.33	9.52	37.44	27.04
부산광역시 Busan	11.91	12.77	30.22	21.98
대구광역시 Daegu	11.13	13.03	36.94	24.07
인천광역시 Incheon	10.80	8.87	32.33	24.20
광주광역시 Gwangju	12.66	16.57	37.65	29.17
대전광역시 Daejeon	13.30	15.03	35.07	25.79
울산광역시 Ulsan	14.10	18.11	37.68	27.87
세종특별자치시 Sejong	11.55	16.18	40.37	31.39
경기도 Gyeonggido	10.98	9.15	32.78	24.54
강원도 Gangwondo	13.43	19.25	41.26	27.85
충청북도 Chungcheongbukdo	12.89	15.59	33.38	25.89
충청남도 Chungcheongnamdo	11.18	14.94	31.61	23.99
전라북도 Jeollabukdo	12.31	15.15	39.30	31.06
전라남도 Jeollanamdo	13.51	14.87	33.45	26.61
경상북도 Gyeongsangbukdo	12.14	14.55	36.66	24.68
경상남도 Gyeongsangnamdo	13.05	17.12	33.80	25.55
제주특별자치도 Jeju	12.67	12.69	28.61	21.71

이용량과 예측 이용량은 <그림 6>, <그림 7>과 같이 나타난다.

3) 소결

대중교통 이용량의 예측을 네 가지 모형을 통해 진행하였으며, 가장 먼저 전국을 기준으로 모델을 설계하였다. RNN 모형의 RMSE 값이 가장 낮아 최적 모형으로 분석되었으며, 이때 도출된 모형을 기준으로 하여 시도별로 대중교통 이용량을 예측하여 비교 분석하였다. 시도별 예측 결과, 서울, 인천, 경기도의 수도권 지역에서는 RNN보다 LSTM의 예측력이 우수하였으며, 비수도권 지역에서는 RNN의 예측력이 우수하였다. 이처럼 수도권 지역과 비수도권 지역에서 예측력이 우수한 모델이 다르게 나타난 것은 코로나19에 의한 영향력의 정도가 다르기 때문으로 사료된다.

수도권의 경우, 서울시를 예로 들어 2019년 기준 대중교통의 수단분담률이 65%에 달하는 것과 같이(서울특별시, 2022), 대중교통 수요가 높아서 단기변동에 대한 영향이 상대적으로 적어 LSTM의 예측력이 우수한 것으로 생각된다. 비수도권의 경우에

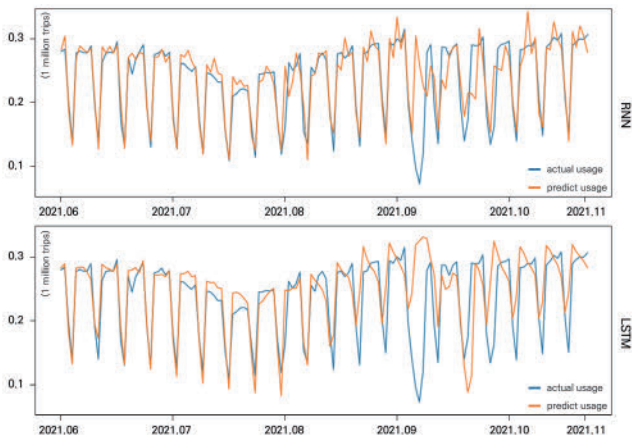


그림 6. 모형별 광주광역시 대중교통 예측 및 실제 이용량

Figure 6. Actual and predict public transportation usage of Gwangju

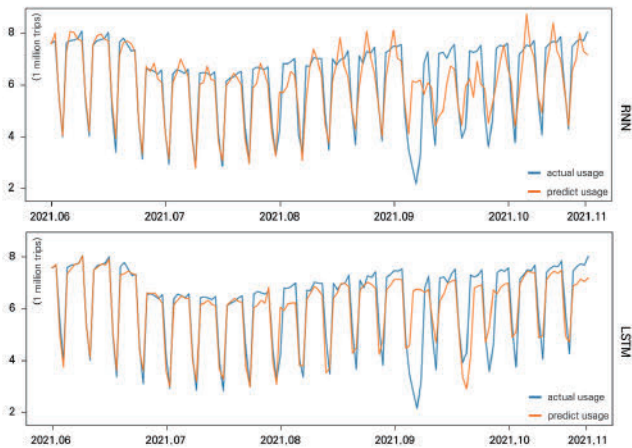


그림 7. 모형별 서울특별시 대중교통 예측 및 실제 이용량

Figure 7. Actual and predict public transportation usage of Seoul

는 상대적으로 코로나19의 확산에 민감하게 반응하며, 이에 따른 대중교통의 영향이 크기 때문에 단기적인 영향력이 있는 상황에서 예측력이 뛰어난 RNN이 두드러진 것으로 판단된다.

V. 결론

코로나19의 발생과 확산은 팬데믹 상황까지 이르렀으며 그로 인한 WHO의 지침은 전 세계에서 감염 예방을 위해 관련한 법적 지침과 제한을 내리게 하였다. 이는 우리의 생활 방식에 많은 변화를 가져왔으며, 교통 패러다임에 즉각적인 변화를 나타내었다. 감염 확산의 위험에 따라 정부는 ‘사회적 거리두기’ 지침을 시행하였으며, 대면 활동이 감소하고 개인의 교통수단 선택에도 영향을 미치는 것으로 나타났다. 이에 따른 통행 패턴은 버스, 지하철과 같은 대중교통 이용량이 감소하고 접촉이 적은 개인형 교통수단의 이용이 증가하는 것으로 확인되었다. 이와 같은 대중교통 이용객의 감소는 대중교통 운전자 입장에서는 큰 손해이며, 승객 역시도 감염의 위험을 감수하고 이동해야 하는 불안함이 존재한다. 이러한 통행의 변화에 따라 코로나19 이후 이용객의 심리적 위축에 따른 대중교통 이용량에 대한 흐름의 파악이 필요할 것으로 예상되며, 이를 예측하는 데 적합한 모형을 알아보고자 하였다.

본 연구는 국내의 시도를 대상으로 코로나19가 발생한 이후 대중교통 이용량 예측에 적합한 모형에 대해 알아보고 비교 분석하였다. 2020년 1월 20일부터 2021년 11월 5일까지 약 2년 동안 코로나19와 대중교통 이용량에 영향을 미치는 변수들을 일자별로 자료를 수집하였으며, 시계열 예측에 적합한 네 가지 모형들을 선정하여 예측하였다. RNN, LSTM, ARIMA, HA 모형을 사용하였으며, 모형의 설명변수로는 대중교통 이용량과 휴일 여부, 일일 강수량, 코로나19 확진자 수, 백신 접종자 수를 활용하였다. 모형의 예측력을 평가하는 평가 지표로 RMSE와 MAPE를 이용하였으며 가장 낮은 값을 갖는 모형이 예측력이 가장 우수한 것으로 확인하였다. 분석 결과, 수도권을 제외한 모든 지역에서 RNN의 평가 지표 값이 가장 낮게 나타났으며, 수도권은 평가 지표가 가장 낮은 모형이 LSTM으로 확인되었다. 이와 같은 결과를 토대로 다음과 같은 시사점을 도출하였다.

먼저, 딥러닝 모형의 설계과정에서 이전의 14일로 하루를 예측하는 경우가 가장 우수하게 나타났다. 이는 사회적 거리두기 개편안 발표 일자가 14일임에 따라 나타나는 것으로 사료된다. 이를 통해, 대중교통의 이용에 사회적 거리두기와 같은 정부의 지침이 사람들의 인식에 영향을 미치는 것을 알 수 있다. 따라서 코로나19 상황에서 대중교통 이용량 예측을 진행할 경우, 코로나19 확산 상황에 따른 정부의 지침(사회적 거리두기 등)에 대한 고려가 필요할 것으로 판단된다.

두 번째, 수도권을 제외한 모든 시도에서 RNN의 예측력이 가장 우수한 것으로 분석되었다. 앞선 연구들을 검토한 결과, <표 1>

에서 확인할 수 있는 바와 같이 코로나19 이전까지 진행된 연구들에서는 대부분 ARIMA와 LSTM 기반의 모델이 대중교통 이용량 예측에 가장 적합한 것으로 분석되었으며, 그 이후에도 계속 ARIMA 모형과 LSTM 모형을 활용한 경우가 예측력이 가장 뛰어난 것으로 확인되었다. 그러나 이처럼 일반적으로는 대중교통 이용량 예측에 딥러닝 모형 중 LSTM의 예측 성능이 우수하지만, 본 연구 결과에서는 RNN의 예측 성능이 더욱 우수한 것으로 분석되었다.

이와 같은 결과는 김경환(2021)에 의하면 일반적으로 RNN이 단기 예측에 비교적 효과적인 순환신경망 모델로 나타나며, 코로나19의 경우 사회적 거리두기가 14일 단위로 변동을 하는 것과 같이 단기적인 영향을 미치기 때문에 단기 예측에 효과적인 RNN 모형이 적합한 것으로 판단된다. 수도권을 제외한 모든 지역에서 대중교통 이용량 예측에 많이 사용된 LSTM이 아니라 RNN이 코로나19의 상황에서 예측이 적합한 것으로 보아, 향후 코로나19와 같은 단기적인 영향이 큰 팬데믹 상황에서 대중교통 이용량 예측은 RNN 모형을 통한 예측이 적합할 것으로 보인다. 한편, 교통부문의 운영자는 교통수단을 운영할 때, 적은 비용으로 최대의 효율을 내고자 할 것이며 이는 시간대에 따른 배차 간격이 매우 중요하게 작용할 것이다. 또한, 차량의 혼잡도 역시 이용량 예측을 기반으로 하여 이용자에게 쾌적한 이용 경험을 가져다줄 수 있으므로 보다 정확한 대중교통 이용량 예측은 중요할 것으로 판단된다. 이에 따라 팬데믹 상황에서의 대중교통 단기 이용량 예측을 활용하여 운행 스케줄링과 배차 간격 조정 등 효율적인 대중교통 운영계획 수립과 운영자의 의사결정에 기여할 것으로 기대된다.

세 번째, 수도권 지역인 서울특별시, 인천광역시와 경기도는 LSTM의 예측 성능이 우수하게 나타났다. 수도권에서 RNN의 예측력이 낮은 것은 출퇴근 등의 필수적인 통행량이 많아 대중교통 이용량에 대해 코로나19에 의한 단기적인 변동이 비교적 작기 때문으로 생각된다. 이는 대중교통 이용량과 통행의 목적이 밀접한 관계에 있음을 시사하며, 이러한 관계는 향후 배차 간격 조정과 직원 일정 관리 등의 대중교통 운행과 관련하여 운송회사 운영자의 의사결정에 고려되어야 할 점으로 보인다.

한편, 본 연구는 대중교통 이용량 예측을 진행하고 적합한 모형을 규명하는 과정에서 다음과 같은 한계점을 가진다. 첫째, 시설별 운영시간의 변화와 사적모임 인원 제한과 같은 사회적 거리두기의 단계 조정이 고려되지 않았다. 이와 같은 질병으로 인한 사회적인 변화는 향후 연구에서 고려되어야 할 것이다.

둘째, 기상자료들은 서로에게 밀접한 관련이 있어서 다중공선성의 문제가 발생하여 본 연구에서는 기상자료 중 강수량만 분석 자료로 활용되었다. 기상상태는 대중교통의 이용에 많은 영향을 미치는데, 이는 향후 연구에 다른 기상적인 요인이 추가로 고려되어야 할 것으로 보인다.

셋째, 본 연구의 백신 접종자 수 자료에서 수집 기간 당시에 백신 공급이 부족하여 2021년 중반을 기점으로 60대 이상 등의 위험군이 우선 접종대상자인 시기였으며, 2021년 8월 이후 11월까지 50대 이상의 2차 접종 시기였던 점을 감안하면, 실제 획득한 자료의 기간에서는 대중교통 이용 대상군의 2차 접종 완료자 수가 미흡했다는 한계가 존재한다.

넷째, 대중교통의 목적통행량으로 버스와 지하철의 이용량이 합쳐져 고려되었다. 전체 대중교통의 이용량 흐름을 파악하기 위해 두 수단이 함께 고려되었지만, 각 수단은 이용 패턴과 혼잡도 등이 다른 양상을 보이므로 이후 연구에서는 보다 면밀한 예측을 위해 구분되어야 할 것으로 생각된다.

다섯째, 코로나19 바이러스만 고려되었기 때문에 다른 질병들의 상황에서도 RNN의 예측력이 우수한지 비교가 어렵다. 이에 따라, 향후 연구에서는 이를 고려하고 추가로 코로나19와 같은 기타 전염병들의 데이터를 활용하여 예측력을 비교 검증하는 연구가 필요할 것으로 판단된다.

인용문헌 References

- 김경환, 2021. "순환신경망을 활용한 bunker가 예측에 관한 연구", 「한국컴퓨터정보학회논문지」, 26(10): 179-184.
Kim, K.H., 2021. "A Study on the Forecasting of Bunker Price Using Recurrent Neural Network", *Journal of the Korea Society of Computer and Information*, 26(10): 179-184.
- 김동규, 2020. "시계열 자료를 활용한 도시철도 수요 예측", 「한국자료분석학회지」, 22(2): 753-765.
Kim, D.K., 2020. "Urban Railway Demand Forecast using Time Series Data", *Journal of The Korean Data Analysis Society*, 22(2): 753-765.
- 배영임·신혜리, 2020. "코로나19, 언택트 사회를 가속화하다", 「이슈 & 진단」, 1-26.
Bae, Y.I. and Shin, H.R., 2020. "COVID-19, Accelerate the Untact Society", *Issue & Analysis*, 1-26.
- 신송현·추상호, 2022. "코로나19로 인한 공유 모빌리티 이용 변화에 미치는 영향요인에 관한 연구: 신도시 주민을 대상으로", 「대한교통학회지」, 40(2): 230-244.
Shin, S.H. and Choo, S.H., 2022. "Exploring the Influencing Factors on Change in Use of Sharing Mobility by the COVID-19 Pandemic: Focused on Residents in New Towns", *Journal of Korean Society of Transportation*, 40(2): 230-244.
- 이수진·채수성·이창희·안수연·장기태, 2020. "코로나바이러스 감염증-19로 인한 통행 변화 분석: 대전광역시를 중심으로", 「교통기술과정책」, 17(4): 17-27.
Lee, S.J., Chae, S.S., Lee, C.H., An, S.Y., and Jang, G.T., 2020. "Effects of COVID-19 on Trip Patterns in Daejeon Metropolitan City, South Korea", *Transportation Technology and Policy*, 17(4):

- 17-27.
6. 장동익·임서현·성낙문, 2020. “코로나19의 육상교통부문 영향 분석 및 대응방향”, 『월간교통』, 265: 6-16.
Jang, D.I., Lim, S.H., and Seong, N.M., 2020. “Analysis of the Impact of COVID-19 in the Land Transportation Sector and Directions for Response”, *Monthly KOTI Magazine on Transport*, 265: 6-16.
 7. 조혜림·윤성범·정영제, 2020. “코로나19 로 인한 서울 통행변화 분석 및 미래 대응방안”, 『교통기술과정책』, 17(3): 46-51.
Cho, H.R., Yun, S.B., and Jeong, Y.J., 2020. “Seoul Transportation Changes and Strategies after COVID-19”, *Transportation Technology and Policy*, 17(3): 46-51.
 8. Anke, J., Francke, A., Schaefer, L.M., and Petzoldt, T., 2021. “Impact of SARS-CoV-2 on the Mobility Behaviour in Germany”, *European Transport Research Review*, 13(1): 1-13.
 9. Fathi-Kazerooni, S., Rojas-Cessa, R., Dong, Z., and Umpai-chitra, V., 2020. “Time Series Analysis and Correlation of Subway Turnstile Usage and Covid-19 Prevalence in New York City”, *arXiv*, 2008.08156.
 10. Guo, J., Xie, Z., Qin, Y., Jia, L., and Wang, Y., 2019. “Short-term Abnormal Passenger Flow Prediction Based on the Fusion of SVR and LSTM”, *IEEE Access*, 7: 42946-42955.
 11. Hochreiter, S. and Schmidhuber, J., 1997. “Long Short-term Memory”, *Neural Computation*, 9(8): 1735-1780.
 12. Jenelius, E. and Cebecauer, M., 2020. “Impacts of COVID-19 on Public Transport Ridership in Sweden: Analysis of Ticket Validations, Sales and Passenger Counts”, *Transportation Research Interdisciplinary Perspectives*, 8: 100242.
 13. Jiao, F., Huang, L., Song, R., and Huang, H., 2021. “An Improved STL-LSTM Model for Daily Bus Passenger Flow Prediction during the COVID-19 Pandemic”, *Sensors*, 21(17): 5950.
 14. Joo, I.T. and Choi, S.H., 2018. “Stock Prediction Model Based on Bidirectional LSTM Recurrent Neural Network”, *The Journal of Korea Institute of Information, Electronics, and Communication Technology*, 11(2): 204-208.
 15. Kingma, D.P. and Ba, J., 2015. “Adam: A Method for Stochastic Optimization”, Paper presented at 3rd International Conference for Learning Representations, USA: San Diego.
 16. OECD, 2020. “COVID-19 Transport Brief: Re-spacing Our Cities for Resilience”.
 17. Przybylowski, A., Stelmak, S., and Suchanek, M., 2021. “Mobility Behaviour in View of the Impact of the COVID-19 Pandemic—Public Transport Users in Gdansk Case Study”, *Sustainability*, 13(1): 364.
 18. Rumelhart, D.E., Hinton, G.E., and Williams, R.J., 1986. “Learning Representations by Back-propagating Errors”, *Nature*, 323(6088): 533-536.
 19. Sha, S., Li, J., Zhang, K., Yang, Z., Wei, Z., Li, X., and Zhu, X., 2020. “RNN-based Subway Passenger Flow Rolling Prediction”, *IEEE Access*, 8: 15232-15240.
 20. Teixeira, J.F. and Lopes, M., 2020. “The Link between Bike Sharing and Subway Use during the COVID-19 Pandemic: The Case-study of New York's Citi Bike”, *Transportation Research Interdisciplinary Perspectives*, 6: 100166.
 21. Tirachini, A. and Cats, O., 2020. “COVID-19 and Public Transportation: Current Assessment, Prospects, and Research Needs”, *Journal of Public Transportation*, 22(1): 1-21.
 22. Willmott, C.J. and Matsuura, K., 2005. “Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance”, *Climate Research*, 30(1): 79-82.
 23. 서울특별시, 2022.12.19. “주요교통통계”, <https://news.seoul.go.kr/traffic/archives/289>
Seoul Metropolitan Government, 2022, December 19. “Key Traffic Statistics”, <https://news.seoul.go.kr/traffic/archives/289>
 24. 교통카드빅데이터 통합정보시스템, “이용량 지표”, 2022.6.2. 읽음. <https://stcis.go.kr/wps/main.do>
Smart Transit Card Information System, “Usage Indicator”, Accessed June 2, 2022. <https://stcis.go.kr/wps/main.do>
 25. 국토교통부, 2021.3.23. “교통카드 빅데이터 분석 결과…대중교통 이용 27% 감소”, <http://www.molit.go.kr/>
Ministry of Land, Infrastructure and Transport, 2021, March 23. “Traffic Card Big Data Analysis Results... 27% Reduction in Public Transportation Use”, <http://www.molit.go.kr/>
 26. 기상자료개방 포털, “중관기상관측(ASOS)”, 2022.6.2. 읽음. <https://data.kma.go.kr/cmmn/main.do>
Weather Data Open Portal, “Automated Synoptic Observing System (ASOS)”, Accessed June 2, 2022. <https://data.kma.go.kr/cmmn/main.do>
 27. 질병관리청, “발생현황”, 2022.6.2. 읽음. <https://ncov.kdca.go.kr/>
Korea Disease Control and Prevention Agency, “Occurrence Status”, Accessed June 2, 2022. <https://ncov.kdca.go.kr/>
 28. 질병관리청, “백신별 일일 접종현황”, 2022.6.2. 읽음. <https://ncv.kdca.go.kr/vaccineStatus.es?mid=a11710000000>
Korea Disease Control and Prevention Agency, “Daily Inoculation Status by Vaccine”, Accessed June 2, 2022. <https://ncv.kdca.go.kr/vaccineStatus.es?mid=a11710000000>

Date Received 2022-08-30
Date Reviewed 2022-11-29
Date Accepted 2022-11-29
Date Revised 2023-01-31
Final Received 2023-01-31