Project 2211 | March 2023

SJSU SAN JOSÉ STATE UNIVERSITY

MTI MINETA TRANSPORTATION INSTITUTE

Attention-Based Data Analytic Models for Traffic Flow Predictions

Kaushal Kumar
Dr. Yupeng Wei

Mineta Transportation Institute

Founded in 1991, the Mineta Transportation Institute (MTI), an organized research and training unit in partnership with the Lucas College and Graduate School of Business at San José State University (SJSU), increases mobility for all by improving the safety, efficiency, accessibility, and convenience of our nation's transportation system. Through research, education, workforce development, and technology transfer, we help create a connected world. MTI leads the Mineta Consortium for Transportation Mobility (MCTM) funded by the U.S. Department of Transportation and the California State University Transportation Consortium (CSUTC) funded by the State of California through Senate Bill 1. MTI focuses on three primary responsibilities:

Research

MTI conducts multi-disciplinary research focused on surface transportation that contributes to effective decision making. Research areas include: active transportation; planning and policy; security and counterterrorism; sustainable transportation and land use; transit and passenger rail; transportation engineering; transportation finance; transportation technology; and workforce and labor. MTI research publications undergo expert peer review to ensure the quality of the research.

Education and Workforce

To ensure the efficient movement of people and products, we must prepare a new cohort of transportation professionals who are ready to lead a more diverse, inclusive, and equitable transportation industry. To help achieve this, MTI sponsors a suite of workforce development and education opportunities. The Institute supports educational programs offered by the Lucas Graduate School of Business: a Master of Science in Transportation Management, plus graduate certificates that include High-Speed and Intercity Rail Management and Transportation Security Management. These flexible programs offer live online classes so that working transportation professionals can pursue an advanced degree regardless of their location.

Information and Technology Transfer

MTI utilizes a diverse array of dissemination methods and media to ensure research results reach those responsible for managing change. These methods include publication, seminars, workshops, websites, social media, webinars, and other technology transfer mechanisms. Additionally, MTI promotes the availability of completed research to professional organizations and works to integrate the research findings into the graduate education program. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

Report 22-63

Attention-Based Data Analytic Models for Traffic Flow Predictions

Kaushal Kumar

Yupeng Wei, PhD

March 2023

A publication of the
Mineta Transportation Institute
Created by Congress in 1991

College of Business
San José State University
San José, CA 95192-0219

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
22-63		

4. Title and Subtitle	5. Report Date
Attention-Based Data Analytic Models for Traffic Flow Predictions	March 2023
	6. Performing Organization Code
7. Authors Kaushal Kumar Yupeng Wei, PhD: ORCID:0000-0001-5070-3300	**8. Performing Organization Report** CA-MTI-2211
9. Performing Organization Name and Address Mineta Transportation Institute College of Business San José State University San José, CA 95192-0219	**10. Work Unit No.** **11. Contract or Grant No.** 69A3551747127
12. Sponsoring Agency Name and Address U.S. Department of Transportation Office of the Assistant Secretary for Research and Technology University Transportation Centers Program 1200 New Jersey Avenue, SE Washington, DC 20590	**13. Type of Report and Period Covered** **14. Sponsoring Agency Code**

15. Supplemental Notes

16. Abstract

Traffic congestion causes Americans lose millions of hours and dollars each year. In fact, 1.9 billion gallons of fuel are wasted each year due to traffic congestion, and each hour stuck in traffic costs about $21 in wasted time and fuel. The traffic congestion can be caused by various factors, such as bottlenecks, traffic incidents, bad weather, work zones, poor traffic signal timing, and special events. One key step to addressing traffic congestion and identifying its root cause is an accurate prediction of traffic flow. Accurate traffic flow prediction is also important for the successful deployment of smart transportation systems. It can help road users make better travel decisions to avoid traffic congestion areas so that passenger and freight movements can be optimized to improve the mobility of people and goods. Moreover, it can also help reduce carbon emissions and the risks of traffic incidents. Although numerous methods have been developed for traffic flow predictions, current methods have limitations in utilizing the most relevant part of traffic flow data and considering the correlation among the collected high-dimensional features. To address this issue, this project developed attention-based methodologies for traffic flow predictions. We propose the use of an attention-based deep learning model that incorporates the attention mechanism with Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks. This attention mechanism can calculate the importance level of traffic flow data and enable the model to consider the most relevant part of the data while making predictions, thus improving accuracy and reducing prediction duration.

17. Key Words	18. Distribution Statement
Traffic congestion Data analysis Machine learning Transportation Predictive Models	No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	28	

Form DOT F 1700.7 (8-72)

DOI: 10.31979/mti.2023.2211

Mineta Transportation Institute
College of Business
San José State University
San José, CA 95192-0219

Tel: (408) 924-7560
Fax: (408) 924-7565

Email: mineta-institute@sjsu.edu
transweb.sjsu.edu/research/2211

Acknowledgments

The work was sponsored in part by the Mineta Transportation Institute. A portion of the research findings has been published in *Sensors* 22.20 (2022): 7994, entitled "Convolutional Long-Short Term Memory Network with Multi-Head Attention Mechanism for Traffic Flow Prediction."

CONTENTS

List of Figures

List of Tables

1. Introduction

One of the problems that many of us who live in urban areas deal with is traffic. The growing urban population is one of the primary factors that contributes to traffic. There is an influx of residents looking for work and opportunity, leading to serious traffic congestion problems. Due to this congestion, there is an increase in fuel consumption which leads to a rise in carbon emissions and air pollution. This also results in additional expenses in terms of time and money. To alleviate traffic congestion, the accurate prediction of traffic flow is crucial. Traffic flow prediction involves analyzing data from various sources, such as road sensors, GPS, and historical traffic patterns, to forecast the movement of vehicles on the road network. This information can then be used to optimize traffic management strategies, such as adjusting traffic signals and diverting traffic to less-congested routes. By accurately predicting traffic flow, traffic congestion can be reduced, resulting in smoother traffic, less fuel consumption, and a reduction in carbon emissions and air pollution. Additionally, reducing traffic congestion can also save time and money for commuters and improve overall road safety. In conclusion, traffic flow prediction is an essential tool in reducing traffic congestion and its associated negative impacts.

Deep learning has recently received significant attention from both academia and industry in traffic flow predictions for several reasons [2, 3]. First, deep learning algorithms can analyze large amounts of data, including both structured and unstructured data, to make more accurate predictions. This is particularly important in traffic flow prediction, as real-time traffic information can be complex and unpredictable. Second, deep learning algorithms can process data in real-time, making predictions faster and more efficient. This is important for traffic flow prediction, as traffic conditions can change rapidly, and a quick response is necessary to manage traffic effectively. Third, deep learning algorithms can automate the process of traffic flow prediction, reducing the need for manual intervention and improving the accuracy of predictions. While promising, current deep learning methods, such as artificial neural networks (ANNs) and long short-term memory (LSTM), have limitations in utilizing the most relevant part of traffic flow data and considering the correlation among the collected high-dimensional features. This results in longer prediction times and less accurate forecasts.

To address this issue, we propose the use of an attention-based deep learning model that incorporates the attention mechanism with Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks. This attention mechanism can calculate the importance level of traffic flow data and enable the model to consider the most relevant part of the data while making predictions, thus improving accuracy and reducing prediction duration. To prove the efficacy of our proposed attention-based deep learning model, we conduct experiments using traffic flow data from the Caltrans Performance Measurement System (PeMS). Our results indicate that the proposed model can make traffic flow predictions with a high degree of accuracy. The following outlines the structure of the remaining work. Section II reviews the existing data-driven methods for traffic predictions. In Section III, we outline the data source, perform a data exploratory analysis, and undergo data pre-processing. Section IV discusses the machine learning methods applied to train the prediction models, and finally, Section V presents the numerical results of the case study, draws conclusions, and outlines future work.

2. Literature Review

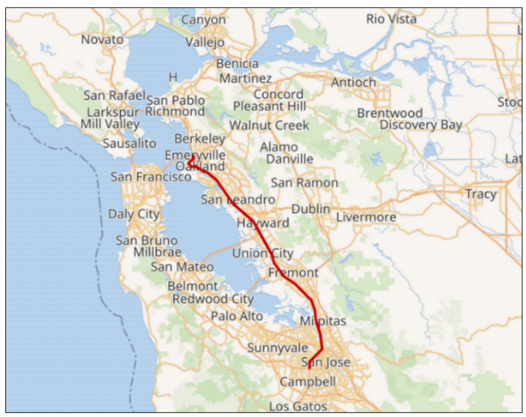
Numerous data-driven methods have been extensively researched and developed for the purpose of traffic flow prediction. These methods utilize various data sources, including road sensors, GPS, and historical traffic patterns, to make predictions about the movement of vehicles on the road network. For example, Hamed et al. [4] utilized an autoregressive integrated moving average (ARIMA) model in their traffic prediction analysis. It predicted the future traffic flows based on the earlier traffic flows. In order to increase prediction accuracy, their research recommended smoothing time series data using lagged moving averages. Davis and Nihan [5] used the K nearest neighbors (KNN) to capture the linear behavior of the historical traffic flow data. A dynamic multi-interval traffic volume forecast model based on KNN nonparametric regression was reported by Chang et al. [10]. Another researcher, El Faouzi [6], used functional estimating techniques to create a kernel smoother for the autoregression function in order to estimate short-term traffic flow. To predict future traffic, Sun et al. [7] developed a local linear regression model. The forecasting of traffic flow was suggested using a Bayesian network technique. For predicting short-term traffic flow, a weighted support vector regression (SVR) technique was presented. Some studies investigate hybrid approaches, which combine different procedures to create adaptive models. Tan et al. [8] claimed that it was possible to estimate traffic flow by combining the moving average (MA) along with other models. The data was integrated using the belief rule-based algorithm, and the resulting traffic data was then used to inform the methodology that was recommended for estimating traffic flow. Chen et al.'s [9] novel GNN methodology for estimating traffic flow distinguished between the density of connected roads using a weighted undirected network. To simulate traffic propagation, a simulation model was developed, and the outcomes were included in the GNN model to predict traffic flow in real-time.

Although many data analysis methods have been investigated to predict traffic flow under different circumstances, these methods have limitations in utilizing the most relevant part of traffic flow data and considering the correlation among the collected high-dimensional features. This results in longer prediction times and less accurate forecasts. To address this issue, we propose the use of an attention-based deep learning model that incorporates the attention mechanism with Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks. This attention mechanism can calculate the importance level of traffic flow data and enable the model to consider the most relevant part of the data while making predictions, thus improving accuracy and reducing prediction duration.

3. Data Description

The Caltrans Performance Measurement System (PeMS) dataset is frequently used for predicting traffic flow. The data acquired from the PeMS database was used for the proposed deep learning model. Traffic data consisting of vehicle miles and vehicle hours was collected every hour for the US I880 freeway. These data were collected by road sensors, the physical devices installed on roads that collect data on traffic volume, speed, and occupancy. In this study, the experiments were conducted using traffic flow data that was gathered during the weekdays of the first eight months of 2022. The first seven months' worth of data was chosen as the training set, and the final month's data was chosen as the test set. The data for the training and test sets were both standardized to prevent generality loss. The data for this study is normalized using data normalization method to ensure that the scales for vehicle miles traveled (VMT) and vehicle hours traveled (VHT).

Figure 1. Location of Freeway 1880 on a Map



The data was obtained from PeMS, a collection of numbers of vehicles at two junctions at an hourly frequency. The CSV file provides four features:

- Date and time
- Lane points
- VHT
- VMT

The traffic data comes from several periods because the sensors on each of these lane locations gathered data at various intervals. Several of the lane points only provided scant or incomplete data. Table 1 describes the output of the data through a python file.

Table 1. Detailed Description of Dataset

Date Time	VMT	VHT	Lane Points	%Observed
4/1/2022 0:00	2156	33	16764	38.4
4/1/2022 1:00	1554	24	16764	38.5
4/1/2022 2:00	1435	22	16764	38.5
4/1/2022 3:00	1760	27	16764	38.5
4/1/2022 4:00	3151	47	16764	38.5

Table 2 presents a summary of the statistics on the gathered traffic flow data together with some statistical analysis. This includes data from Saturday and Sunday, in addition to public holidays. The volume of freeway traffic and the overall number of data points are the figure's minimum and count, respectively.

Table 2. Basic Statistical Analysis of Dataset

SL No	VMT	VHT
Count	4325.000	4325.000
Mean	7524.762	123.633
Std	3665.674	65.015
Min	902.000	13.000
25%	3662.000	55.000
50%	8997.000	144.000
75%	10766.00	180.000
Max	12288.000	236.000

In order to train the model, we employed historical load data, timestamp data on the number of vehicles traveling on US Interstate 880 in California, feature extraction from the datasets provided, and historical load data. Table 3 presents the descriptive analysis of the dataset. We specifically considered all factors that describe date and time, such as a month, day, hour, day of the week, and holiday, because traffic intensity tends to alter with time and date. Hour, day, and month forms are all in sequence format, which AI systems can directly utilize.

Table 3. Descriptive Analysis of Datasets

Date Time	VHT	Lane points	Year	Month	Date no	Hour
2022-04-01 00:00	33	16764	2022	4	1	0
2022-04-01 01:00	24	16764	2022	4	1	1
2022-04-01 02:00	22	16764	2022	4	1	2
2022-04-01 03:00	27	16764	2022	4	1	3
2022-04-01 04:00	47	16764	2022	4	1	4

Figures 2 and 3 represent the VHT data from the python output of traffic flow in two lane points on different days of the month and days of the week, respectively.

Figure 2. Graphical Presentation of VHT Traffic Flow by Day of the Month.

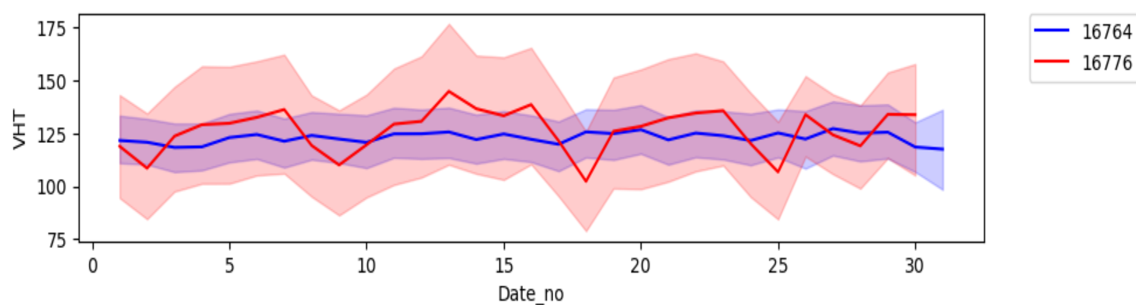
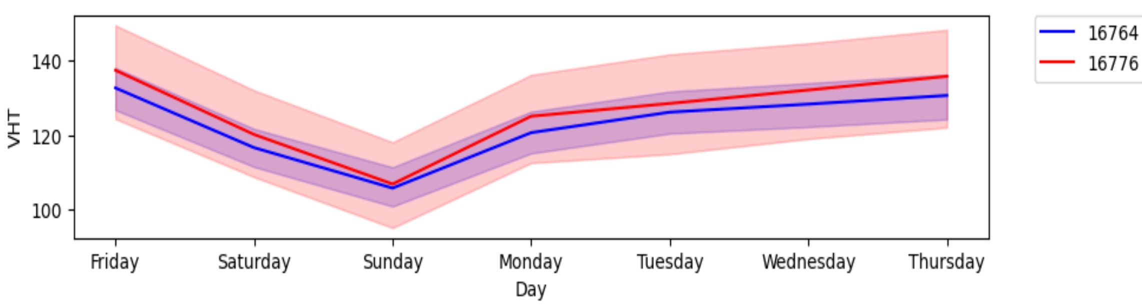
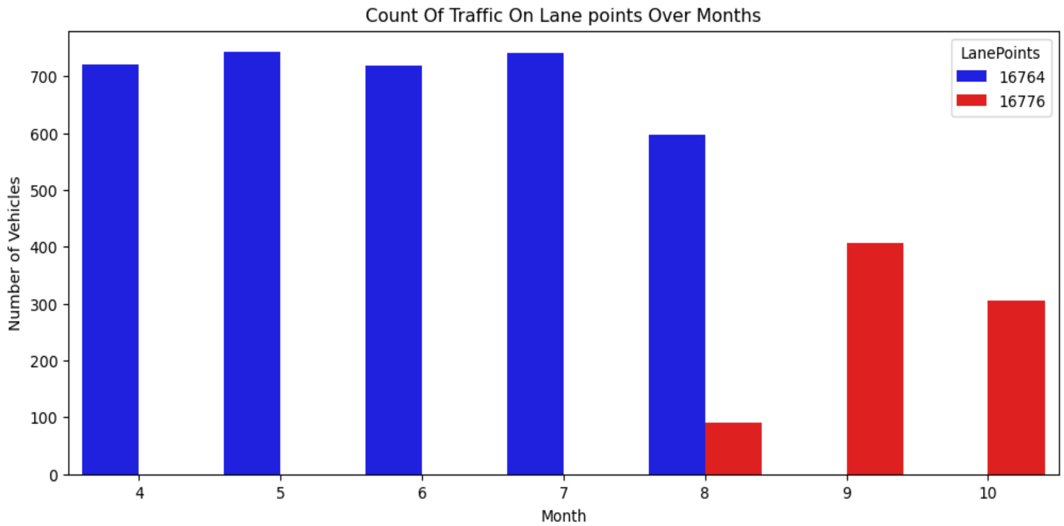


Figure 3. Graphical Presentation of VHT Traffic flow by Day of the Week



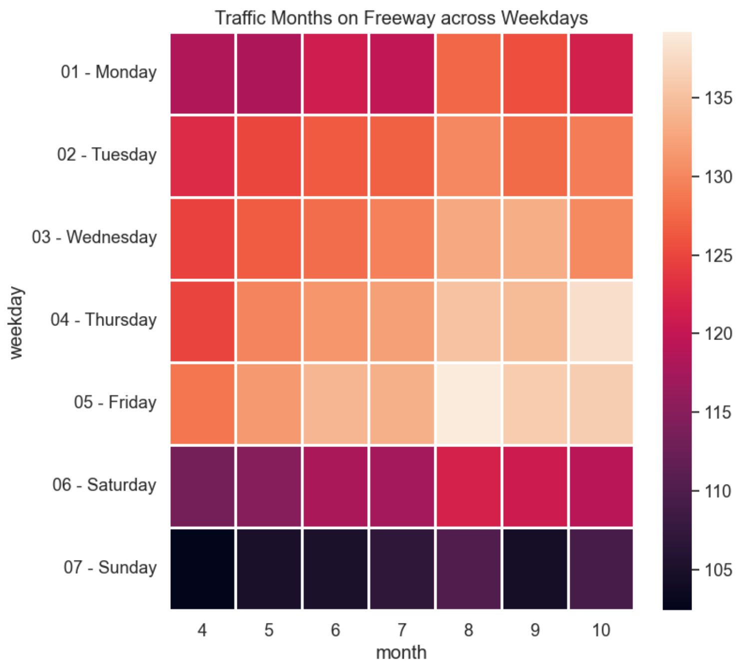
The trend of the first lane points in Figure 4 appears to be uniform throughout the day of the month. However, data for the second lane points only becomes available starting from September. To determine the presence of seasonality, a closer examination of the date and time composition is necessary. Figure 4 also reveals that Lane 1 has a higher vehicle count compared to Lane 2.

Figure 4. Traffic Flow Data Month Wise in two Lane Points



The following finds are observed based on Figure 5. All two-lane points trend upward on a monthly basis. We can see that around June, there is an increase in the first and second lane points, and we assume this may be related to summer vacation and related activities. There is good consistency in the data on a monthly basis across all dates. We may observe that there are peaks in the morning and evening, and a fall in activity during the night, for a given day. Due to fewer automobiles on the roads on Sundays than other days of the week, traffic is less congested. The traffic is steady from Monday through Friday. All the two-lane points' data did not cover the same time period. Lane Points 1 and 2 have different monthly trends. Lane Point 1 has a stronger weekly seasonality than Lane Point 2.

Figure 5. Traffic Flow Data Month Wise in two Lane Points



4. Methodology

We integrated the attention mechanism with the Gated Recurrent Unit (GRU) and Long-Term Short Memory (LSTM) for traffic flow predictions. The attention mechanism is a type of neural network architecture that selectively focuses on key parts of the input to make predictions. It is optimized for handling inputs with varying lengths and assigns varying levels of importance to different parts of the sequence. The mechanism involves encoding the input, calculating attention weights, generating context vectors, and decoding the output. Attention weights determine the significance of each input element for the prediction, while context vectors are a fusion of the input and attention weights. A simplified illustration of the attention mechanism can be seen in Figure 6. Figure 7 shows an example of the integration of the attention mechanism with the GRU model. More details about the integration of the attention mechanism with deep learning models can be found in [1].

Figure 6. A Simplified Illustration of the Attention Mechanism

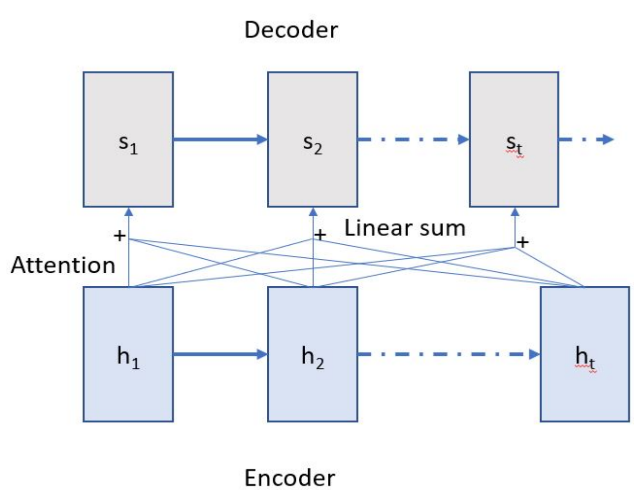
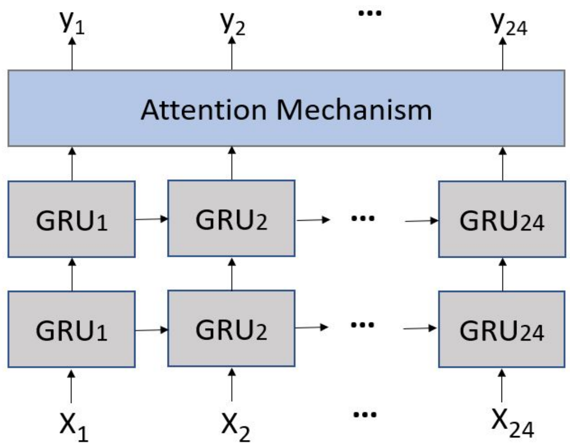


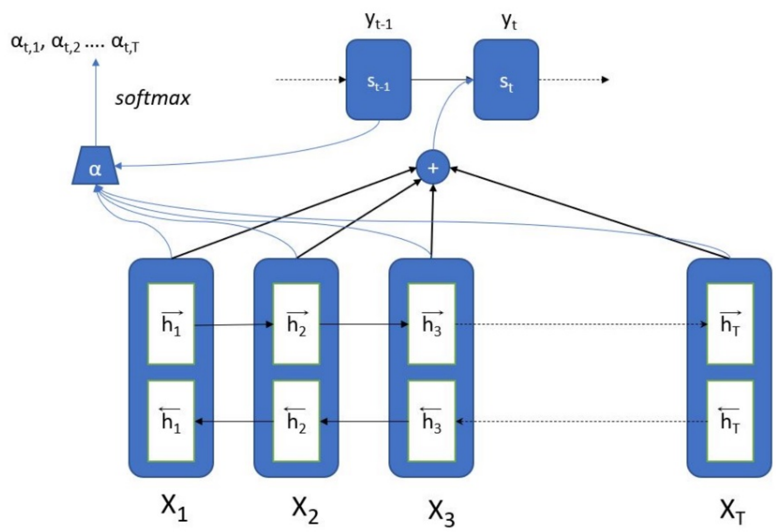
Figure 7. An Example of the Integration of the Attention Mechanism with the GRU Model



4.1 Attention-based Gated Recurrent Unit

The attention mechanism can be integrated with the GRU model by using the attention weights to weight the hidden state vectors produced by the GRU model and generate a context vector that represents the relevant information from the input sequence. This context vector can then be concatenated with the hidden state of the GRU and used as input to a fully connected layer to produce the final prediction. The attention weights are typically computed using a feedforward neural network that takes the current hidden state of the GRU and the input sequence as input and outputs the attention weights. Figure 8 shows an example of the integration of the attention mechanism with the GRU.

Figure 8. The Integration of the Attention mechanism with the GRU



4.2 Attention-based Long Short-Term Memory (LSTM)

The Long Short-Term Memory (LSTM) model developed by Hochreiter and Schmidhuber was one of the first and most effective methods for handling vanishing gradients. Figure 9 shows a basic LSTM computation cell. The attention mechanism can be integrated with the LSTM model in a similar way as it is integrated with the GRU model. The main difference is that the attention weights are computed using the hidden state and cell state of the LSTM instead of just the hidden state of the GRU.

Figure 9. The basic LSTM Computational Cell

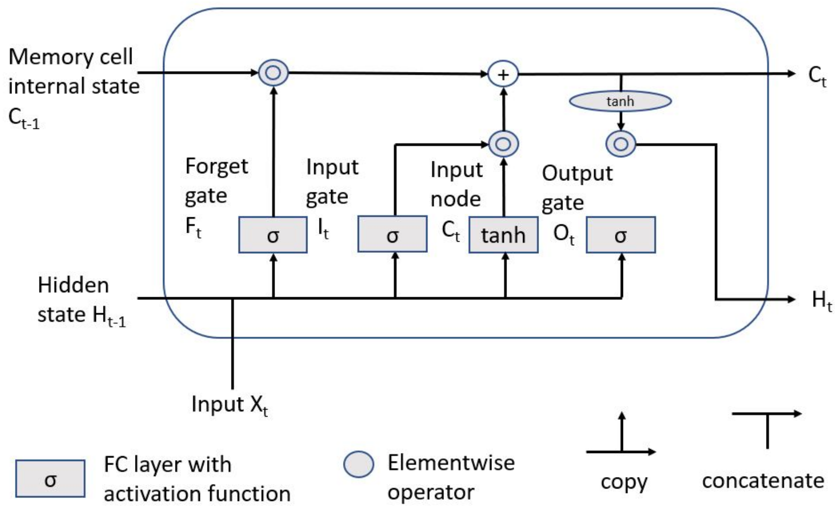
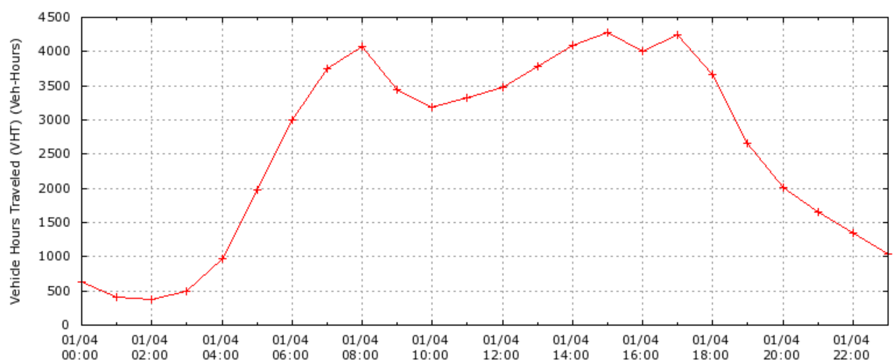
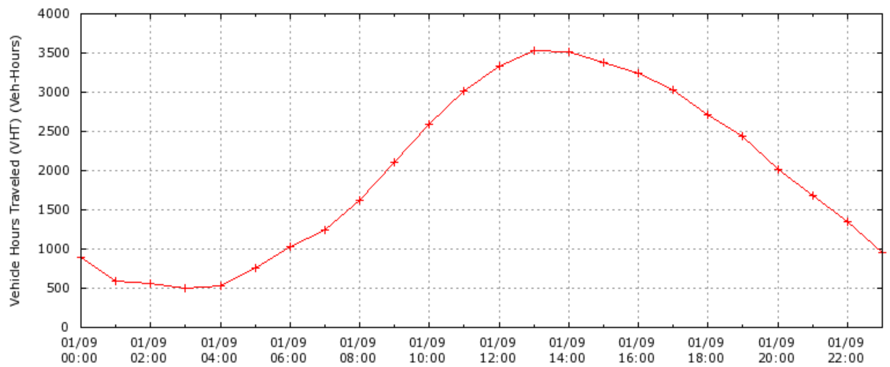


Figure 10. Traffic Flow over Time for Weekdays



Figures 10 and 11 show the flow of traffic over weekdays and weekends, respectively. These figures are collected from the data of Caltrans traffic flow over the study period. From them, we can see that VHT and VMT data significantly increased during the daytime.

Figure 11. Traffic Flow over Time for Weekends



The suggested model is evaluated using three performance indices: mean absolute error (MAE), mean square error (MSE), and root mean squared error (RMSE). They are defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}|$$

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n}}$$

where y_i is the observed traffic flow, $\hat{y}$ is the predicted traffic flow, and n is the number of samples.

5. Results & Future Work

The Figures 12 and 13 present the comparison between the predicted traffic flow and the actual traffic flow in two lanes of a selected highway using the attention-based GRU model. Upon observing these figures, it can be concluded that the proposed attention-based GRU model exhibits a remarkable capability in capturing the fluctuations of the traffic flow. Additionally, the model demonstrates a relatively high level of accuracy in predicting the traffic flow. The attention mechanism incorporated into the GRU model allows it to focus on the most relevant information in the input sequence, thereby improving the accuracy of the prediction. As a result, the attention-based GRU model is capable of accurately tracking changes in traffic flow trajectory and making predictions with a high level of precision.

Figure 12. GRU Model Output for the First Lane

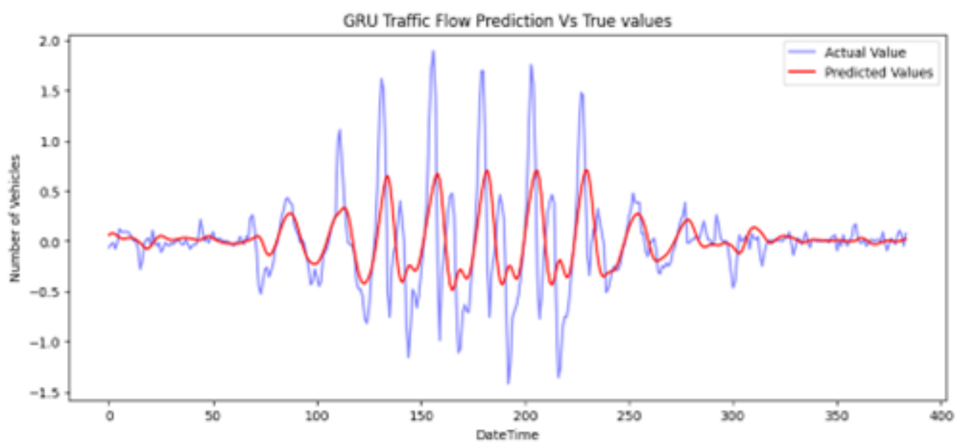
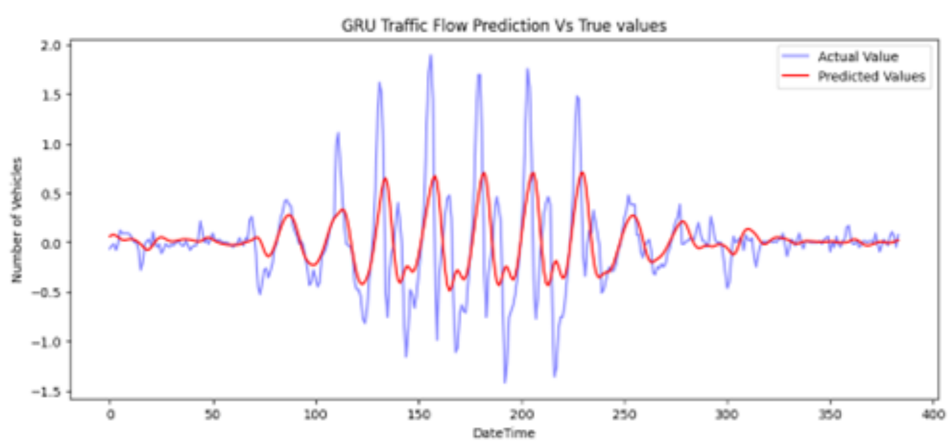


Figure 13. GRU Model Output for the Second Lane



The comparison between the predicted and actual traffic flow in two lanes of a chosen highway is displayed in Figures 14 and 15, which were generated using an attention-based LSTM model. Upon reviewing these visualizations, it becomes evident that the attention-based LSTM model has a remarkable ability to accurately depict the fluctuations in traffic flow. This is a testament to the model's effectiveness in capturing and predicting the complex patterns in real-world traffic data. Additionally, the comparison between the two models also demonstrates that the attention-based LSTM model surpasses the attention-based GRU model. The predicted traffic flow from the attention-based LSTM model is found to be more closely aligned with the actual traffic flow, as opposed to the predicted traffic flow from the attention-based GRU model, which is less accurate. This highlights the superiority of the attention-based LSTM model in predicting traffic flow patterns.

Figure 14. LSTM Model Prediction for the First Lane

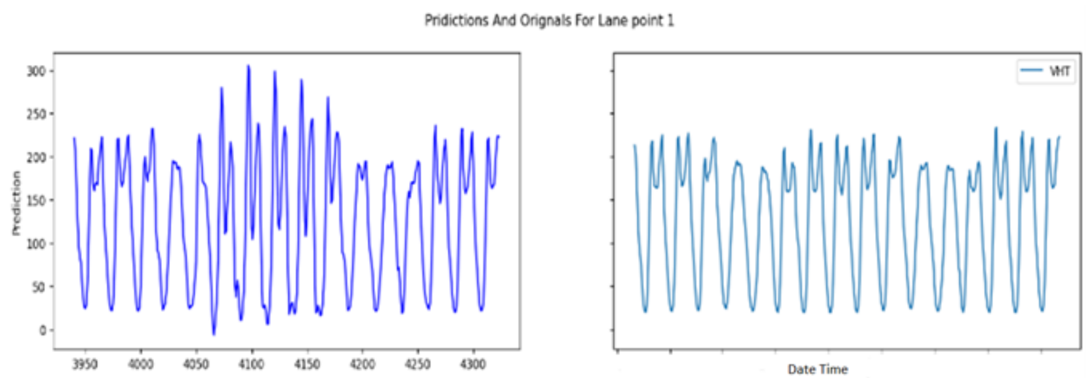
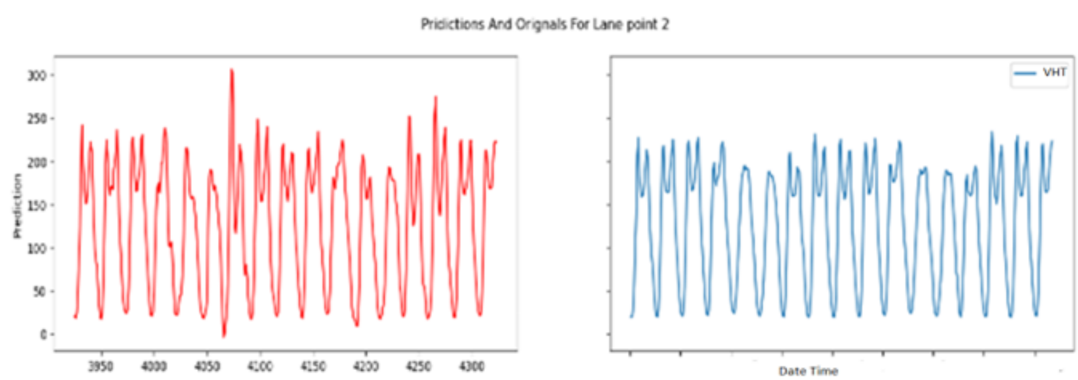


Figure 15. LSTM Model Prediction for the Second Lane



In this study, we aimed to enhance traffic flow prediction by integrating attention mechanism into GRU and LSTM models. The attention mechanism calculates the significance of traffic data and allows the model to concentrate on the most crucial information while making predictions, leading to improved accuracy and faster prediction times. To validate the effectiveness of the proposed attention-based deep learning model, we conducted experiments using data from PeMS. Our findings demonstrate that the proposed model can accurately predict traffic flow. Additionally, our results show that the attention-based LSTM model outperforms the attention-based GRU model in terms of traffic flow predictions. In future work, we plan to extend our research to predicting traffic flow in more complex conditions.

Bibliography

1. Wei, Yupeng, and Hongrui Liu. "Convolutional Long-Short Term Memory Network with Multi-Head Attention Mechanism for Traffic Flow Prediction." *Sensors* 22.20 (2022): 7994.

2. Jiang, Luofeng. "Traffic Flow Prediction Method Based on Deep Learning." *Journal of Physics: Conference Series* Vol. 1646. No. 1. IOP Publishing, 2020.

3. Kang, Zifeng, et al. "Learning dynamic graph embedding for traffic flow forecasting: A graph self-attentive method." *2019 IEEE Intelligent Transportation Systems Conference* (ITSC). IEEE, 2019.

4. Hamed, Mohammad M., Hashem, Al-Masaeid R., and Zahi, Bani Said M. "Short-term prediction of traffic volume in urban arterials." *Journal of Transportation Engineering* 121.3 (1995): 249–254.

5. Davis, Gary A., and Nihan, Nancy L. "Nonparametric regression and short-term freeway traffic forecasting." *Journal of Transportation Engineering* 117.2 (1991): 178–188.

6. El Faouzi, Nour-Eddin. "Nonparametric traffic flow prediction using kernel estimator." *Transportation And Traffic Theory*. Proceedings of the 13th International Symposium on Transportation and Traffic Theory, Lyon, France, 24–26 July 1996. 1996.

7. Sun, Hongyu, et al. "Use of local linear regression model for short-term traffic forecasting." *Transportation Research Record* 1836.1 (2003): 143–150.

8. Tan, Man-Chun, et al. "An aggregation approach to short-term traffic flow prediction." *IEEE Transactions on Intelligent Transportation Systems 10.1* (2009): 60–69.

9. Chen, Chenyi, et al. "The retrieval of intra-day trend and its influence on traffic prediction." *Transportation research Part C: Emerging Technologies* 22 (2012): 103–118.

10. Chang, H., et al. "Dynamic near-term traffic flow prediction: system-oriented approach based on past experiences." *IET Intelligent Transport Systems* 6.3 (2012): 292–305.

About the Authors

Kaushal Kumar

Kaushal Kumar began pursuing his MS degree in Industrial and Systems Engineering at San Jose State University in 2022. He is currently working with National Renewable Energy Laboratory (NREL) as a battery supply chain intern.

Dr. Yupeng Wei

Dr. Wei is an Assistant Professor in Industrial and Systems Engineering at San Jose State University (SJSU). His research interests include optimization, machine learning, and data analytics and their applications in quality monitoring, anomaly detection, and degradation prognostics.

MTI FOUNDER

Hon. Norman Y. Mineta

MTI BOARD OF TRUSTEES

Directors