

# BIS(Bus Information System) 정확도 향상을 위한 머신러닝 적용 방안 연구

## A Study on the Application of Machine Learning to Improve BIS (Bus Information System) Accuracy

장 준 용\* · 박 준 태\*\*

\* 주저자 : 세종시청 대중교통과 주무관

\*\* 교신저자 : 한국교통대학교 교통시스템공학과 연구교수

Jun yong Jang\* · Jun tae Park\*\*

\* Public Transportation Division, Sejong City hall

\*\* Dept. of Transportation Systems Engineering, University of Transportation Korea

† Corresponding author : Jun tae Park, pjt724@ut.ac.kr

Vol. 21 No.3(2022)  
June, 2022  
pp.42~52

pISSN 1738-0774  
eISSN 2384-1729  
<https://doi.org/10.12815/kits.2022.21.3.42>

Received 18 March 2022  
Revised 8 April 2022  
Accepted 2 June 2022

© 2022. The Korea Institute of  
Intelligent Transport Systems. All  
rights reserved.

### 요 약

BIS(Bus Information System) 서비스는 대도시를 포함하여 중소도시까지 전국적으로 확대 운영되는 추세이며, 이용자의 만족도는 지속적으로 향상되고 있다. 이와 함께 버스도착시간 신뢰성 향상 관련 기술개발, 오차 최소화를 위한 개선 연구가 지속되고 있으며 무엇보다 정보 정확도의 중요성이 부각되고 있다. 본 연구에서는 기계학습 방법인 LSTM을 이용하여 정확도 성능을 평가하였으며 기존 칼만필터, 뉴럴 네트워크 등 방법론과 비교하였다. 실제 여행시간과 예측값에 대해 표준오차를 분석한 결과 LSTM 기계학습 방법이 기존 알고리즘에 비해 정확도는 약 1% 높고, 표준오차는 약 10초 낮은 것으로 분석되었다. 반면 총 162개 구간 중 109개 구간(67.3%) 우수한 것으로 분석되어 LSTM 방법이 전적으로 우수한 것은 아닌 것으로 나타났다. 구간 특성 분석을 통한 알고리즘 융합시 더욱 향상된 정확도 예측이 가능할 것으로 판단된다.

핵심어 : 버스도착정보, 머신러닝, 예측 알고리즘, 장단기기억네트워크, 평균제곱오차

### ABSTRACT

Bus Information System (BIS) services are expanding nationwide to small and medium-sized cities, including large cities, and user satisfaction is continuously improving. In addition, technology development related to improving reliability of bus arrival time and improvement research to minimize errors continue, and above all, the importance of information accuracy is emerging. In this study, accuracy performance was evaluated using LSTM, a machine learning method, and compared with existing methodologies such as Kalman filter and neural network. As a result of analyzing the standard error for the actual travel time and predicted values, it was analyzed that the LSTM machine learning method has about 1% higher accuracy and the standard error is about 10 seconds lower than the existing algorithm. On the other hand, 109 out of 162 sections (67.3%) were analyzed to be excellent, indicating that the LSTM method was not entirely excellent. It is judged that further improved accuracy prediction will be possible when algorithms are fused through section characteristic analysis.

Key words : Bus Arrival Information, Machine Learning, Prediction Algorithm. Long Short-Term Memory Units, RMSE

## I. 서론

버스도착시간 예측은 도시 대중교통시스템의 정보기술을 실현하는데 중요한 요소로, 승객의 여행계획을 합리적으로 조정하고, 버스의 효율적인 운영을 통해 좋은 이미지를 확립과 대중교통의 매력을 높이도록 도움을 주고 있다. 과거 2000년대 초반 현재 정거장과 다음 정거장 정보를 송출할 수 있는 수준이었으나 현재 이용자가 원하는 버스의 실시간 위치와 도착 시간까지 제공하고 있으며 스마트폰 앱을 통해 편리하게 정보를 수득 할 수 있어 만족도가 높은 서비스를 제공하고 있다. 각종 보도자료에 의하면 BIS(Bus Information System) 서비스를 제공하고 있는 서울시, 부천시, 안산시 등 각 지자체 조사 결과 만족도 85%를 상회하고 있고, 특히 버스도착시간 정보정확도는 이용자가 매우 중요하게 여기는 만족도 지표로 정확도 향상을 위한 관련 연구가 지속적으로 진행되고 있다.

도착시간 분석 및 예측을 위한 알고리즘으로 대부분의 버스정보시스템(BIS)은 이력 및 실시간 데이터를 기반으로 칼만필터(Kalman Filter, 이하 KF), 신경망(Neural Network, 이하 NN), 회귀분석(Regression Analysis), 이동평균(Moving Average) 등이 이용되고 있으며 이외에도 여러 가지 방법들이 활용되고 있다. 반면 노선 구간(정류소 to 정류소)별 도로환경과 시간대별 특성이 다르게 변화함에 따라, 실시간적 특성에 맞는 알고리즘을 선택하여 적용하는 방안도 고려되어야 한다. 이는 특정 알고리즘이 모든 구간의 예측시간을 대표하기 어려울 수 있기 때문이다. 또한 알고리즘의 경우 다양한 분야에서 머신러닝 활용이 지속적으로 증가, 버스도착시간 예측 알고리즘으로 활용이 가능할 것으로 판단된다. LSTM(Long short-term memory Units, 이하 LSTM) 기계학습의 경우 시계열 또는 시퀀스 데이터를 처리할 때 매우 유용한 알고리즘으로, 버스도착시간 예측은 최신 자료 학습 기반으로 예측되어야 하므로 적절한 분석방법론으로 활용이 가능하다.

본 연구에서는 LSTM 러닝 머신을 활용하여 BIS 정확도를 기존 알고리즘과 비교하고자 한다. 도착 예정 정보는 실시간으로 제공되어야 하므로, 구간별, 시간대별 최적 알고리즘을 선정하기 위해 평균 제곱근 편차(Root Mean Square Error, 이하 RMSE) 비교를 통해 정확도가 높은 알고리즘으로 선정하는 방안이 필요하다.

## II. 선행연구 고찰

버스정보 안내 시스템(BIS)의 신뢰성을 향상시키기 위해서는 버스도착시간 추정의 정확성이 요구된다. 주로 가중이동평균법을 활용한 연구가(Kim et al., 2009; Roh et al., 2010) 수행되었으며 개별/단일노선추정보다 정확도가 향상되는 것으로 제시되고 있으나 버스정류소 서비스시간이 포함되지 않고 버스유형(광역,순환 등)에 따라 모형 적용 결과가 상이함이 단점으로 지적되었다. 그 외 가중이동평균법과 칼만필터링이 결합한 퓨전알고리즘 적용 정확도 추정 연구(Wi, 2013), 정류장·경로기반(Chen et al., 2009) 연구, 커널 회귀식을 적용한(Baker and Nied, 2014) 공간적 특성 계량 연구 등 다양한 방법론을 적용한 연구가 진행되었으며 공통적으로 모든 시간대를 아우르는 정확도 산정의 어려움 극복과 과대·과소 추정에 따른 신뢰도 향상이 보완되어야 함을 향후 연구로 제시하였다. 또한 칼만필터와 Support Vector Regression을 결합(K-SVR)하여 인접하는 두 버스정류장의 예측 정확도를 향상(Zhang et al., 2021) 시킬수 있으나 버스 헤드웨이가 가급적 일정한 시간대 효과가 있는 것으로 나타났다. 날씨와 같이 통행시간의 변동성이 큰 특수한 상황, 반복지연 및 비반복지연에 의한 낮은 버스도착시간 예측 정확성이 개선(Patricia et al., 2015; Yang et al., 2016) 되어야 함을 지적하고 있다.

최근 인공지능 학습 기반 정확도 예측 연구가 다양한 분야에서 수행되고 있다. 이와 관련하여 버스 통행시간 예측시 비지도학습의 SOM 알고리즘을 활용한 연구에서 학습데이터로 예측시 정확도가 우수하며 패턴

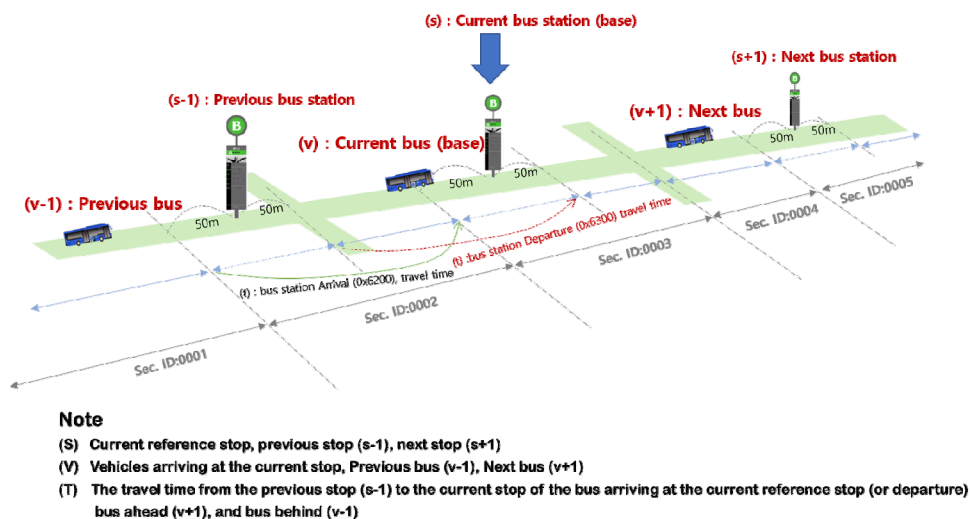
분석을 통해 정밀도 높은 구간 통행속도 상태 예측이 가능함을 제시하였다.(Lee et al., 2017; Chang et al., 2019) 반면 주로 사용하고 있는 분석 알고리즘 대비 실증 데이터 기반 기계학습 알고리즘 비교 연구는 미흡한 실정으로 기계 및 딥 러닝 알고리즘의 예측 정확도가 SVR, OLS보다 우수한 정확도를 나타내는 것으로 분석되고 있다.(Dancho and Titus, 2020; Ashwini et al., 2022) 인공지능 RNN(Recurrent Neural Network) 기반 Deep Feature 추출 프레임워크를 사용하는 경우 시공간 상관관계 포착이 가능하여 기존 SVR 모델 대비 4.82% 예측 정확도가 향상되는(Yuan et al., 2020) 것으로 분석되어 다양한 인공지능 및 기계학습 적용의 필요성이 부각되고 있다.

### III. 연구 방법론

#### 1. 기계학습 적용 및 데이터 연관 관계

머신러닝의 기본 개념은 데이터를 기반으로 해서 어떤 가설(공식)을 만들어 낸 다음, 그 가설에서 나온 값이 실제 측정값과의 차이(Cost)가 최소한의 값을 가지도록 변수에 대한 값을 컴퓨터를 이용해서 찾은 후, 찾아진 값을 가지고 학습된 모델을 정의해서 예측을 수행하는 것이다. 서울시 BIS의 정확도 향상을 위한 최적의 기계학습 방법을 찾기 위해 데이터간의 연관관계를 파악하고, 이 관계를 바탕으로 기계학습 방법론별 성능을 측정하고 기존 알고리즘과 비교를 하는 방식을 설정하였다.

정류소 출도착정보는 정류소 구간의 여행시간과 승하차인원, 앞차와의 시간간격등의 정보가 산출되는 시점이며, 시민들에게 제공되는 도착정보가 생성되는 시점 정보이다. 본 연구에서도 이 정보를 활용하였으며 정류소 출발정보는 데이터간의 연관관계를 파악하고, 기존 알고리즘과 비교를 위해 정류소 도착정보로 기계학습 성능을 측정하였다. 서울BIS의 정류소는 진입시 발생하는 “정류소도착(0x6200)” 정보와 승객의 승하차 후 차량이 출발하여 진출시 발생하는 “정류소출발(0x6300) 정보로 구분되며, 진입/진출의 거리는 원격센터에서 지정하게 된다(기본값 50m).



<Fig. 1> Conceptual diagram of the analysis section

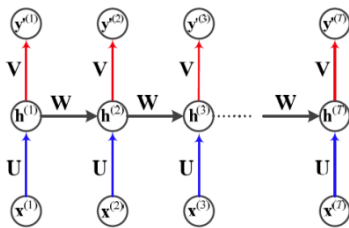
버스도착정보는 특정 정류소를 통과한 차량들의 정보를 이용하여 다음 도착예정차량의 시간을 추정하는 것으로 현재 차량(v)의 정보는 입력값으로, 뒷차량(v-1)의 여행시간을 참(결과)값으로 사용하였다. 또한 데이터 속성간 연관관계로 입력값과 참(결과)값을 연결시켜 데이터간의 연관성을 파악하는 의사결정트리(Decision Tree)를 이용하여, 예측 정확도와 표준오차를 측정하고 속성간의 연관성 정도를 파악하고자 하였다.

<Table 1> Definition of data processing for machine learning

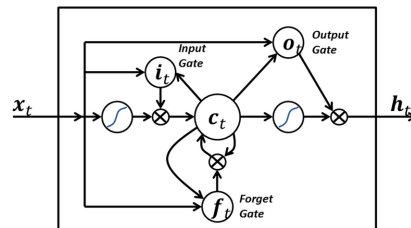
Table name	Definition	Note
Section ID	Section of bus station	-
Current point of time	Departure/arrival time of the bus station	Input value
Route ID	Route ID	-
Bus / Vehicle ID : (v)	Bus / Vehicle ID : (v)	-
Travel time(s) : (t)	Required time : Previous-current bus station	Input value
Travel distance (m)	Distance : Previous-current bus station	-
Travel speed (km/h)	Space speed : Previous-current bus station	-
Number of passengers, getting on	# of boarding : current bus station	-
Number of passengers, getting off	# of getting off : current bus station	-
Number of Passenger, on-board	# of Passenger, after getting on and off : current bus station	Input value
Time interval between bus(min)	Time difference with previous passing bus	Input value
Previous bus (v-1) Travel time:(t)	-	True value

## 2. LSTM 머신러닝 모델 구축 및 성능 측정

전통적 Neural Network는 이전에 일어난 사건을 바탕으로 나중에 일어나는 사건을 유추하지 못한다. 이를 개선한 Recurrent neural network (이하 RNN)은 스스로 반복하며 이전 단계의 정보를 지속되도록 한다. LSTM 은 RNN의 한 종류로 긴 의존 기간(시퀀스) 문제를 다루는데 적합하도록(Schmidhuber and Hochreiter, 1997) RNN을 변형하여 설계한 네트워크이다. LSTM의 구조는 RNN과 달리 게이트(Gate)라는 요소를 통해 정보를 더하거나 제거함으로써 선택적으로 정보를 저장할 수 있다. LSTM의 게이트는 크게 forget gate( $f_t$ ), input gate( $i_t$ ), output gate( $o_t$ )의 3개로 구성되며, 모든 게이트는 sigmoid 연산을 수행하여 0에서 1사이의 값을 출력한다. forget gate는 셀 상태에서 현재 정보의 제거 정도, input gate는 현재 정보의 반영 정도, output gate는 데이터의 출력을 결정한다. 수식(3)과 수식(6)은 각각 셀 상태 갱신과 출력 결과를 나타낸다.



<Fig. 2> RNN structure



<Fig. 3> LSTM structure

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \dots\dots\dots (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \dots\dots\dots (2)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \dots\dots\dots (3)$$

$$C_t = f_t^* C_{t-1} + i_t^* \tilde{C}_t \dots\dots\dots (4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \dots\dots\dots (5)$$

$$h_t = o_t^* \tanh(C_t) \dots\dots\dots (6)$$

모델 구축은 Python과 keras LSTM을 사용하였다. 데이터셋 구축은 구간ID, 발생시각, 통행시간이며 학습 데이터 67%, 테스트 33% 비율로 구성하였다. 측정 기준은 테스트 데이터 중 입력값으로 “기준시각, 여행시간”을, 참(결과)값으로 “뒷차량 여행시간”을 사용하여 예측값과 참값을 비교하였다.

기계학습 성능은 학습과 테스트로 구분하고 테스트 결과를 측정하며 최적 방법을 선정하기 위해 4가지 성능 측정 모델을 사용하여 정확도와 RMSE를 측정하였다. 첫째, 서포트벡터머신(Support Vector Machine, 이하 SVM) 모델로 데이터 포인트와 구분하는 선형 결정 경계(초평면)을 찾아서 데이터를 분류하는 모델이다. 둘째, 의사결정트리(Decision Tree, 이하 DT)는 예측자 값이 훈련된 가중치에 비교되는 분기 조건으로 구성되고 분기 수와 가중치 값은 훈련 프로세스에서 결정하며 셋째, 뉴럴 네트워크(Neural Network, 이하 NN)의 경우 연결성이 높은 뉴런 네트워크로 구성되어 입력을 원하는 출력에 연관시켜 입력이 정확한 응답에 매핑되도록 연결 강도를 반복 수정하는 방식으로 훈련을 통해 적용 가능하다. 넷째, 장단기 기억 네트워크 LSTM은 순환 신경망(Recurrent neural Network)의 장기학습 단점을 보완한 것으로 시계열 기반 순환신경망모델이다. 그 외 이동평균, 칼만필터, 신뢰평균은 서울BIS 도착정보를 생성한 이력자료 받아 활용하였다.

기계학습 성과와 기존 알고리즘의 비교는 BIS 정확도이력데이터로 정확도와 RMSE를 비교하였다. 정확도의 경우 절대평가를 위해 현재 서울시 BIS의 정확도 기준인 ±120초와 비교하며 기준값 변경시의 정확도 변화까지 포함하였다. 성능 측정에 사용한 모델은 <Table 2>와 같다.

<Table 2> Performance Measurement Model

Table name	SVM	DT	NN	LSTM
Parameters	C=10, cache_size=200, class_weight=None, coef0=0.0, decision_function_shape=None, degree=3, gamma='auto', kernel='rbf', max_iter=-1, probability=False, random_state=None, shrinking=True, tol=0.001, verbose=False	class_weight=None, criterion='gini', max_depth=None, max_features=None, max_leaf_nodes=None, min_impurity_split=1e-07, min_samples_leaf=1, min_samples_split=2, min_weight_fraction_leaf=0.0, presort=False, random_state=None, splitter='best'	activation='logistic', alpha=0.0001, batch_size='auto', beta_1=0.9, beta_2=0.999, early_stopping=False, epsilon=1e-08, hidden_layer_sizes=(100), learning_rate='constant', learning_rate_init=0.001, max_iter=2000, momentum=0.9, nesterovs_momentum=True, power_t=0.5, random_state=None, shuffle=True, solver='adam', tol=0.0001, validation_fraction=0.1, verbose=False, warm_start=False	Sequential() LSTM(12, input_shape=(1, 1)) Dense(1) loss='mean_squared_error', optimizer='adam', metrics=['accuracy'] epochs=10, batch_size=1, verbose=1
Note	scikit learn	scikit learn	scikit learn	karas, tensorflow

## IV. 기계학습 적용 분석 결과

### 1. 데이터 연관 관계 분석

기준시각과 여행시간은 학습과 예측에 대한 기본정보로 필수 요소이며 기준시각과 여행시간 외 정확도에 영향을 미치는 속성을 분석하였다. 정확도는 정확히 1초 단위까지 맞은 것을 기준으로 여행시간 예측에 영향을 미치는 속성셋을 결정하기 위해 원시자료로부터 추출가능한 현재시각, 요일구분, 구간거리, 여행시간, 승하차인원, 재차인원, 앞차간격 속성을 Attribute set으로 조합·구분하여 정확도와 표준오차를 살펴보았다. 표준오차 차이는 거의 없고 어느 한 속성이 절대 우위를 점하지 못하는 것으로 나타나 필수 요소 외 나머지 속성은 연관성이 아주 미미한 것으로 판단된다. Attribute1의 경우 정확도가 0.1% 향상되었으나 Attribute2와 비교하였을 때 큰영향을 미치지 않는 것으로 나타났다. 이는 특정 구간의 여행시간 예측에 가장 중요한 요소로 Attribute2가 가장 중요하고 결정적인 속성임을 알 수 있다.

<Table 3> Results of data association analysis


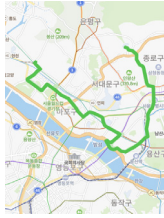
Table name	Attribute1	Attribute2	Attribute3	Attribute4	Attribute5
Input value	·Current point of time ·Day of the week ·Route ID ·Travel distance ·Travel time ·# of getting on and off ·# of on-board Passengers ·Time interval of Preceding bus	·Current point of time ·Travel time	·Current point of time ·Route ID ·Travel time	·Current point of time ·Travel time ·# of on-board Passengers	·Current point of time ·Travel time ·Time interval of Preceding bus
True value	Previous bus(v+1), Travel time:(t)				
Accuracy	2.2%	2.1%	2.1%	2.1%	2.0%
Sta. error	50s	50s	49s	50s	51s
Dataset	76,258 (11/7, 11/8, 11/9 Departure information about the bus station)				
	Number of learnings : 57,147		Number of tests : 19,110		

### 2. 모델별 정확도 및 표준오차 성능

알고리즘 모델별 학습과 테스트 과정을 거쳐 정확도를 측정하고 모델별 BIS 정확도와 비교하였다. 분석 대상은 109번 간선노선(우이동~광화문), 7016번 지선노선(수색동~상명대)을 대상으로 2021년 11월 7일에서 9일까지 3일간 데이터를 분석하였으며 특정 정류소 구간에 도착하는 1전 정류소 차량으로 한정하였다. 109번 노선의 경우 총운행거리 28.16km이며 평일 기준 배차간격은 12~15분이다. 7016번 노선의 경우는 총운행거리 56.0km, 평일 기준 배차간격 5~12분이다.

분석 결과 LSTM 모델이 가장 안정적이고 우수한 것으로 나타났다. 오차 기준 변화( $\pm 20s$  단위씩 증가)에 따라 정확도 차이가 커지고 있으며, 표준오차가 정확도에 영향을 미치는 것으로 나타났다.

알고리즘 모델별 학습 후 테스트 시행 결과 LSTM 모델이 36.4초로 표준오차가 가장 낮았으며, 총 162개 구간 중 109개 구간에서 67.3% 비중으로 우수한 것으로 도출되었다. 기존 알고리즘 중에서는 신뢰평균이 뉴럴 네트워크모델 뒤를 이어 3위로 나타났다. 시계열기반의 순환신경망 모델인 장단기기억네트워크 LSTM 기

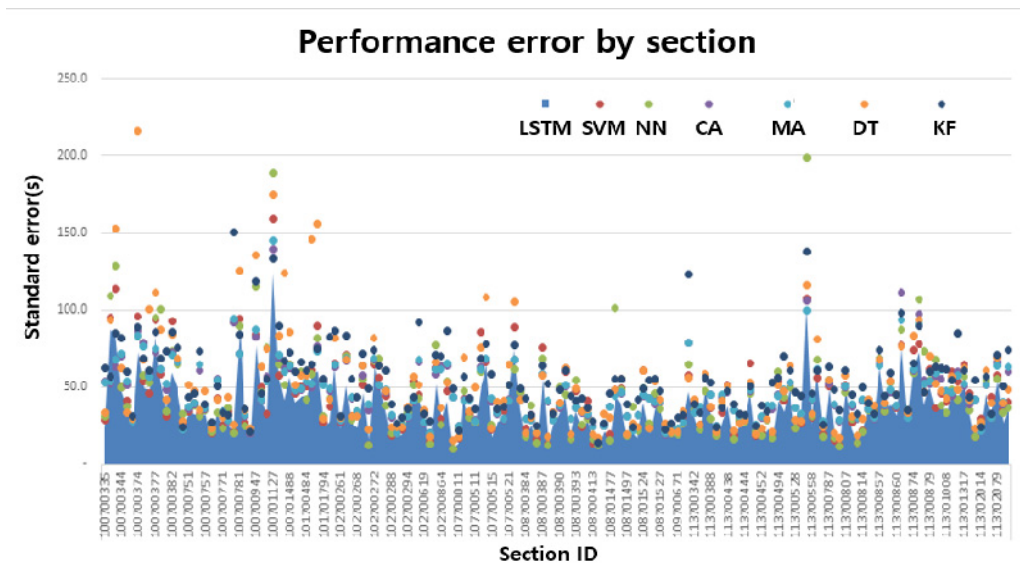
Bus No. 109	Bus No. 7016
 <ul style="list-style-type: none"> <li>·Starting point - Ui-dong, Gangbukgu</li> <li>·End point - KT Gwanghwamun branch, Jongnogu</li> <li>·Total operating distance - 28.16km</li> <li>·Bus interval - 12~15min.(Weekdays)</li> </ul>	 <ul style="list-style-type: none"> <li>·Starting point - Susaekdong, Eunpyeonggu</li> <li>·End point - Sangmyung University, Jongnogu</li> <li>·Total operating distance - 56.0km</li> <li>·Bus interval - 5~12min.(Weekdays)</li> </ul>

<Fig. 4> Overview of bus routes

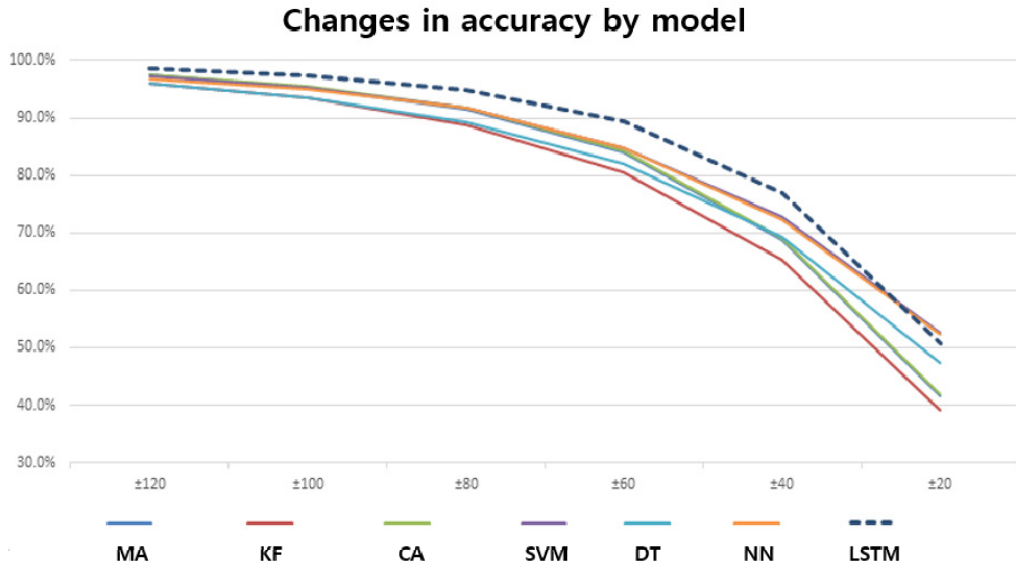
계획수행방법이 기존 알고리즘에 비해, 정확도는 약 1% 높고, 표준오차는 약 10초 정도 낮은 것으로 분석되었다. 이 결과는 1전 정류소에 위치한 차량에 대한 성능치로써, 평균 1~5전의 차량에 대한 도착정보를 생성하는 서울시 BIS의 특성상 10~50초의 오차를 줄일 수 있는 것으로 나타났다.

<Table 4> The results of the accuracy measurement by model

Table name	# of learnings	# of tests	Accuracy(%)			
			±120s	±60s	±40s	±20s
Moving average (MA)	-	22,747	97.4	84.1	68.9	41.6
Kalman filter (KF)			96.0	80.4	65.2	39.1
Confidence average (CA)			97.4	84.2	69.0	42.0
Support Vector Machine (SVM)	57,148	19,110	97.3	84.7	72.7	52.6
Decision tree (DT)			96.0	82.0	69.3	47.4
Neural networks (NN)			96.8	84.7	72.3	52.3
Long short-term memory Units (LSTM)			98.5	89.4	76.8	50.8



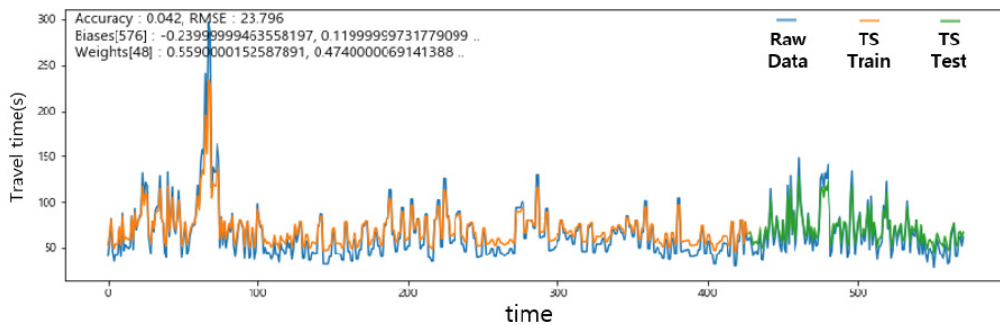
<Fig. 5> Performance error by section



<Fig. 6> Change in accuracy by model

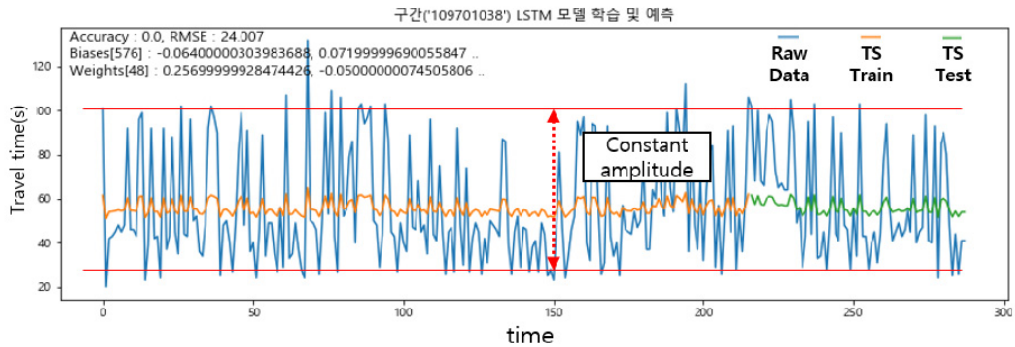
### 3. LSTM 모델 적용 특성

장단기기억네트워크 LSTM 모델은 주기별 일정 패턴을 가지는 경우 학습과 예측이 우수한 것으로 판단된다. LSTM 예측이 떨어지는 구간은 NN > CA > MA 순으로 예측정확도가 높은 것으로 검토되었다. 학습(TS Train)과 예측(TS Test)이 잘된 경우와 그렇지 않은 경우를 비교하면 진폭이 일정하고, 편차가 큰 경우 학습 결과가 좋지 않는 현상이 발생하는 것으로 나타났다. 이는 교통신호, 지체체로 이해 패턴이 불규칙한 경우 나타나는 특성으로 검토되었으며 일정한 진폭내에서 2~3가지의 일정값들이 반복적으로 출현함에 따라 예측에 적합하지 않는 결과로 학습되는 것으로 판단된다. 이 경우 평균에 가까운 값으로 예측되었으나 특정 몇 개 구간의 경우에만 10~20초의 차이가 발생하였을 뿐, 대부분은 구간별로 우수한 타 모델과의 예측값이 1~3초 사이의 편차를 나타내었다.



<Fig. 7> An example of good LSTM learning





<Fig. 8> An example of poor LSTM learning

분석 결과를 요약하면 정확도 향상은 결과적으로 표준오차를 줄이는 것으로 LSTM은 1전의 차량예측시 표준오차가 10초 정도 낮고 모든 구간 대비 67%로 우수한 것으로 나타났다. 서울 시계내에의 도착정보 제공 시 대부분 차량은 1~5전 정류장에 위치하며 이는 평균적으로 10~50초 정도 도착정보 오차가 줄어들 것으로 예상된다. 속성과의 관계를 살펴보면 승하차/시격 등은 예측에 상관관계가 적다. 이는 정류소 도착정보에 반영되어 있기 때문이며 상관관계가 큰 것은 시간대/여행시간으로 지역에 따라 시간대/노선ID/여행시간 등도 고려할 필요가 있다.

또한 기존 알고리즘 중 신뢰평균 방법이 모든 구간대비 9.3%로 이동평균 1.9%보다 우수한 것으로 신뢰평균은 이동평균대비 민감도가 높아 도로의 상황이 적절히 반영되어 일부 구간에서 활용가치가 있다고 판단된다. 보완점으로 버스도착시간 예측은 정류장 구간을 기준으로 하고 있으며, 정류장 구간별 도로 상황에 맞는 적절한 알고리즘을 반영 할 수 있어야 할 것이다. 다음으로 표준오차 결과에서 나타나듯 특정 구간들은 다른 모델이 우수한 특징을 보이고 있어, 구간별 특성에 따라 모델을 선택하는 방안도 필요하다. 선택의 기준은 특정 정류소 1개를 기준으로 5~10개 차량의 도착정보에 대한 표준오차를 산출하여 표준오차가 가장 적은 모델을 자동 선택하도록 하는 방안이 효과적으로 판단되며 상시 표준오차를 측정하여 가장 우수한 알고리즘으로 예측 여행시간을 제공하는 기술적 보완이 필요하다.

## V. 결 론

버스 도착 시간 예측은 도시 대중교통시스템의 정보기술을 실현하는데 중요한 요소로, 승객의 여행계획을 합리적으로 조정하고, 버스의 효율적인 운영을 통해 좋은 이미지를 확립과 대중교통의 매력을 높이도록 도움을 주고 있다. 서울시 BIS는 2003년 버스종합사령실(BMS) 구축을 시작으로 현재 실시간 버스도착시간을 예측하여 제공하고 있으며 정보 정확도에 대한 다양한 평가가 진행되다. 분석 방법으로는 운행 이력 및 실시간 데이터를 기반으로 칼만필터(KF), 신경망(NN), 회귀분석(Regression), 이동평균(MA) 등의 알고리즘이 주로 활용하고 있다.

LSTM은 시계열적 정보 기반 최신 자료를 학습하는 방식으로, 예측 및 추세분석의 장점이 있는 알고리즘이다. 본 연구에서는 BIS 정확도에 대해 칼만필터 등 기존 분석방법론과 LSTM 기계학습방법의 예측 정확도를 비교하였다. 분석 대상으로 서울시 109번, 7016번 2개 노선의 정류소 출도착정보를 사용하였고 BIS 정확

도 이력 데이터로 정확도와 평균제곱근오차(RMSE)를 비교한 결과 LSTM 기계학습방법이 기존 알고리즘에 비해, 정확도는 종합적으로 약 1% 높고, 표준오차는 약 10초 낮은 것으로 분석되었다. 그러나 LSTM이 모든 구간에서 적합함을 보이지는 않는 것으로 나타났다. 신호, 지정체로 인해 패턴이 일정하지 않은 경우가 특히 예측 정확도가 낮은 것으로 나타났으며 세부적으로는 버스 통행시간의 진폭이 일정하고, 편차가 큰 경우 학습 결과가 좋지 않는 현상이 발생하는 것으로 나타났다. 서울시 BIS데이터는 원시수집자료의 경우 1일 6천 만건이 넘는 상황으로 본 연구에서는 전구간에 대한 데이터를 구축 및 알고리즘별 분석을 수행하지 못하였다. 이에 지선 및 간선 각 1개 노선 분석으로 한정하여 분석 노선 확대를 통한 추가적 검증이 필요하다.

또한 세부 구간별 기하구조, 노선굴곡도, 버스 주행 방향 외부요인 등 본 연구에서 고려하지 못한 다양한 요인의 영향이 있을 것으로 향후 구간별 심층 특성분석을 예측 영향요인으로 반영해야 할 것이다. 이러한 점을 고려시 정류소 구간별 도로상황에 맞는 적절한 알고리즘을 평가하여 LSTM 기계학습방법과 결합하여 사용하는 최적화 융합 연구가 필요하다. 실무적으로 LSTM 적용에 대한 효과 구간과 비효과 구간에 대한 정의가 필요하며 LSTM 대비 타 알고리즘 적용 구간에 대한 단계적 연구가 필요하다.

## REFERENCES

- Ashwini, B. P., Sumathi, R. and Sudhira, H. S.(2022), “Bus Travel Time Prediction: A Comparative Study of Linear and Non-Linear Machine Learning Models”, *Journal of Physics Conference Series*, 2161, Canada.
- Baker, C. B. and Nied, A.(2014), “Predicting bus arrival using One Bus Away real-time data”, *Computer Science*, p.204.
- Chang, H. H., Yoon, B. J. and Lee, J. S.(2019), “A Study on Development of Bus Arrival Time Prediction Algorithm by using Travel Time Pattern Recognition”, *Journal of the Korean Society of Civil Engineers*, vol. 39, no. 6, pp.833-839.
- Chen, X., Lei, Y., Yushi, Z. and Jifu, G.(2009), “Analyzing urban bus service reliability at the stop, route, and network levels”, *Transportation Research Part A Policy and Practice*, vol. 43, no. 8, pp.722-734.
- Dancho, P. and Titus, Z.(2020), “Real-Time Public Transportation Prediction with Machine Learning Algorithms”, *IEEE International Conference on Consumer Electronics(ICCE)*, USA.
- Kim, W. G., Roh, C. G., Heo, M. G. and Son, B. S.(2009), “A Development of Bus Arrival Time Estimation Model”, *Seoul Studies*, vol. 10, no. 2, pp.21-29.
- Lee, J. Y., Gu, E. M., Kim, H. J. and Jang, K. T.(2017), “Long-term Prediction of Bus Travel Time Using Bus Information System Data”, *Journal of Korean Society of Transportation*, vol. 35, no. 4, pp.348-359.
- Li, Y., Pree, T. and Poon, T.(2021), “A Real-Time Bus Arrival Time Prediction System Based on Spark Framework and Machine Learning Approaches: A case study in Chiang Mai”, *Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering*, Thailand.
- Patricia, H., Justin, A., Yvonne, C. and Richard, H.(2015), “Transit Travel Time Reliability: Shifting the Focus from Vehicles to Customers”, *Transportation Research Record: Journal of the*

*Transportation Research Board*, vol. 2535, no. 1, pp.35-44.

- Roh, C. G., Kim, W. G. and Son, B. S.(2010), “Development of Bus Arrival Time Estimation Model by Unit of Route Group”, *Journal of Korea Transportation Research Society*, vol. 28, no. 1, pp.135-142.
- Schmidhuber, J. and Hochreiter, S.(1997), “Long short-term memory”, *Neural Computation*, vol. 9, no. 8, pp.1735-1780.
- Wi, J. R.(2013), *A Study on the Weighted Value Application Standard of Fusion Algorithm for the Estimate Bus Arrival Time Prediction*, Master's Thesis, University of Seoul.
- Yang, M., Chen, C., Wang, L. and Yan, X.(2016), “Bus Arrival Time Prediction Using Support Vector Machine with Genetic Algorithm”, *Neural Network World*, vol. 26, no. 3, pp.205-217.
- Yuan, Y., Shao, C., Cao, Z., He, Z., Zhu, C., Wang, Y. and Jang, V.(2020), “Bus Dynamic Travel Time Prediction: Using a Deep Feature Extraction Framework Based on RNN and DNN”, *Electronics*, vol. 9, no. 11, pp.1-20.
- Zhang, X., Yan, M., Xie, B., Yang, H. and Ma, H.(2021), “An automatic real-time bus schedule redesign method based on bus arrival time prediction”, *Advanced Engineering Informatics*, vol. 48, 101295, pp.1-15.