Traffic Volume Prediction based on Memory Based RNN

Lokesh Chandra Das, U00740183

COMP 8118-M50

Introduction

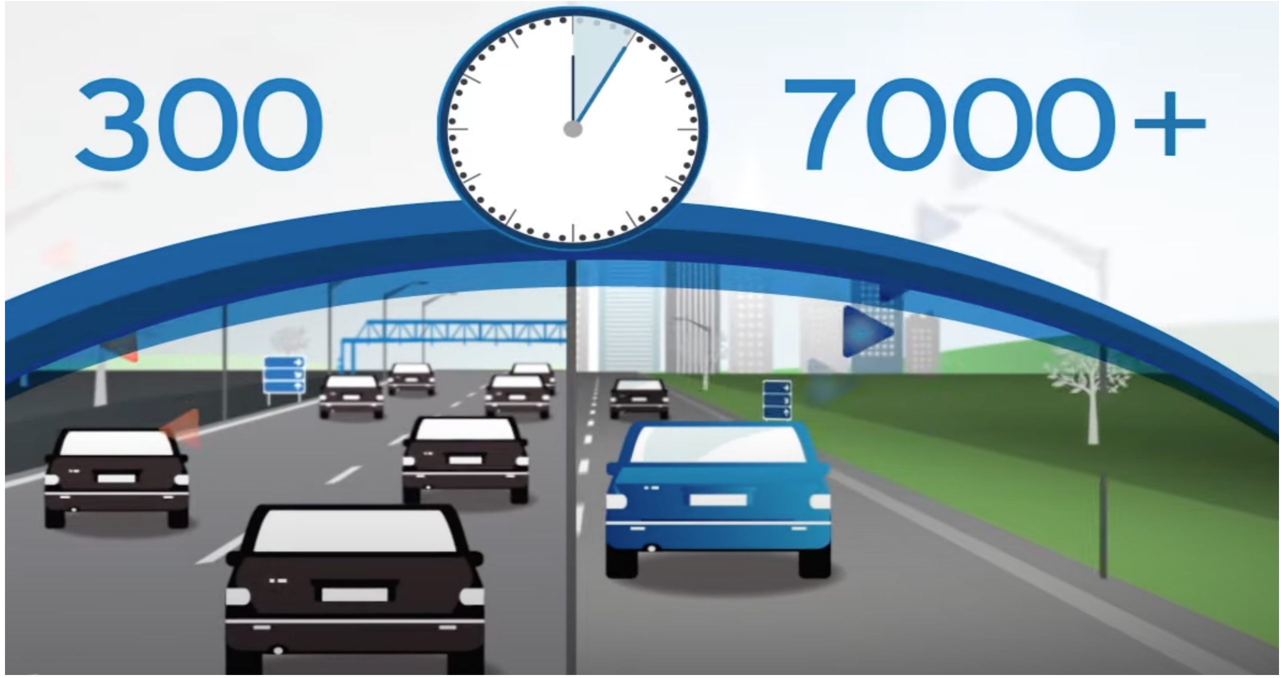
- Traffic forecasting aims to anticipate future traffic conditions (e.g., traffic volume, vehicle speed) based on past and current traffic situations.
- Accurate traffic flow predictions are critically important
  - Mitigate traffic congestions, improve traffic efficiency, and ensure better traffic safety
- Predicting traffic flow is a challenging task due to the spatial-temporal dependency and pattern
  - Spatially: Dynamic traffic conditions of nearby roads
  - Temporally: Weather conditions, rush-hours, weekends and holidays
- We propose a **memory-based Recurrent Neural Network approach** to estimate the future traffic volume given the past observations.

Related Works

- Parametric Methods:
  - *Chen et al*. used ARIMA for traffic flow predictions.[1]
  - *Kumar et al.* designed a Kalman Filter to forecast the traffic flow.[2]
  - *Dong et al.* used Gradient Boosting based Decision Tree method to perform short-time traffic flow prediction.[3]
- Non-Parametric Approaches:
  - *Zhen et al.* proposed a Graph Multi-Attention network for traffic prediction.[4]
  - *Zhao et al*. apply LSTM to forecast short-term traffic.[5]
  - *Fu et al.* used LSTM and GRU for traffic flow prediction.[6]

[1] Chen, C., Hu, J., Meng, Q., & Zhang, Y. (2011, June). Short-time traffic flow prediction with ARIMA-GARCH model. In *2011 IEEE Intelligent Vehicles Symposium (IV)* (pp. 607-612). IEEE.
[2] Kumar, S. V. (2017). Traffic flow prediction using Kalman filtering technique. *Procedia Engineering*, *187*, 582-587.
[3] Dong, X., Lei, T., Jin, S., & Hou, Z. (2018, May). Short-term traffic flow prediction based on XGBoost. In *2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS)* (pp. 854-859). IEEE.
[4] Zheng, C., Fan, X., Wang, C., & Qi, J. (2020, April). Gman: A graph multi-attention network for traffic prediction. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 01, pp. 1234-1241).
[5] Zhao, Z., Chen, W., Wu, X., Chen, P. C., & Liu, J. (2017). LSTM network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, *11*(2), 68-75.
[6] Fu, R., Zhang, Z., & Li, L. (2016, November). Using LSTM and GRU neural network methods for traffic flow prediction. In *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)* (pp. 324-328). IEEE.

Proposed Method



- Traffic volume prediction is a time-series problem.
  - Predict future traffic volume based on empirical data
- LSTM and GRU based deep neural network

Dataset(1/2)

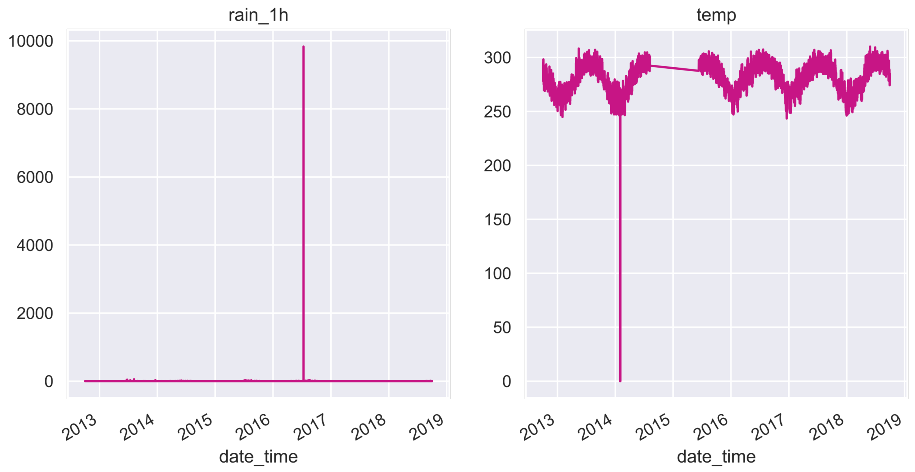
date_time	holiday	temp	rain_1h	snow_1h	clouds_all	weather_main	weather_description	traffic_volume
2012-10-02 09:00:00	None	288.28	0.0	0.0	40	Clouds	scattered clouds	5545
2012-10-02 10:00:00	None	289.36	0.0	0.0	75	Clouds	broken clouds	4516
2012-10-02 11:00:00	None	289.58	0.0	0.0	90	Clouds	overcast clouds	4767
2012-10-02 12:00:00	None	290.13	0.0	0.0	90	Clouds	overcast clouds	5026
2012-10-02 13:00:00	None	291.14	0.0	0.0	75	Clouds	broken clouds	4918

- Hourly Interstate 94 Westbound traffic volume for MN DoT ATR station 301.
  - **Metro-Interstate-Traffic-Volume.**[*]
- Hourly weather features and holidays included for impacts on traffic volume
- Multivariate, Sequential, Time-Series
  - No. of features: 9
  - Total Records: 48204

[*] https://archive.ics.uci.edu/ml/datasets/Metro+Interstate+Traffic+Volume

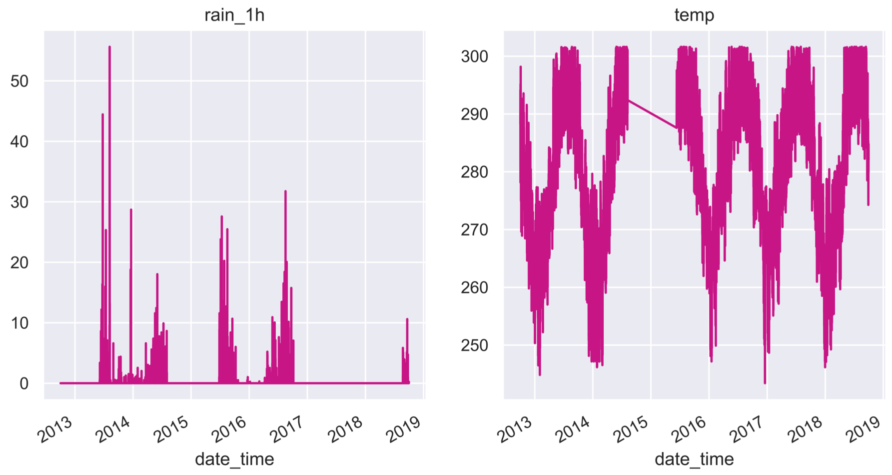
Dataset(2/2)

- Outlier exists in the dataset



Feature *rain_1h* and *temp* with outlier

- Interquartile range removes the outlier



Feature *rain_1h* and *temp* without outlier

Data Preprocessing

- Data are in different scales
  - Max value is 7280 and min is 0.00227
- Features are scaled in [0,1]
- Used past t-hours of data to predict the next $'n'$ hours traffic volume
- RNN Input:

# of records

# of features

Feature, X: (N, L, n)

past observation

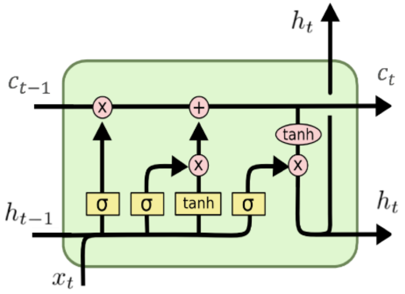
Target, Y: (N, 1)

	holiday	temp	rain_1h	snow_1h	clouds_all	weather_main	weather_description	traffic_volume
count	47284.0	47284.000000	47284.000000	47284.000000	47284.000000	47284.000000	47284.000000	47284.000000
mean	1.0	280.839466	0.132735	0.000227	49.790986	2.605173	16.564631	3232.676233
std	0.0	12.451778	1.012858	0.008247	39.064691	2.787757	8.916584	1989.588639
min	1.0	243.390000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	1.0	272.020000	0.000000	0.000000	1.000000	0.000000	10.000000	1151.000000
50%	1.0	281.930000	0.000000	0.000000	64.000000	1.000000	17.000000	3319.000000
75%	1.0	291.450000	0.000000	0.000000	90.000000	5.000000	27.000000	4919.250000
max	1.0	301.620000	55.630000	0.510000	100.000000	10.000000	37.000000	7280.000000

- Categorical attributes are converted to numerical values.
- Data Split:
  - Training and Validation sets: 2012-10-20 to 2017-12-31
  - Testing set: 2018-01-01 to upwards

Memory Based RNN

LSTM*

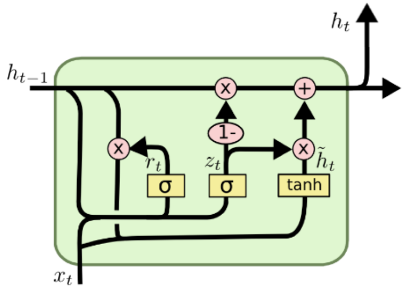


LSTM
(Long-Short Term Memory)

$$f_t = \sigma_g\,(W_f \times x_t + U_f \times h_{t-1} + b_f)$$
$$i_t = \sigma_g\,(W_i \times x_t + U_i \times h_{t-1} + b_i)$$
$$o_t = \sigma_g\,(W_o \times x_t + U_o \times h_{t-1} + b_o)$$
$$c'_t = \sigma_c\,(W_c \times x_t + U_c \times h_{t-1} + b_c)$$
$$c_t = f_t \cdot c_{t-1} + i_t \cdot c'_t$$
$$h_t = o_t \cdot \sigma_c(c_t)$$

- 3-Gates: Let the information through
  - Forget Gate: Decide what to forget
  - Input Gate: Decide what value to update
  - Output Gate: decide what to be output

GRU*



GRU
(Gated Recurrent Unit)

$$r_t = \sigma(W_r.x_t + W_r.h_{t-1} + b_r)$$
$$z_t = \sigma(W_z.x_t + W_z.h_{t-1} + b_z)$$
$$\tilde{h}_t = tanh(W_h.x_t + r_t * W_h.h_{t-1} + b_z)$$
$$h = z_t * h_{t-1} + (1 - z) * \tilde{h}_t$$

- 2-Gates:
  - Reset Gate: Determines how to fuse new input with the previous memory
  - Update Gate: How much of the previous memory remains

❖ GRU is memory efficient and faster compared to LSTM. However, LSTM is more accurate when dataset is complex and has longer sequences.

* Image: https://github.com/roomylee/rnn-text-classification-tf

Training Details

Evaluation Metrics

- MSE

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

- MAE

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

- MAPE

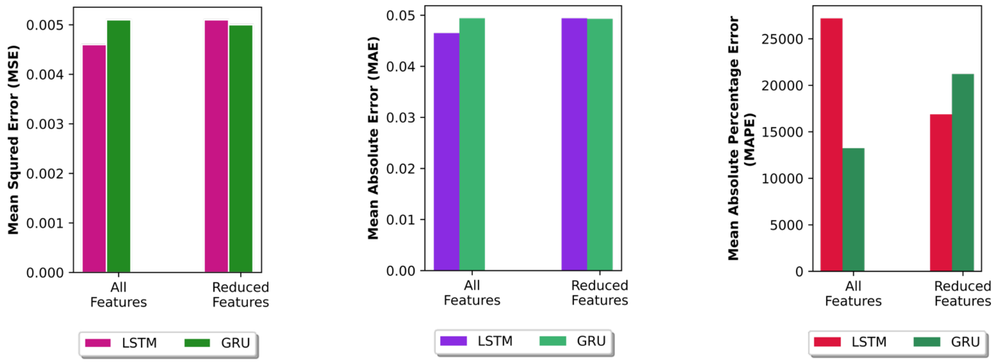
$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\%$$

Experimental Setting

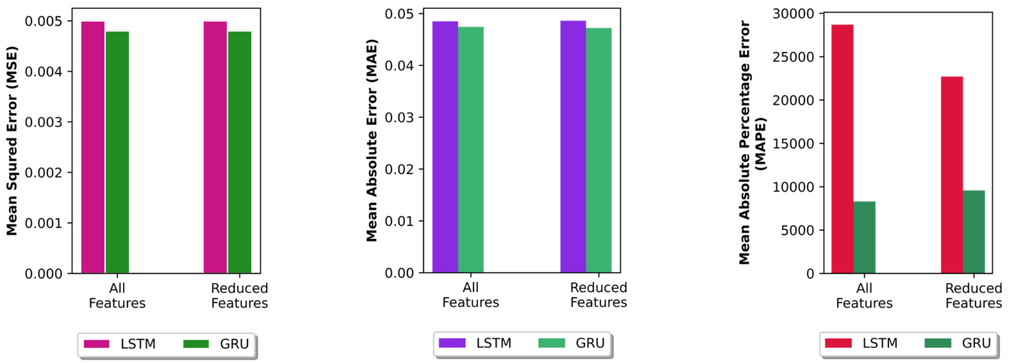
Neural Network Settings	
Input Shape	(# of Records, Time Step, # of Features)
Hidden Layers	4 Hidden Layers with 128, 64, 32, 16 units of neurons in each layer respectively
Activate Function	tanh
Batch Size	64
Learning Rate	0.0001 with decay rate 1e-5
Optimizer	Adam
Epochs	300

- Total 4 Experiments:
  - 6 past observations
  - 24 past observations
- Run with all features vs reduced features.
  - Reduced Features:
    - *Temp, rain_1h, clouds_all, Traffic volume*

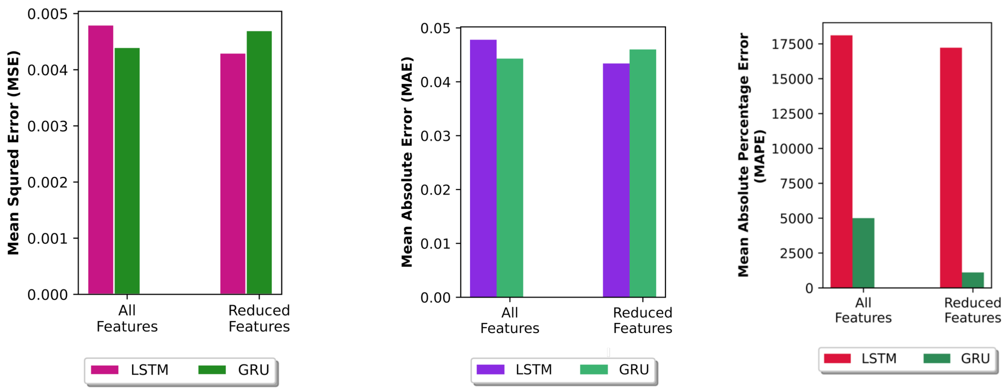
Experimental Results(1/4)



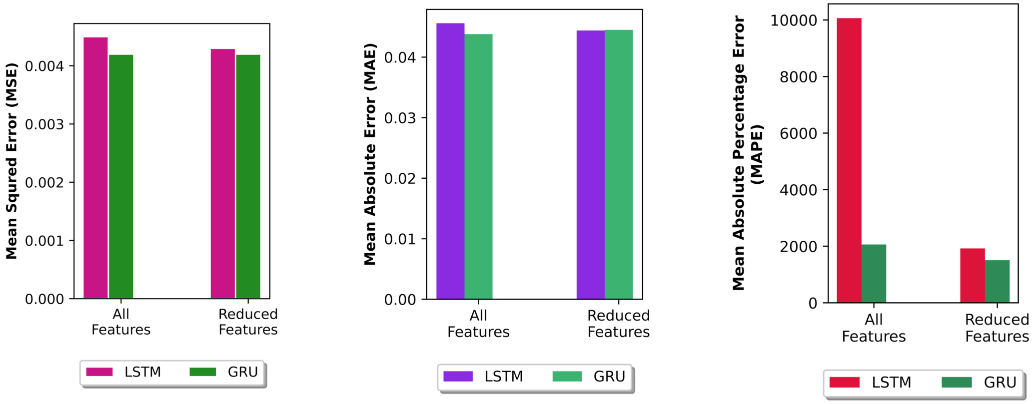
NN settings: [128, 64,32,16], epochs: 300 (24timestep)



NN settings: [128, 64,32,16], epochs: 300 (6 timestep)



NN settings: [256, 128,64,32], epochs: 500 (24timestep)



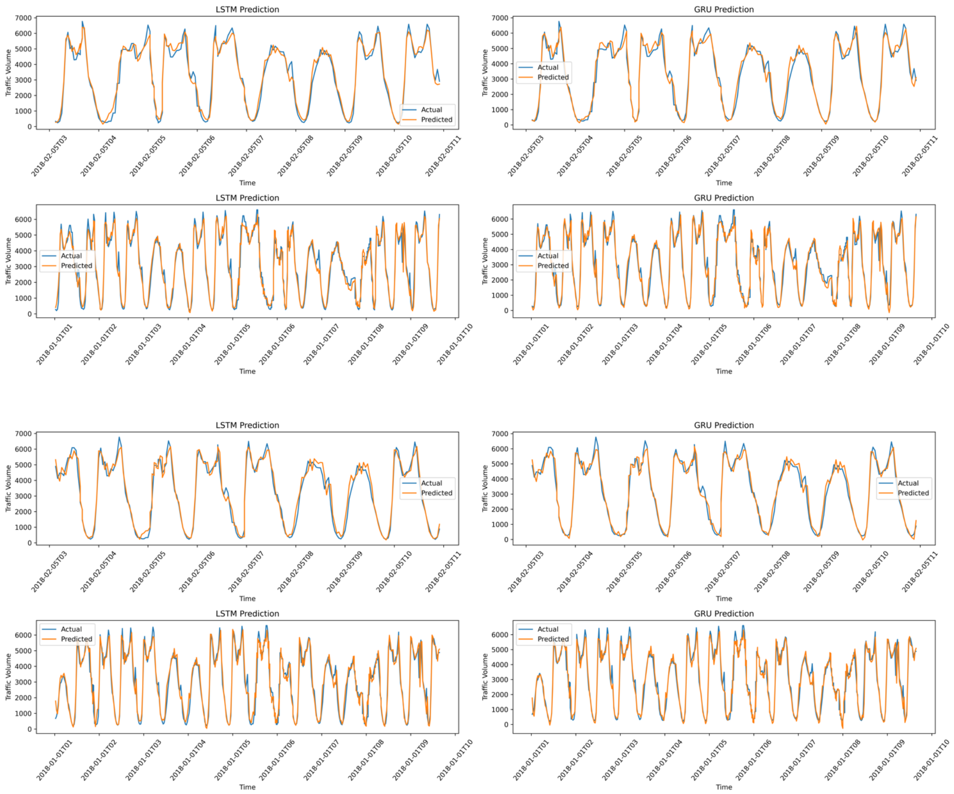
NN settings: [256, 128,64,32], epochs: 500 (6 timestep)

- Empirically, LSTM and GRU outperform ARIMA model.[*]

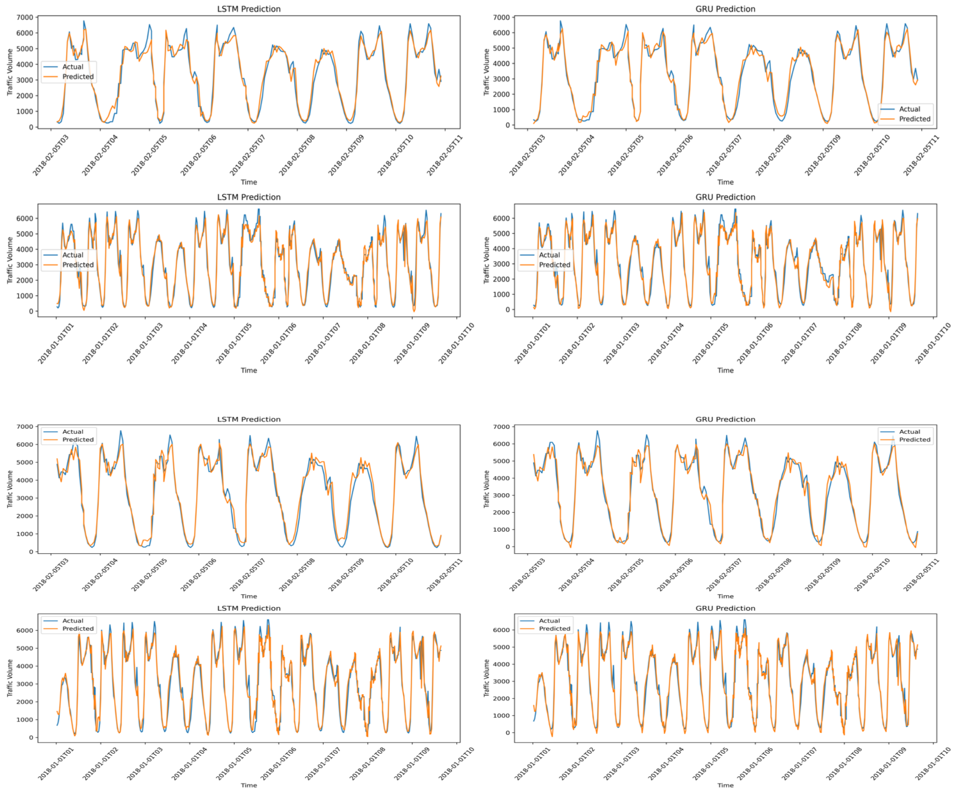
[*] Fu, R., Zhang, Z., & Li, L. (2016, November). Using LSTM and GRU neural network methods for traffic flow prediction. In *2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)* (pp. 324-328). IEEE.

Experimental Results(2/4)

All features (timestep 24, 6)



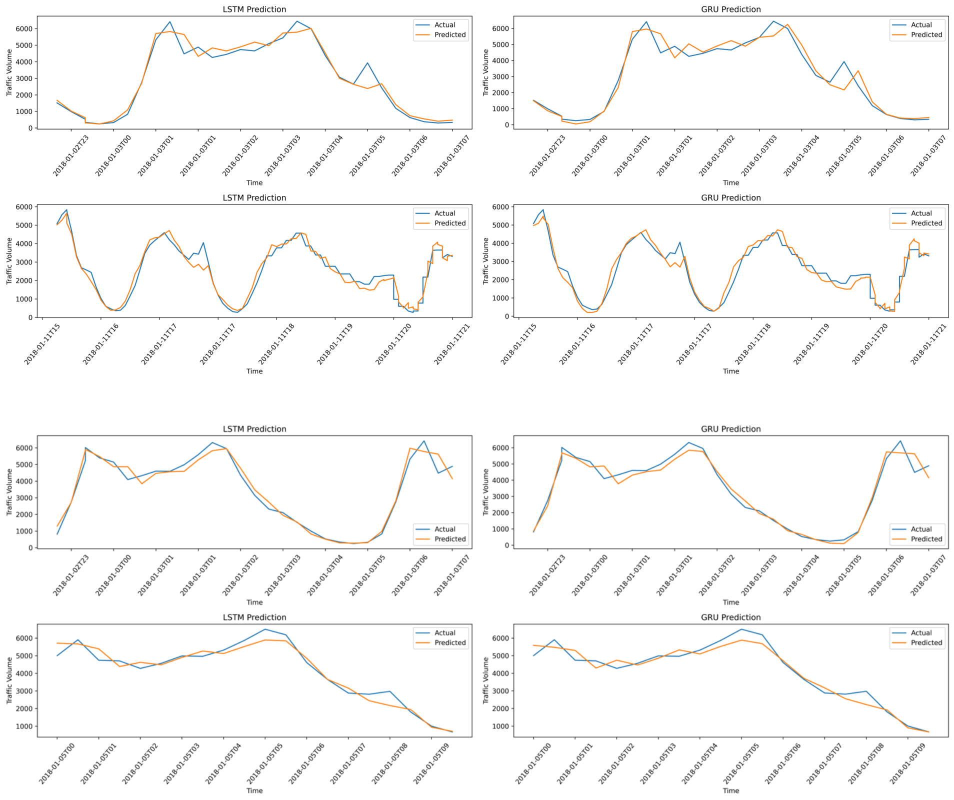
Reduced Features (timestep 24, 6)



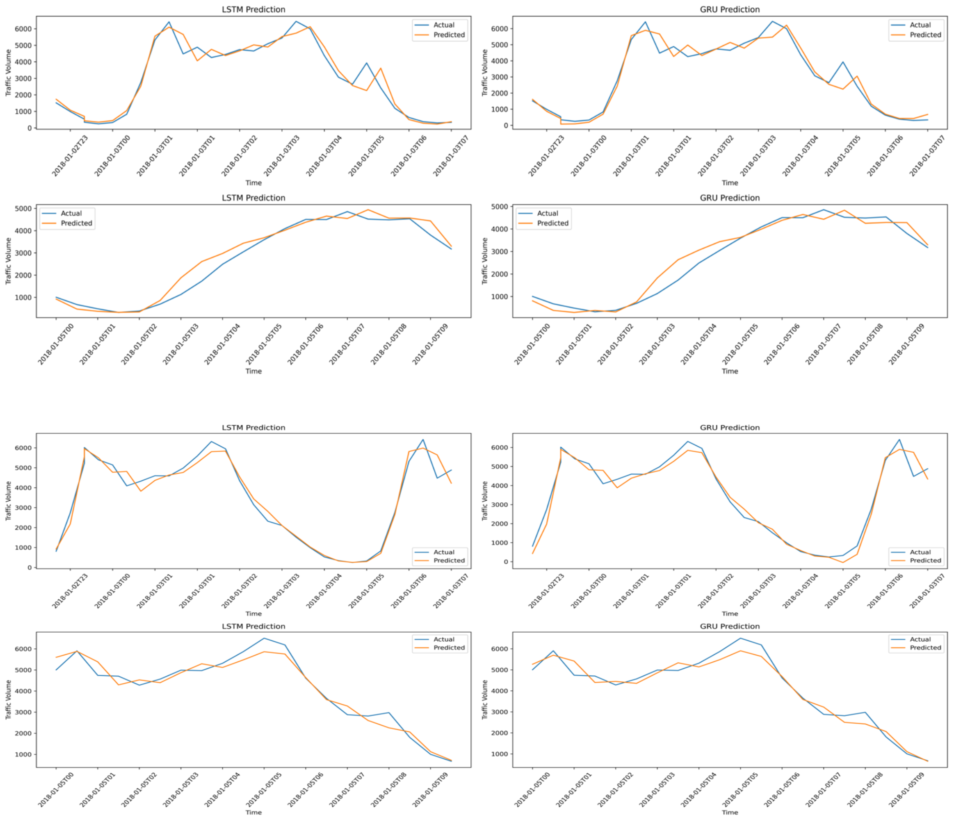
NN settings: [128, 64,32,16], epochs: 300

Experimental Results(3/4)

All features (timestep 24, 6)



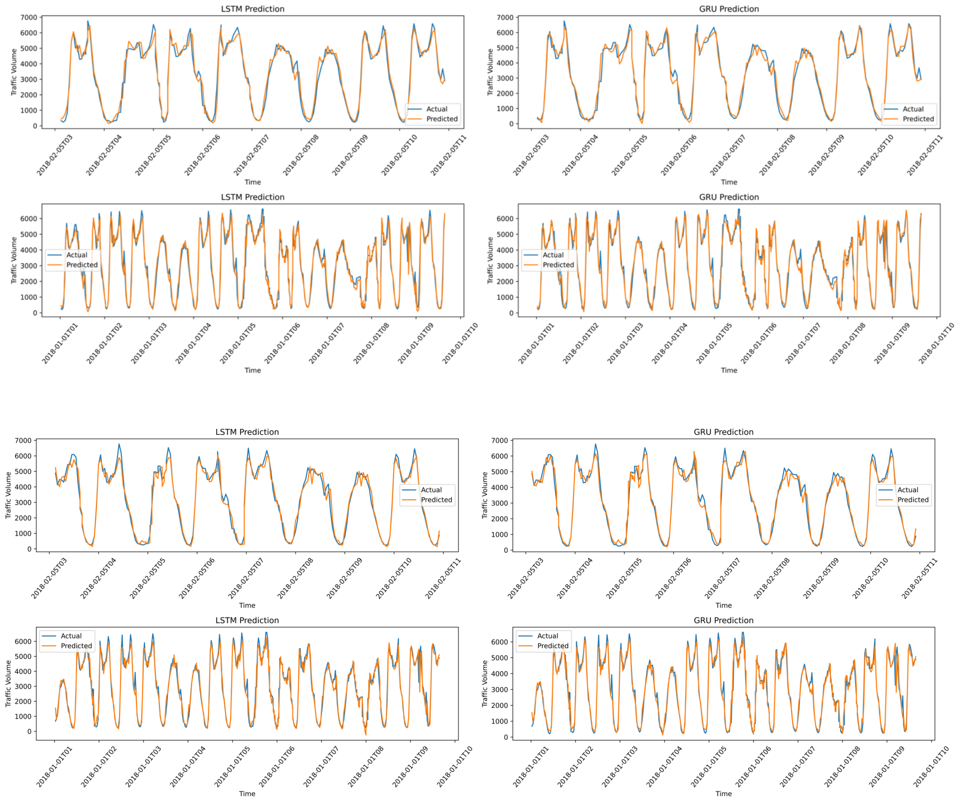
Reduced Features (timestep 24, 6)



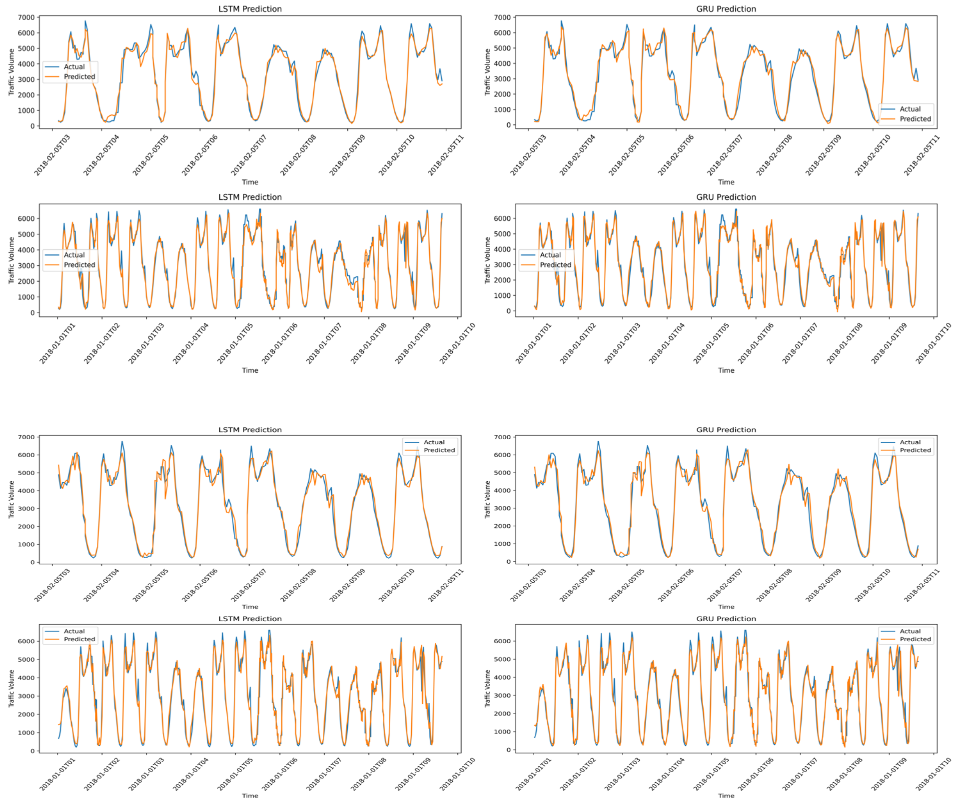
NN settings: [128, 64,32,16], epochs: 300

Experimental Results(4/4)

All features (timestep 24, 6)



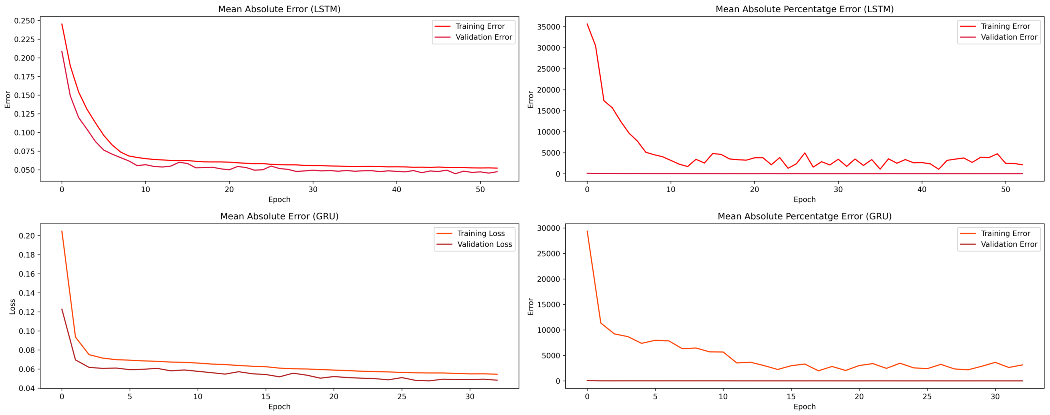
Reduced Features (timestep 24, 6)



NN settings: [256, 128, 64,32], epochs: 500

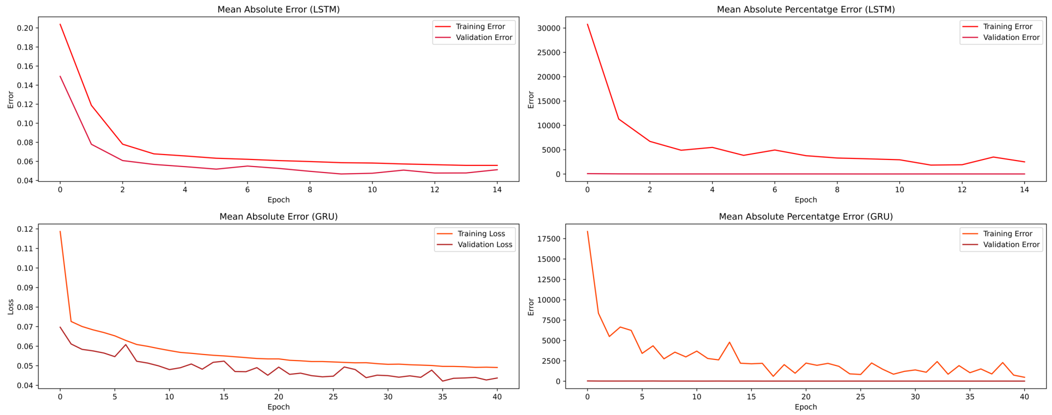
Experimental Results (MAE/MAPE)

All features used



NN settings: [128,64,32,16]: epoch 300 (24 timestep)

- **4 Features**
  - **Temp, rain_1h, clouds_all, Traffic volume**



NN settings: [256, 128,64,32]: epoch 500 (24 timestep)

Conclusion & Future Works

- Traffic Volume Prediction problem is formulated using Long-short term memory (LSTM) and Gated Recurrent Unit (GRU)
- GRU and LSTM outperforms the parametric models (ARIMA) and vanilla RNN.
- GRU perform better than LSTM if the model does not need to keep very long history of observations.