A Deep Learning Model to Forecast the Impact of COVID-19 on Traffic Demand in Salt Lake County

Aaron Wang

Skyline High School

11

Introduction

- Since early March 2020, the global COVID-19 pandemic has had an enormous impact on various aspects of society. In the United States, there have been over 70.7 million cases and 866 thousand deaths, as of January 23, 2022. In addition to illness and loss of life, the pandemic has had a significant effect on traffic across the United States. During the early stages of the pandemic, traffic was significantly reduced due to travel restrictions imposed by governments, fear of getting sick, lower levels of economic and social activity, and shutdowns of schools and businesses. Later on, as the pandemic progressed, traffic volume demonstrated a gradually increasing trend. It is clear that the COVID pandemic has a significant effect on traffic demand.
- This project seeks to predict the impact of COVID-19 on vehicular traffic in Salt Lake County during the pandemic. Traffic data from freeways in Salt Lake County will be used to quantify the traffic patterns. Deep learning models will be developed using data that could influence traffic demand to forecast the traffic demand patterns in the near future, using a novel approach combining machine learning with graph theory. Traditional machine learning forecasting methods will also be tested. These models will potentially be able to help transportation agencies prepare for changes in traffic demand in the near future.

Methodology

- This project focuses on freeway traffic patterns in Salt Lake County. County-wide vehicle miles traveled (VMT) of freeways within the county were used to quantify traffic flow. The VMT data was collected from the UDOT Performance Measurement System (PeMS) from January 2019 to the last week of November 2021, providing traffic data before and during the pandemic.

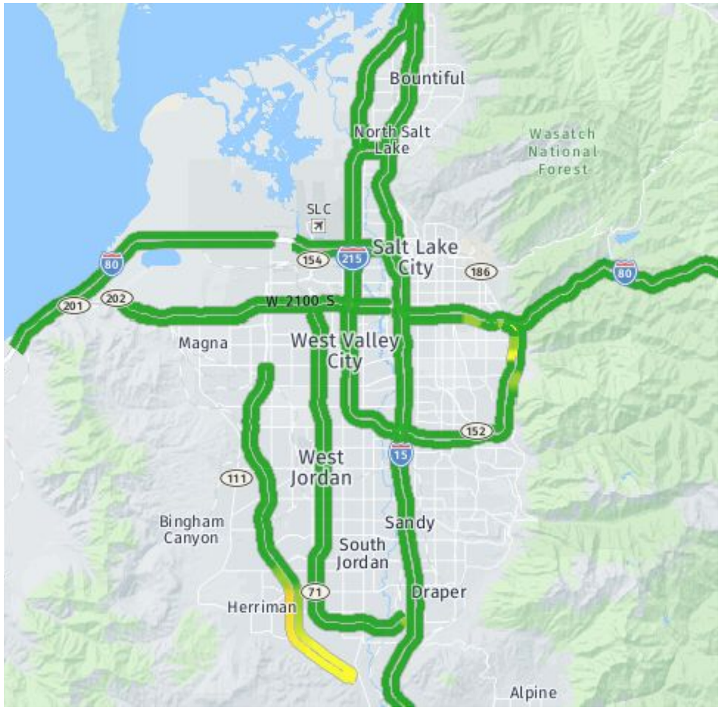

FIGURE 5. Freeways in Salt Lake County

Data

- Various factors related to the pandemic and vehicular traffic were collected as explanatory variables.
- Daily new COVID-19 cases and the percentage of fully vaccinated individuals over five years of age were collected from the Utah Department of Health.
- The daily news sentiment index, obtained from the Federal Reserve Bank of San Francisco, is based on analysis of economic-related news from 24 major U.S. newspapers, providing information about overall sentiment as a quantifiable number. The monthly unemployment rate of Salt Lake County was obtained from the Utah Department of Workforce Services.
- Weather conditions, including temperature, precipitation, and snow depth were collected from the National Oceanic and Atmospheric Administration. The weather observing station at the Salt Lake International Airport was selected for weather data in Salt Lake County.
- Pandemic-related policies are another important factor. State and county-level policies, including stay at home orders, mask mandates, school closings, and a state of emergency, can restrict or influence travel. If a policy was effective at a certain time, a dummy variable was set to 1; otherwise, it was set to 0.
- All data were aggregated by week. Economic and weather data were collected from January 2019 to the last week of November 2021, while pandemic related data and policies were collected once they became available until the last week of November 2021.

Prediction Model: LSTM

- A Long Short-Term Memory (LSTM) model is a type of recurrent neural network (RNN) that can handle long-term memory better than traditional RNNs. It is commonly used for time series forecasting. An LSTM has a chain of repeating modules of neural networks, with each module containing a cell, input gate, output gate, and forget gate.
- The LSTM model utilizes its forget gates to remove bias toward recent events. A given LSTM unit will take in the previous cell state and perform various activations, multiplications, and concatenations. Then, the forget gate calculates how much of the current cell state and its input should be passed along, forgetting the rest.
- Note that while the activation function normally used in LSTM is the hyperbolic tangent function (tanh), the rectified linear unit (ReLU) activation function was adopted for better performance in this model.
- Figure 3 shows an example of an LSTM cell and its various gates and functions.

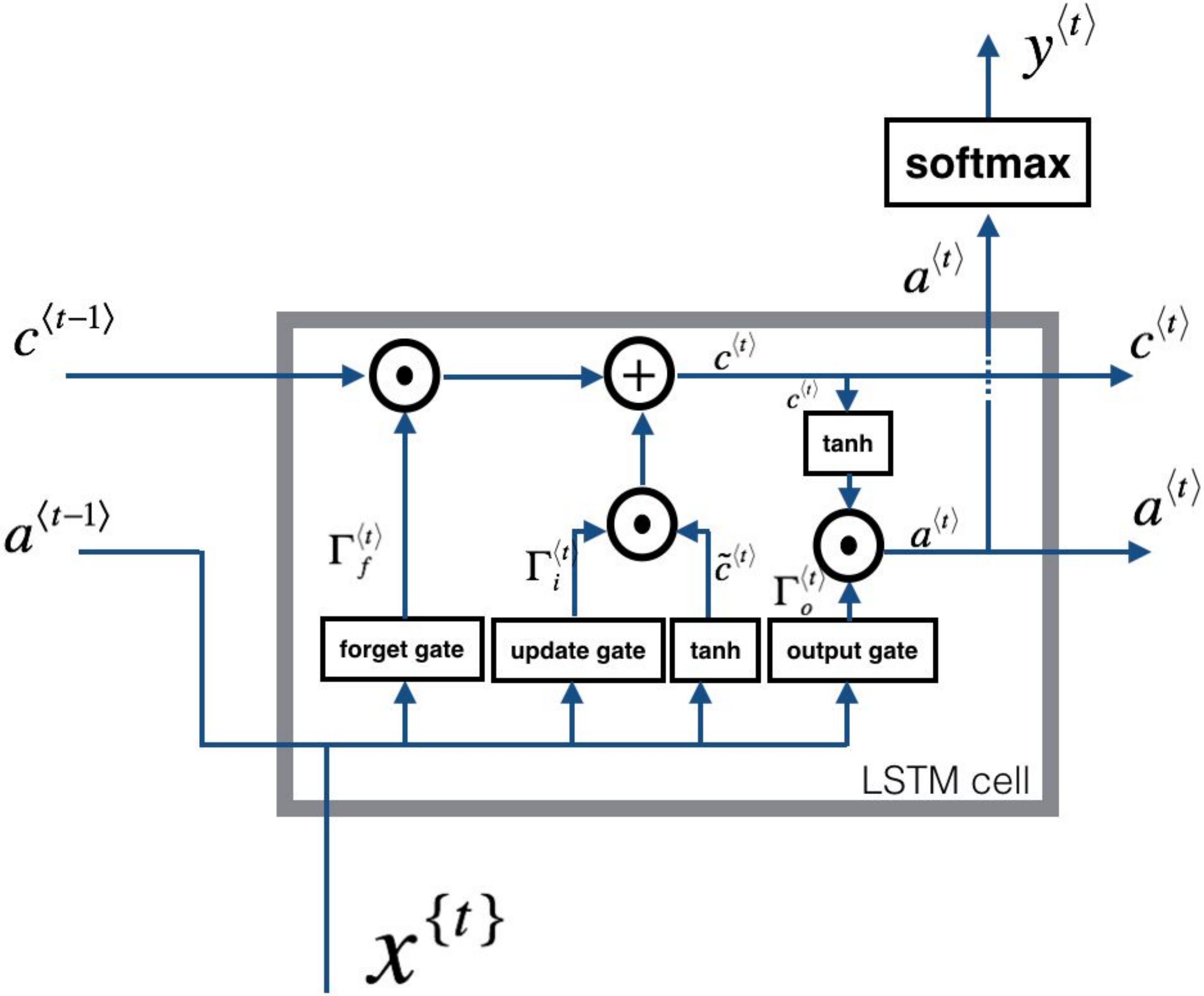

FIGURE 1. Diagram of an LSTM cell.

Prediction Model: GCN-LSTM

- Graph Neural Networks (GNN) are increasing in popularity due to their ability to incorporate graphs and capture the relationships between various factors.
- The Graph Convolutional Network (GCN) can be used to perform machine learning while incorporating network structures such as graphs. As shown in Figure 3, the model consists of two parts: the graph convolution network (GCN) and the LSTM.
- The GCN layer has trainable parameters weight matrix **W** and bias vector **b**, with inputs of node features matrix **F** and the normalized graph adjacency matrix **A'**, as shown in Figure 5.
- The graph adjacency matrix **A'** represents the connections between the nodes. **A'** is normalized so that each node's contribution is proportional to how connected the node is in the graph. An element-wise non-linear function, in this case ReLU, is applied to **A'FW+b**. The output can then be used as input to another neural network layer. (Elinas, 2019)
- The knowledge graph used in this project is displayed in Figure 6, where the edge from node *i* to node *j* exists if the node *i* has a possible impact on node *j*. The adjacency matrix **A'** is a binary matrix showing the existence of edges between nodes.

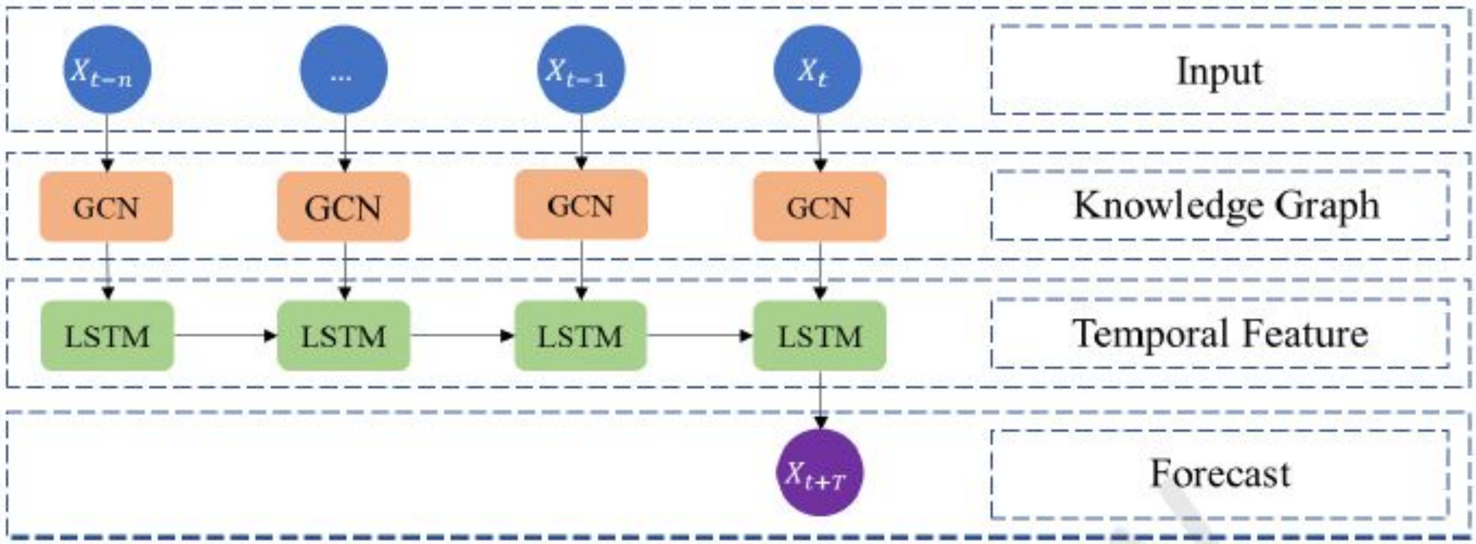

FIGURE 2. GCN-LSTM Model Structure

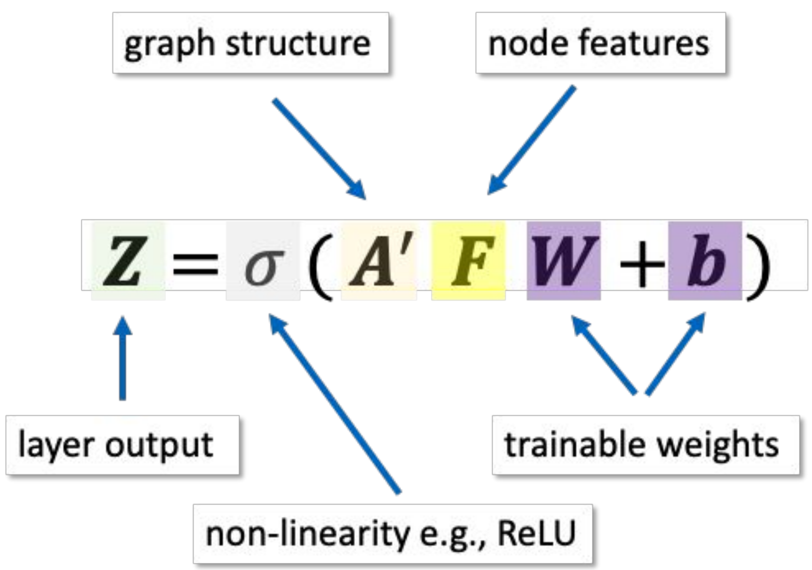

FIGURE 3. GCN Trainable Parameters

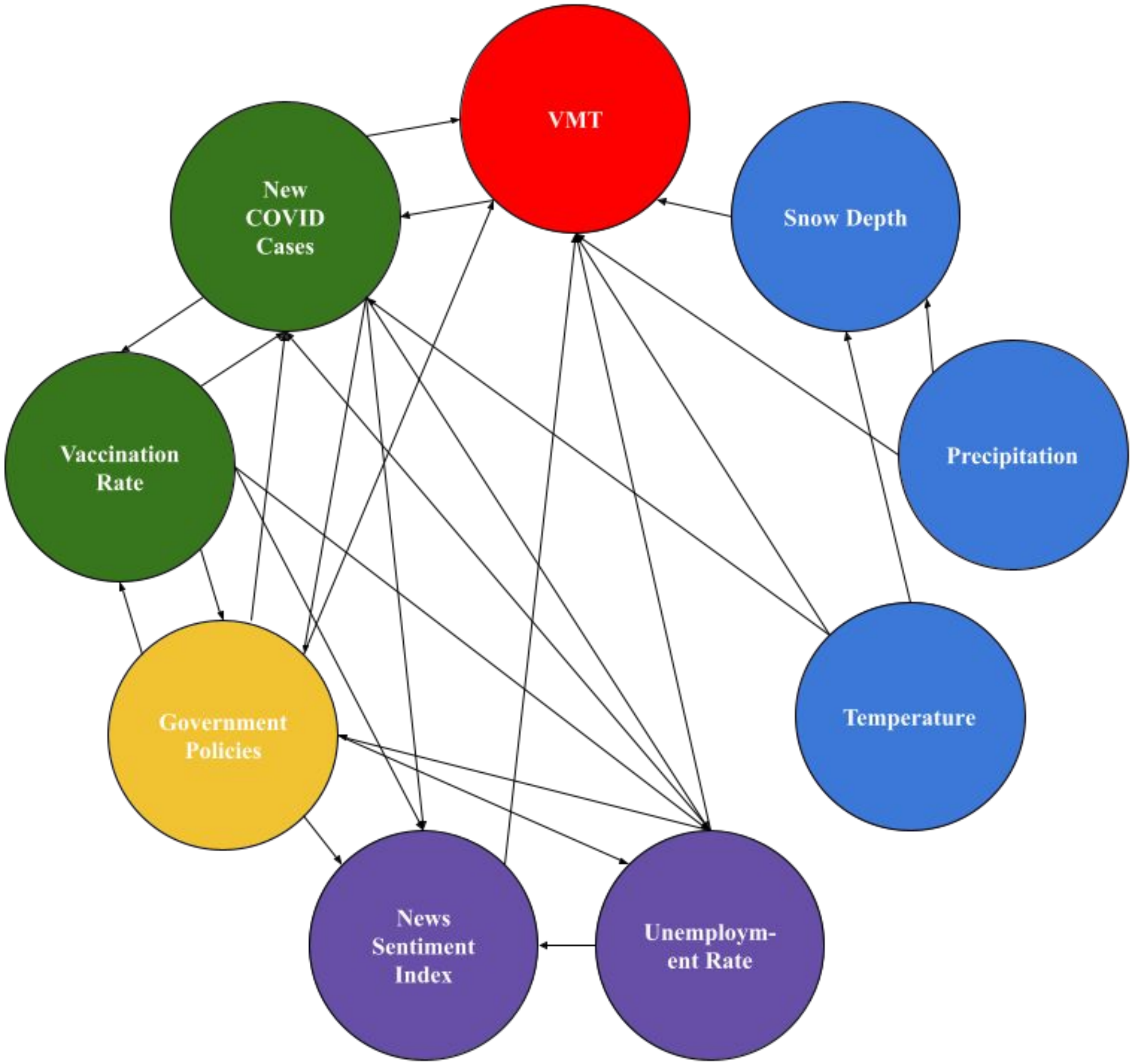

FIGURE 4. GCN Knowledge Graph Depiction

Model Training

- Data since the start of the pandemic were used to develop the model. Eighty eight weeks of data were used for training, and the most recent four weeks of data were used for testing. During the training, an early stopping technique was employed to prevent overfitting. Both models used only data from the past week to make predictions one week in the future.
- The loss function used in training was the mean squared error between the predicted factors and the observed ones. The Adam optimizer was selected to minimize the loss. A persistence model was selected as the benchmark. It calculates the future value of a time series under the assumption that nothing changes between the current time and the forecast time. It should be noted that while both models are capable of predicting all input factors, only VMT was used to evaluate model performance.
- Two evaluation metrics were employed, which are root mean squared error (RMSE) and mean absolute percentage error (MAPE), and are calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(VMT_t - \widehat{VMT_t}\right)^2}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{VMT_t - \widehat{VMT_t}}{VMT_t}\right|$$

Model Architectures

- The Python programming language was used to create the models, with the use of the Tensorflow, Keras, and StellarGraph machine learning libraries. Table 1 shows the hyperparameters used in the GCN-LSTM and LSTM.

TABLE 1. Model Configuration

Model		**LSTM**	**GCN-LSTM**
Layer Configuration	GCN	N/A	16
	LSTM	40	60
Learning Rate		0.01	0.001
Batch Size		4	5
Optimizer		Adam	Adam
Look Back		One Week	One Week

Data Analysis

- We can see that at the start of the pandemic in early March 2020, there was a sharp drop in VMT. Since about June 2020, VMT has gradually increased. There was another drop that coincided with the highest COVID case count around November and December 2020, although that may be influenced by other factors as well, such as winter weather. More recently, we can see a gradual decreasing trend in VMT as COVID cases rise and Omicron spreads. Note that data from December 2021 and January 2022 were not used in the training of the model.

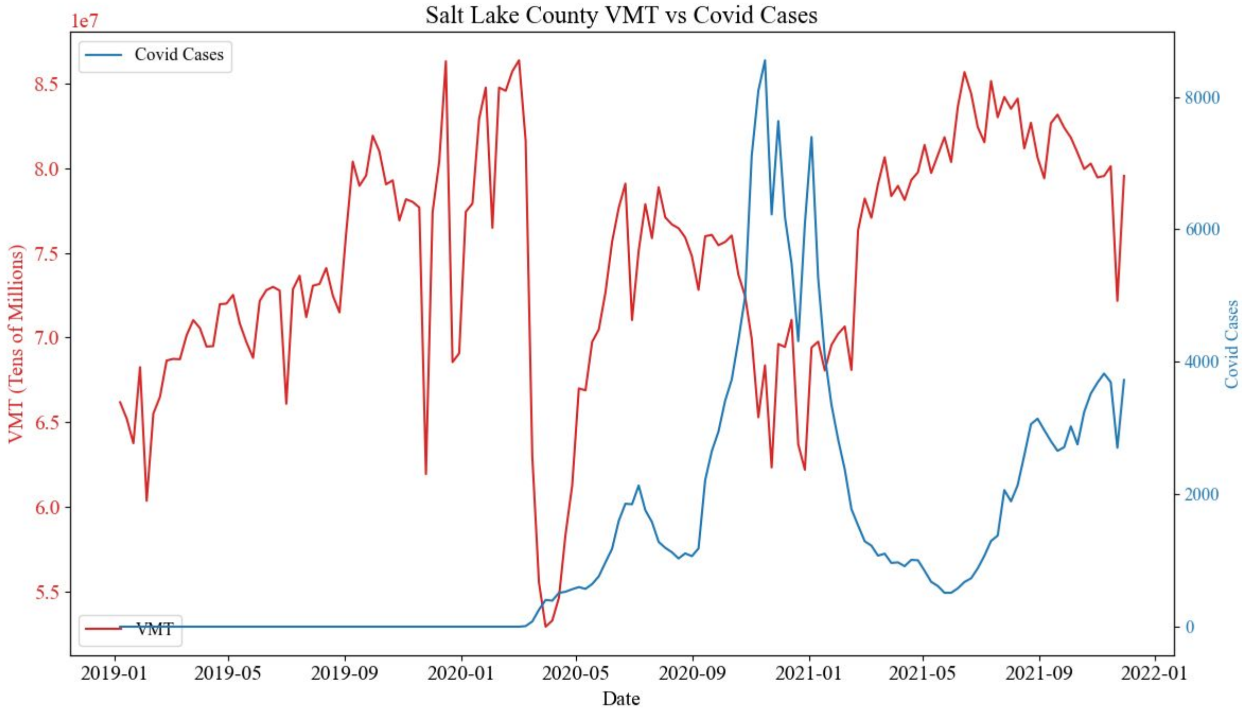

FIGURE 6. VMT Versus Number of New COVID Cases

Results

- Table 2 shows the performance of the LSTM and GCN-LSTM models. Both are able to capture the general increasing trend of VMT after the initial drop at the start of the pandemic. However, the GCN-LSTM has smaller prediction errors compared to the LSTM, outperforming it by 6.30% in terms of MAPE and 2.33% in terms of RMSE.
- Both the LSTM and GCN-LSTM significantly outperform the persistence model benchmark by 86.23% and 87.10% in terms of MAPE, respectively. As shown in Figure 7, there is a sudden drop of VMT from the second week to the third week in the testing data. Both the GCN-LSTM and LSTM had poor capability in predicting this drop.

TABLE 2. Model Performance

Model	**Persistence**		**LSTM**		**GCN-LSTM**	
County	RMSE (10^6 Miles)	MAPE	RMSE (10^6 Miles)	MAPE	RMSE (10^6 Miles)	MAPE
Salt Lake	5.4317	22.98%	3.3292	3.16%	3.2516	2.96%
Performance Improvement	**Persistence / LSTM**		**Persistence / GCN-LSTM**		**LSTM / GCN-LSTM**	
	RMSE (10^6 Miles)	MAPE	RMSE (10^6 Miles)	MAPE	RMSE (10^6 Miles)	MAPE
	-38.71%	-86.23%	-40.14%	-87.10%	-2.33%	-6.30%

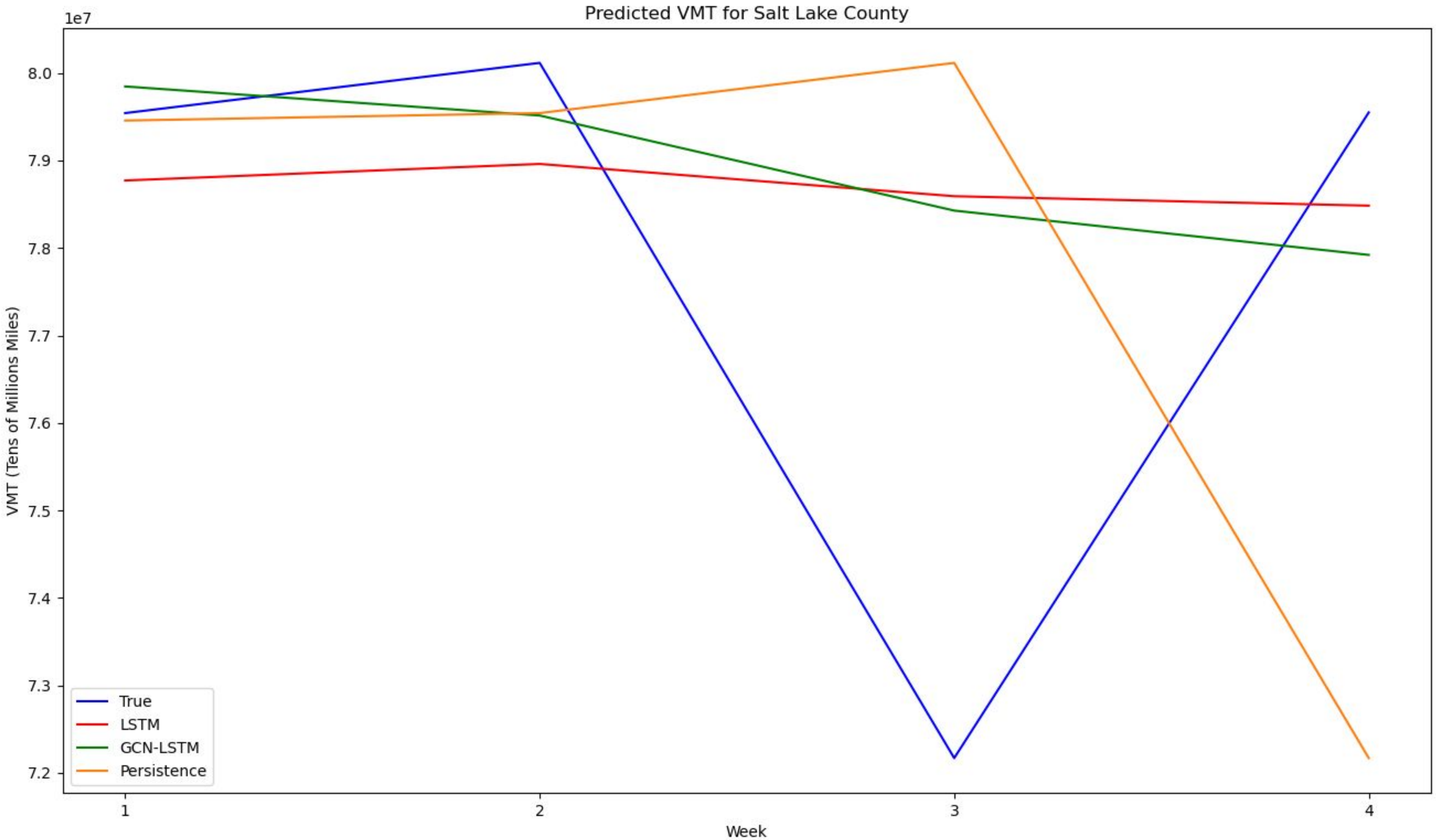

FIGURE 7. Prediction Results of the Models

Conclusion

- Both the GCN-LSTM and LSTM perform significantly better than the persistence model. Persistence models fail to accurately forecast future VMT due to the rapid developments of the pandemic, as well as the variation of weekly VMT in general. For example, the VMT of Salt Lake County dropped by 9.9% (7.9 million miles) from the second to third week of the test dataset. While this occurred near the Thanksgiving holiday, which may have influenced VMT, it suggests that it will be difficult to make long-term traffic forecasts during the pandemic.
- The GCN-LSTM performs the best compared to the LSTM and persistence models. This supports the idea that incorporating human knowledge improves the model's prediction ability.
- There is a sudden drop of VMT from the second to third week, then a sharp rise from the third to fourth week. Both the GCN-LSTM and LSTM were unable to accurately predict the sharp changes. This may be a result of uncaptured randomness, but it is a potentially problematic issue that should be investigated further.
- Because both the GCN-LSTM and LSTM had optimal performance by looking back only one week, a simpler model, such as an AutoRegressive Integrated Moving Average (ARIMA) or a plain Recurrent Neural Network (RNN) might produce comparable or even better results.

Next Steps

- With certain data only available up to November 2021, this project was not able to take into account the recent developments of the pandemic, including the emergence of the Omicron variant, recent school closings, and the availability of a booster shot. Further research could utilize more recent data to improve the models.
- Holiday dates were not considered in these models. However, the incorporation of holiday dates may increase the accuracy of the models, since VMT can be heavily influenced by holiday travel. Vehicle class data, such as Truck VMT, was also not utilized in this project. Future work could investigate the trends in these data, which might relate to an increase in online shopping during the pandemic.
- While the models in this project were fine-tuned manually, better performing models may exist. An optimization algorithm, such as Bayesian Optimization, could be used to search for potentially better hyperparameters that may improve the models. Other models, such as AutoRegressive Integrated Moving Average (ARIMA), plain Recurrent Neural Networks (RNNs), or variations of Graph Neural Networks (GNN), might also be considered to predict traffic demand.
- A similar model could be developed for other forms of transportation, such as air travel, which may help airline businesses predict and prepare for changes in demand. By extending this project to other geographic areas, such as Utah County, the entire State of Utah, and the rest of the United States, as well as specific freeway sections, the prediction model could be used by departments of transportation to prepare for changes in traffic patterns in the near future.

References

- Kim, Karl. Impacts of COVID-19 on transportation: Summary and synthesis of interdisciplinary research, Transportation Research Interdisciplinary Perspectives, Volume 9, 2021, 100305, ISSN 2590-1982, https://doi.org/10.1016/j.trip.2021.100305.
- Glaeser, E. L., C. Gorback, and S. J. Redding. "How Much Does COVID-19 Increase with Mobility? Evidence from New York and Four Other U.S. Cities." 2020. https://doi.org/10.3386/W27519.
- Hochreiter, S., J. Schmidhuber, L. M. Gambardella, and J. Schmidhuber. "Long Short-Term Memory." Neural Computation, Vol. 9, No. 8, 1997, pp. 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
- Olah, Christopher. “Understanding LSTM Networks.” *Understanding LSTM Networks -- Colah's Blog*, 27 Aug. 2015, https://colah.github.io/posts/2015-08-Understanding-LSTMs/. Accessed Jan. 24, 2022.
- Omerbsezer. “Omerbsezer/BASICLSTM: The Aim of This Implementation Is to Help to Learn Structure of Basic LSTM (LSTM Cell Forward, LSTM Cell Backward, Etc..).” *GitHub*, https://github.com/omerbsezer/BasicLSTM.
- Elinas, Pantelis. “Knowing Your Neighbours: Machine Learning on Graphs.” *Medium*, Stellargraph, 1 Nov. 2021, https://medium.com/stellargraph/knowing-your-neighbours-machine-learning-on-graphs-9b7c3d0d5896. Accessed Jan. 24, 2022.
- CSIRO’s Data61. Stellargraph machine learning library. https://github.com/stellargraph/stellargraph, 2018.
- Chollet, Francois et al. Keras, 2015. https://github.com/keras-team/keras
- PeMS @ UDOT. https://udot.iteris-pems.com/. Accessed Jan. 24, 2022.
- “Case Counts: Coronavirus.” *Coronavirus.utah.gov*, https://coronavirus.utah.gov/case-counts/. Accessed Jan. 24, 2022.
- Seasonally Adjusted Unemployment Rates by County. https://jobs.utah.gov/wi/update/une/. Accessed Jan. 24, 2022.
- Federal Reserve Bank of San Francisco. “Daily News Sentiment Index.” *Federal Reserve Bank of San Francisco*, Federal Reserve Bank of San Francisco. https://www.frbsf.org/economic-research/indicators-data/daily-news-sentiment-index/. Accessed Jan. 24, 2022.
- Find a Station | Data Tools | Climate Data Online (CDO) | National Climatic Data Center (NCDC). https://www.ncdc.noaa.gov/cdo-web/datatools/findstation. Accessed Jan. 24, 2022.
- Special Orders | Coronavirus. https://coronavirus.utah.gov/special-orders/. Accessed Jan. 24, 2022.