Prediction of Traffic Congestion Using a Time-Series Model and Spatiotemporal Data: A Case Study of the Atlanta Downtown Connector

Yunsik Kim
Dongguk University, yunsik97@dgu.ac.kr

Follow this and additional works at: https://digitalcommons.kennesaw.edu/thegeographicalbulletin

Part of the Geography Commons

Recommended Citation
Kim, Yunsik (2023) "Prediction of Traffic Congestion Using a Time-Series Model and Spatiotemporal Data: A Case Study of the Atlanta Downtown Connector," The Geographical Bulletin: Vol. 64: Iss. 2, Article 7. Available at: https://digitalcommons.kennesaw.edu/thegeographicalbulletin/vol64/iss2/7

This Article is brought to you for free and open access by the Active Journals at DigitalCommons@Kennesaw State University. It has been accepted for inclusion in The Geographical Bulletin by an authorized editor of DigitalCommons@Kennesaw State University. For more information, please contact digitalcommons@kennesaw.edu.
Traffic congestion can be described as the additional travel time or delay experienced beyond what is typically encountered in light or unrestricted traffic conditions (Lomax et al., 1997). According to recent statistics, urban areas worldwide suffer due to the traffic congestion, and American drivers, on average, wasted 51 hours in 2022, costing $869 (Schaper, 2023). Traffic congestion also causes time loss, mental stress, and the added pollution to global warming (Akhtar & Moridpour, 2021).

Since the transportation system plays an important role in understanding, designing, and managing urban areas (Chen et al., 2020), many recent studies have been conducted to predict traffic volume or traffic congestion.

To analyze and predict transportation systems and traffic congestion which are recognized as geographical phenomena, research utilizing GeoAI, an interdisciplinary fusion of geospatial sciences and artificial intelligence (AI), is on the rise. GeoAI, equipped with spatial analysis expertise and AI capabilities, possesses the capacity to decipher previously elusive patterns within the dynamics in urban areas. There have already been increasingly collaborative GeoAI studies for urban activities (Hu et al., 2015; Kim & Yang, 2023), remote sensing (Huang et al., 2018; Reichstein et al., 2019), and transportation systems (Kim & Wang, 2016; Lee et al., 2019). GeoAI also can be used to resolve problems in human environmental systems and their interaction, with a focus on spatial contexts (Gao, 2021). While various algorithms such as support vector regression (SVR) (Ge et al., 2019), convolutional neural network (CNN) (Zhang et al., 2019), and K-nearest neighbor (KNN) (Habtemichael & Cetin, 2016), long short-term memory (LSTM) for time series analysis is widely employed in predicting traffic congestion due to the temporal characteristics of traffic volume at a specific area or point. Huang et al. (2019) suggested traffic congestion prediction based on bus driving time using LSTM and used the congestion index and classification model to analyze and classify the level of congestion. Ranjan et al. (2020) proposed a hybrid neural network architecture combining CNN, LSTM, and Transpose CNN to predict network-wide congestion level and obtained city-wide traffic data from an open-source online web service. Shin et al. (2020) developed a series of processes for traffic congestion prediction using the LSTM model, including the process of getting rid of outliers and correcting missing values in the collected data. Through these processes, it was found that their model performed better than other deep neural network models.

However, efforts to predict traffic congestion face limitations tied to the data. Basically, the three elements comprising traffic congestion encompass intensity, extent, and duration, where...
intensity signifies the congestion’s severity, usually quantified as a rate, duration pertains to the period during which the transportation system experiences congestion, and extent encompasses the geographical distance of congested roads or the number of affected travelers or vehicles (Lomax et al., 1997; Falcoccio & Levinson, 2015). Multiple metrics for explaining traffic congestion have been developed such as travel time index (TTI), vehicle miles traveled (VMT), volume capacity (V/C) ratio, and peak traffic period duration (PTPD) (Rao & Rao, 2012; Rahman et al., 2022). Although these metrics consider individual traffic congestion elements, they do not represent multiple components at the same time. To address this gap, Seong et al. (2023) suggested a comprehensive traffic congestion index that can handle multiple elements, including intensity, duration, and extent, which is called ‘distanceTime’ unit. In this study, distanceTime unit data was used as historical traffic data to predict traffic congestion in the urban area.

In prior research, obtaining historical traffic volume data typically involved installing sensors on roadways or gathering information from vehicles equipped with Global Positioning System (GPS) technology. Nonetheless, these data collection approaches are beset by two significant challenges: issues of data reliability and cost. While an array of data sources, including traffic sensors and GPS data, are used for congestion modeling, their coverage can be inconsistent, with some regions or road segments lacking adequate data representation (Li et al., 2020). Furthermore, data quality and reliability can be compromised by sensor malfunctions, incomplete reporting, or inaccuracies in the collected data. Timeliness is also a restriction, as real-time data may not always be readily accessible (Altintasi et al., 2017). Another constraint related to data collection is the high cost associated with traffic data acquired based on sensors or GPS equipped devices. The data from fixed sensors are also difficult to obtain and access because it requires special permission (Ranjan et al., 2020). In such cases as well, the distanceTime data can tackle the mentioned limitations. The distanceTime data is acquired through Google Maps. In an internet-connected environment, anyone can access Google Maps, and researchers can collect data for free within specified limits by using the application programming interface (API) for traffic data collection.

Based on the backgrounds and related studies as mentioned above, this research aims to propose a framework for predicting traffic congestion using the GeoAI method based on cost-effective spatiotemporal data. To achieve this research purpose, the following specific steps are undertaken: 1) acquiring low-cost traffic data to establish a spatiotemporal traffic dataset from the online map service—Google Maps; 2) identifying the traffic congestion patterns in urban areas with significant congestion; 3) developing a time-series analysis model that incorporates the characteristics of traffic data to predict traffic congestion in the study area. This paper is structured as follows. Section 1 introduces the research background and previous studies associated with prediction of traffic congestion. Section 2 provides an explanation of the research area and methodology employed in this study, along with details on the model implementation. Section 3 presents the results of this study, and Section 4 discusses the contributions of this research and gaps pertaining to predicting traffic congestion. In Section 5, the conclusion of this study is presented.
Figure 1. Study area: Atlanta Downtown Connector (Interstate highway 75/85).

Figure 2. Real-time traffic data obtained from the Google Traffic Layer (Atlanta metropolitan area).
To calculate the length of the pixels covered by the road in the traffic images, the image and the road centerline in shapefile format were overlaid, and then the number of pixels that represented the road in the image was counted. In the case of the Atlanta metropolitan area, each pixel was calculated to be approximately 20.14 meters long by dividing the total length of the road in the shapefile by the number of pixels denoting the road. Next, the number of pixels was separated into color categories—green (free flow), orange (light congestion), red (medium congestion), and dark red (heavy congestion)—that indicate the level of traffic congestion at that pixel. By counting the number of color-coded pixels, the extent of traffic conditions at specific time intervals were quantified. This pixel-level analysis enabled the creation of a comprehensive dataset reflecting traffic variations. To summarize traffic congestion in the study area, empirical weighting values were assigned to each color category—orange, red, and dark red—after a review of previous studies on the Google Traffic Layer (Google, n.d.; Bian et al., 2016; García-Ramírez, 2020; Ji et al., 2021). These weighting values were as follows: 0.25 for orange, 0.5 for red, and 1.0 for dark red. These values were applied to represent the severity of traffic congestion in the Atlanta Downtown Connector. Thus, mileHours, the traffic congestion index utilized in this study, is calculated using the following formula based on the previous study (Seong et al., 2023):

\[ \tau = 24 \times P \times C \times (0.25 \times R_{orange} + 0.5 \times R_{red} + 1.0 \times R_{darkred}) \]

where \( \tau \) is traffic congestion amount in mileHour, \( P \) indicates length of one pixel in meters translated from the number of pixels. \( C \) is the meter to mile conversion factor (i.e., 0.00062137), and \( N \) is the average number of congestion pixels from 144 10 min samples a day. The value 24 was multiplied to calculate a daily congestion amount with the mileHours unit. The traffic congestion dataset generated through this process serves as the foundation for the machine-learning model, allowing development of a prediction system for traffic congestion in the study area.

### 2.2. Methodology

#### 2.2.1. Research Outline

This study comprises three steps for predicting traffic congestion in a heavily congested road section. The first step involves data collection and preprocessing, which includes a feature engineering process for a time series analysis model. Using the data obtained via the Google Maps API, the traffic congestion amount was computed in ‘mileHour’ units. Subsequently, temporal variables were generated to build the time series analysis model for forecasting traffic congestion. The second step entails identifying the data distribution pattern. In other words, it was intended to understand the distribution of traffic congestion data collected every 10 minutes in a time series. The third stage is prediction of traffic congestion using the time series analysis model. In this study, congestion prediction was conducted separately for the northbound and southbound of the Atlante Downtown Connector. The predictions were then analyzed by comparing them with typical traffic data provided by Google Maps.

### 2.2.2. Long Short-Term Memory

In this study, the prediction of traffic congestion was approached using a deep learning algorithm known as LSTM, which is a specific type of Recurrent Neural Networks (RNNs). RNNs and LSTM have gained widespread popularity in time series forecasting tasks due to their ability to capture temporal dependencies in sequential data. RNNs are a class of neural networks designed to work with sequential data, where the order of data points carries valuable information (Giles et al., 2001). Unlike feedforward neural networks, RNNs have connections that loop back on themselves, allowing them to maintain a hidden state that captures information from previous time steps. LSTM is a specialized variant of RNN that addresses some of the limitations of traditional RNNs, such as the vanishing gradient problem (Hochreiter & Schmidhuber, 1997). LSTM is equipped with gating mechanisms that enable them to learn and remember patterns over long sequences, making them particularly effective for modeling time series data with complex temporal dependencies. It takes a series of items and processes them step by step. At each step, it decides what to keep or forget from previous steps, what new information to add, and what to pass on to the next step. It does this by using three special gates—forget gate, input gate, and output gate—that control its memory. This way, LSTM can understand and remember patterns and relationships in a sequence of data, making it great for tasks like predicting future values in a time series or understanding the context in a sentence for tasks like language translation or speech recognition.

#### 2.2.3. Building a Multivariate and Time-Lagged LSTM Model

Traffic volume often exhibits seasonality, which means it follows recurring patterns or cycles over specific time intervals. Daily traffic volume typically shows clear patterns with peak hours during morning and evening rush hours when people commute to and from work. Traffic tends to be lower during the night. Traffic also can vary from day to day. Weekdays generally have higher traffic amounts than weekends. Fridays may experience increased traffic as people travel for weekend getaways.

In consideration of the seasonality of the traffic data, the temporal factors have been included as explanatory variables in the traffic congestion prediction model. Specifically, day of the week and time of day are taken into account by converting them into sine and cosine values (Bhaskaran et al., 2013). This technique is commonly used in time series analysis to preserve cyclical nature of time-related data and provide a machine learning model with more meaningful input features for tasks. All possible positions around a circle are accounted for when both sine and cosine functions are considered. Sine and cosine functions are also orthogonal, meaning they are uncorrelated with each other. This property can be beneficial in machine learning models because it avoids introducing multicollinearity, which can lead to unstable model estimates. A time-lagged LSTM model was also built, considering the characteristic that the current traffic volume in time series analysis is continuous from the previous time step's traffic volume. In this research,
the data from the previous one hour was utilized to predict traffic congestion amounts at a given time. The time-lagged input data is described in Figure 3.

![Figure 3. Conceptual diagram of time-lagged input data.](image)

In this study, the LSTM model was implemented in the Python environment using the TensorFlow and Keras libraries. The model is constructed as a Sequential neural network, allowing for the orderly addition of layers. The first layer, an LSTM layer with 32 units, serves the primary sequence processor. It is set to return sequences, ensuring the output retains the sequential information. The second LSTM layer, also with 32 units, follows the first. The final layer in the model is a Dense layer featuring a single unit. This is ideal for regression tasks, as it yields a singular output. This layer is responsible for generating predictions based on the sequential data processed by the LSTM layers. To optimize the model’s performance during training, it is compiled with the Adam optimizer, and the model is trained to minimize the mean squared error (MSE) loss, a suitable choice for regression problems. Additionally, to mitigate overfitting or low efficiency problems during the training process, a 5-fold cross-validation technique was applied to the time series analysis model.

### III. RESULTS

#### 3.1. Patterns of Traffic Congestion in the Atlanta Downtown Connector

Table 1 shows descriptive statistics for traffic congestion amount in the Atlanta Downtown Connector.

<table>
<thead>
<tr>
<th></th>
<th>Northbound</th>
<th></th>
<th>Southbound</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>0.1769</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation</td>
<td>0.2085</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>0</td>
<td>Minimum</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>1.01</td>
<td>Maximum</td>
</tr>
</tbody>
</table>

Table 1. Descriptive statistics for traffic congestion amount (unit: weighted mileHour).

In this study, weighted mileHours was used as the unit to represent traffic congestion levels. This metric combines weights based on the traffic congestion levels as depicted in Google Maps, as explained in Section 2. When examining the Downtown Connector separately for the northbound and southbound directions, for the northbound, the average congestion amount is 0.1769 mileHours, with a maximum congestion amount of 1.01 mileHours. In the southbound, the average congestion amount is 0.1043 mileHours, with a maximum congestion amount of 1.008. In this case, the values presented in Table 1 represent traffic congestion amounts sampled every 10 minutes.

As mentioned in Section 2, traffic volume exhibits seasonality with a periodicity of 24 hours and a weekly cycle. Figure 4 illustrates the traffic congestion amounts from April 2022 to March 2023, along with the data trendline, demonstrating the seasonality of traffic congestion.

According to Figure 4, the traffic congestion volume in the northbound direction is higher than that in the southbound direction. When comparing the traffic congestion patterns over the week, congestion is most severe on Thursdays and Fridays, while during the weekends, congestion tends to be less noticeable compared to weekdays. In the next step, these data were used to predict traffic congestion for the week following March 31, 2023.

#### 3.2. Prediction of Traffic Congestion

For predicting traffic congestion in the Atlanta Downtown Connector, the multivariate time-lagged LSTM model was executed, using time variables and traffic congestion amounts as input data. Model training and predictions were carried out separately for both northbound and southbound directions, and the results are as follows. Model training consisted of 20 epochs at each cross-validation step, totaling 100 training iterations. First, regarding the model training results for northbound, the training process took approximately 12 minutes and 48 seconds, and the average MSE across the 5 cross-validation runs was approximately 0.0015. For the southbound model training, it took approximately 12 minutes and 13 seconds, and the average MSE for cross-validation was around 0.0009. The subsequent step encompassed utilizing the trained models to carry out experiments for forecasting traffic congestion from Saturday, April 1, 2023, to Friday, April 7, 2023. The outcomes of the traffic congestion predictions and observed values are depicted in Figure 5 below.

Both the northbound and southbound predictions yielded results that closely resembled the existing traffic congestion patterns. A shared characteristic was the presence of peaks during the morning and afternoon on weekdays, reflecting rush hours from Monday to Friday. However, this was not the case on weekends because there are fewer commuters compared to weekdays. When comparing the observations and predictions, the graph of the predictions closely follows the pattern of the observations, but does not show as much details as the observations. This resulted from differences in the tuning of the parameters and training of the LSTM model. Calculating the root mean square error (RMSE) for the observed and predicted values, it is 0.1585 for the northbound and 0.1047 for the southbound, respectively.
In order to explore the feasibility of the results derived from the time-series prediction model, the patterns of predicted values were compared with typical traffic information provided by Google Maps. Figure 6 below shows typical traffic information provided by Google Maps, with examples taken from Thursday at 8:30 AM and 4:00 PM. It is evident in both the predicted results from Figure 5 and the typical traffic in Figure 6 that traffic congestion is more severe in the afternoon compared to the morning. This can be attributed to commuters staggering their morning commute to avoid the rush hour, whereas in the evening, commuting times tend to cluster, resulting in more severe congestion. Furthermore, during daytime, the downtown area experiences a significant influx of people, contributing to even more severe congestion during the afternoon rush hour.

The difference between the historical data constructed in this study and the prediction results lies in the fact that while the historical data recorded a maximum traffic congestion amount of approximately 1, the prediction results only forecasted a maximum amount slightly above 0.4. This suggests that the LSTM model, which predicts based on long-term data, might have reflected the characteristics of losing some information during the training and prediction processes. This issue can potentially be addressed during the model construction step by adjusting the hyperparameters of LSTM, exploring the combination with other model architectures such as CNN, or considering the utilization of different time series analysis algorithms in the future study.

IV. DISCUSSION AND CONCLUSIONS

As cities grow and populations surge, the intricate transportation systems become more critical than ever. This study delved into the realm of time-series machine learning to forecast congestion in the heavily congestion road section and suggested a comprehensive framework that includes data acquisition through web map service, Google Maps, the calculation of a traffic congestion index considering multiple congestion factors simultaneously, and traffic congestion prediction utilizing this congestion index. To perform predictions using LSTM, it is essential to consider the periodicity of traffic congestion. Therefore, in this study, alongside traffic congestion data in mileHours, the transformation of 24 hours in a day and 7 days in a week was implemented into sine and cosine functions as input variables for model training and traffic congestion prediction.

To summarize the model training and prediction results, the predictions for the traffic congestion patterns during the week following April 1, 2023, were consistent with the historical data. Thursday and Friday exhibited the highest levels of traffic congestion, and on weekdays, traffic volume peaks were observed during the morning and afternoon rush hours. Furthermore, when compared to the typical traffic provided by Google Maps, it was observed that the actual traffic congestion patterns closely resembled the predicted traffic congestion patterns. The model training showed errors of approximately 0.0015 and 0.0009 for the northbound and southbound models, respectively. In this study, a time series prediction model was constructed using traffic congestion measured in mileHours unit and temporal variables. Assuming the procedure of this study is applied to a different research area, data collection for the selected region can be achieved through Google Maps. If the data and the pixel size of obtained images are collected, it opens up possibilities for research in other regions or comparative analyses between different cities.
Figure 5. Prediction results of traffic congestion in the Atlanta Downtown Connector.
However, a couple of limitations were identified during this study. Firstly, the predicted traffic congestion values in the predictions were lower than the actual traffic congestion values observed in the Atlanta Downtown Connector. Secondly, when examining the predicted results graph in Figure 5, it is noticeable that the congestion graph appears smoother than the actual traffic congestion graph. These are attributed to limitations inherent in the LSTM algorithm and should be addressed in future research through adjustment to hyperparameters or algorithm modifications. Recently, various algorithms for time series prediction, aside from LSTM, have been developed. For instance, XGBoost, LightGBM regression models using concepts of gradient boosting can be employed for traffic volume prediction. Additionally, a model known as Transformer (Vaswani et al., 2017) has gained attention as an alternative to LSTM. Transformer addresses one of LSTM’s drawbacks, which is the difficulty of parallel processing due to sequential data input. Transformer is capable of preserving important information effectively during the training process.

This study holds the following implications. Firstly, it addresses the issues associated with traditional sensor-based data collection by utilizing readily accessible online web map services for data collection. Additionally, it incorporates the temporal characteristics of traffic data as explanatory variables in the prediction model. The model’s performance effectively captures existing traffic congestion patterns, making it a valuable tool for traffic congestion prediction. This research may assist traffic researchers and stakeholders involved in congestion management and serve as a reference for citizens planning their travel routes on real roads.

ACKNOWLEDGEMENTS

Author thanks Dr. Jeong Seong at University of West Georgia for support on the research design and improvement of this paper.

REFERENCES


Bhaskaran, K., Gasparini, A., Hajat, S., Smeeth, L., &
Kim: Prediction of Traffic Congestion Using a Time-Series Model and Sp

Prediction of Traffic Congestion Using a Time-Series Model and Spatiotemporal Data: A Case Study of the Atlanta Downtown Connector


**Yunsik Kim**, a master’s graduate from Dongguk University, specializes in GeoAI for analyzing spatiotemporal dynamics of urban environments. As a visiting scholar at the University of West Georgia, he involved in a project on traffic big data analysis, employing the GeoAI techniques.