Multivariate time series forecasting - Predicting traffic

Lam Ngo

April 24, 2022

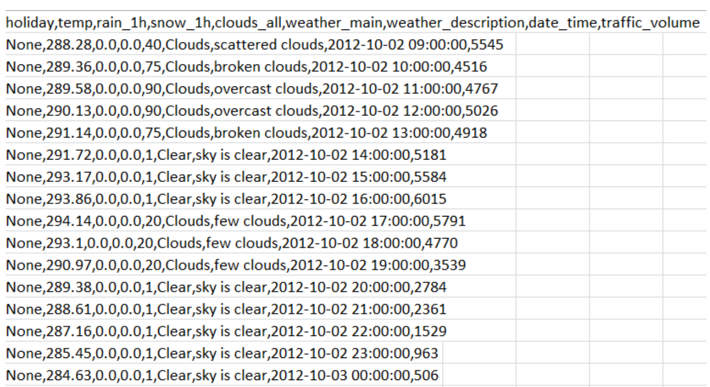
1 Task/Problem

Different algorithms will be used to forecast traffic volumes on a interway. This is a common task in time series forecasting. It has wide-ranging applications. Forecast can be fed to Intelligent Transportation Systems and Smart navigation systems. In Intelligent Transportation Systems it can e.g. used for congestion pricing. Roads that are forecasted to have higher traffic can increase tolls to incentivise drivers to take a different road. In navigation systems, forecasts can be used to optimize the route that is recommended to the user.

I will use state-of-the-art algorithms and compare their performance with baselines. These state-of-the-art algorithms were not yet tested on the data set I used.

2 Data

I obtained traffic data from [4]. There are about 48000 observations and 9 features.



```
holiday,temp,rain_1h,snow_1h,clouds_all,weather_main,weather_description,date_time,traffic_volume
None,288.28,0.0,0.0,40,Clouds,scattered clouds,2012-10-02 09:00:00,5545
None,289.36,0.0,0.0,75,Clouds,broken clouds,2012-10-02 10:00:00,4516
None,289.58,0.0,0.0,90,Clouds,overcast clouds,2012-10-02 11:00:00,4767
None,290.13,0.0,0.0,90,Clouds,overcast clouds,2012-10-02 12:00:00,5026
None,291.14,0.0,0.0,75,Clouds,broken clouds,2012-10-02 13:00:00,4918
None,291.72,0.0,0.0,1,Clear,sky is clear,2012-10-02 14:00:00,5181
None,293.17,0.0,0.0,1,Clear,sky is clear,2012-10-02 15:00:00,5584
None,293.86,0.0,0.0,1,Clear,sky is clear,2012-10-02 16:00:00,6015
None,294.14,0.0,0.0,20,Clouds,few clouds,2012-10-02 17:00:00,5791
None,293.1,0.0,0.0,20,Clouds,few clouds,2012-10-02 18:00:00,4770
None,290.97,0.0,0.0,20,Clouds,few clouds,2012-10-02 19:00:00,3539
None,289.38,0.0,0.0,1,Clear,sky is clear,2012-10-02 20:00:00,2784
None,288.61,0.0,0.0,1,Clear,sky is clear,2012-10-02 21:00:00,2361
None,287.16,0.0,0.0,1,Clear,sky is clear,2012-10-02 22:00:00,1529
None,285.45,0.0,0.0,1,Clear,sky is clear,2012-10-02 23:00:00,963
None,284.63,0.0,0.0,1,Clear,sky is clear,2012-10-03 00:00:00,506
```

Figure 1: Data sample

The target value is the hourly traffic volume (in cars) for MN DoT ATR station 301, roughly midway between Minneapolis and St Paul, MN from 2012 to 2018. The features that can be used to predict the traffic are the following: *holiday* is a categorical variable which gives the name of the holiday at that day, *temp* is the average temp in kelvin, $rain_1h$ is the amount in mm of rain, $snow_1h$ is the amount in mm of snow that occurred in the hour, $clouds_all$ is the percentage of cloud cover, $weather_main$ is a categorical variable which gives a textual description of the weather, $weather_description$ is a categorical variable that gives a more detailed textual description of the weather and $date_time$ is the local time. A summary of the numerical features can be seen in figure 2

	temp	rain_1h	snow_1h	clouds_all	traffic_volume
count	48204.000000	48204.000000	48204.000000	48204.000000	48204.000000
mean	281.205870	0.334264	0.000222	49.362231	3259.818355
std	13.338232	44.789133	0.008168	39.015750	1986.860670
min	0.000000	0.000000	0.000000	0.000000	0.000000
25%	272.160000	0.000000	0.000000	1.000000	1193.000000
50%	282.450000	0.000000	0.000000	64.000000	3380.000000
75%	291.806000	0.000000	0.000000	90.000000	4933.000000
max	310.070000	9831.300000	0.510000	100.000000	7280.000000

Figure 2: Numerical features

I use *sklearn* for label encoding and scaling to zero variance and mean.

3 Related literature

Traffic forecasting is difficult as non-expected events like bad weather can happen. Therefore algorithms that can incorporate these events will be successful [16]. That is why multivariate models are useful for the data set in this project.

I am not the first to do traffic forecasting. Non-neural network attempts include seasonal ARIMA [13], hidden markov model [10] and support vector regression [1]. Several attempts using neural networks have been done such as using LSTM/RNN [2] [17], autoencoder [8] and graph neural networks [18]. I have not found a paper that uses my methods as seen in 4 on the data that I used.

There is a plethora of models for multivariate time series forecasting. A good overview of state-of-the-art neural network models can be found here [12]. Therefore, I selected deep learning models that make sense to me. As seen in Chapter 4, I use DA-RNN [11] and Informer [19]. The reason I use the Informer model is that it is an improvement over the transformer [14] which was covered intensively in our course. DA-RNN is an architecture that I personally found interesting as it is easy to understand and intuitive.

Traffic forecasts are used in real-time intelligent transportation systems environment to increase the operational efficiency and capacity of transportation systems [15].

4 Methods

5 methods are used to predict *traffic volume*. For I I use 2 non deep learning methods *LinearRegression* and *XGBoost*. For deep learning methods, I use a vanilla LSTM and 2 state-of-the-art methods, namely a Dual-Stage Attention-Based Recurrent Neural Network (DA-RNN) [4] and a $Informer$ [19].

4.1 Linear Regression

I used a linear regression of the traffic volume on all features and the last 48 values of traffic volume. This is an easy baseline.

4.2 XGBoost

XGBoost [3] is a tree boosting algorithm. It uses additive regression trees to minimize a regularized loss function. It uses gradient boosting to iteratively add new trees. This algorithm is further optimized for sparse data and is quite efficient and parallelizable. Empirically, XGBoost has performed quite well. This is a harder baseline.

4.3 LSTM

A vanilla LSTM [5] was used.

4.4 Dual-Stage Attention-Based Recurrent Neural Network

The Dual-Stage Attention-Based Recurrent Neural Network (DA-RNN) [11] is a 2 stage model with the usual encoder-decoder architecture. It resembles human perception as humans first select features that seem to be important. And so does DA-RNN. Model like ARIMA [9] do not see any feature series as more important.

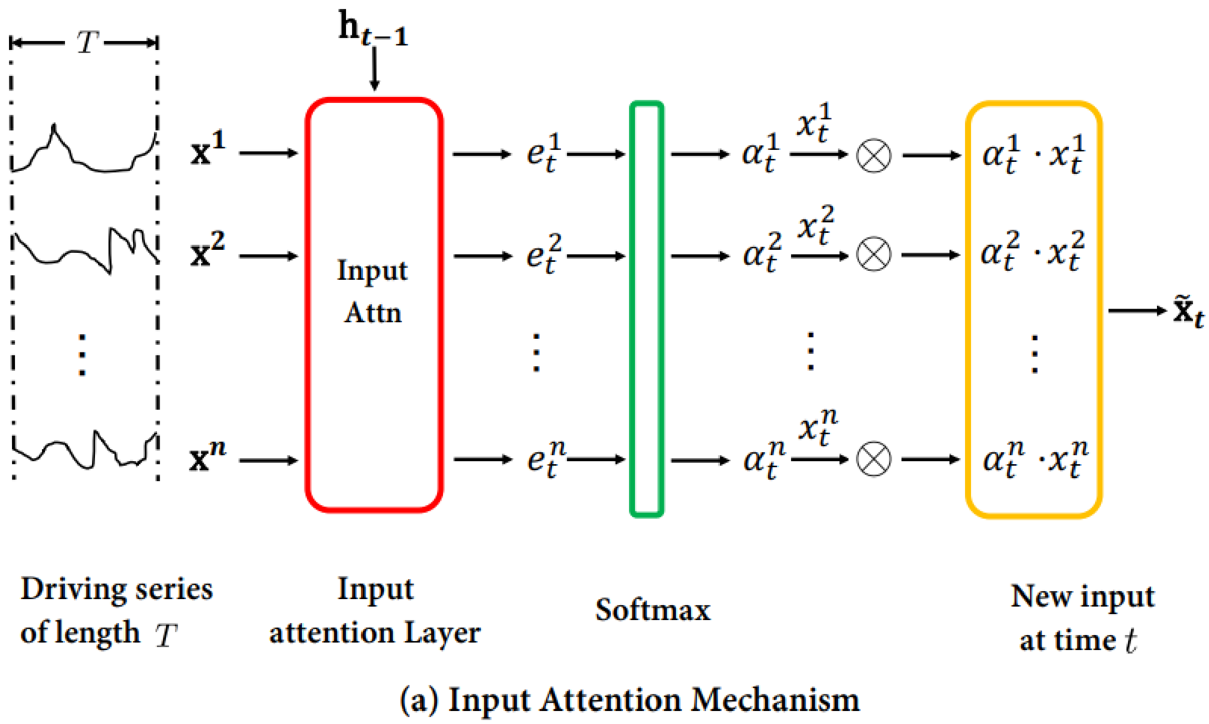


Figure 3: First stage DA-RNN

The first step of the DA-RNN can be seen in figure 3. All the different feature series (”driving series”) are used as input to the attention layer in combination with the previous hidden state. The output of the attention layer is then put through a softmax layer and one gets the attention towards certain feature series. These weighted inputs are then used for the second stage 4

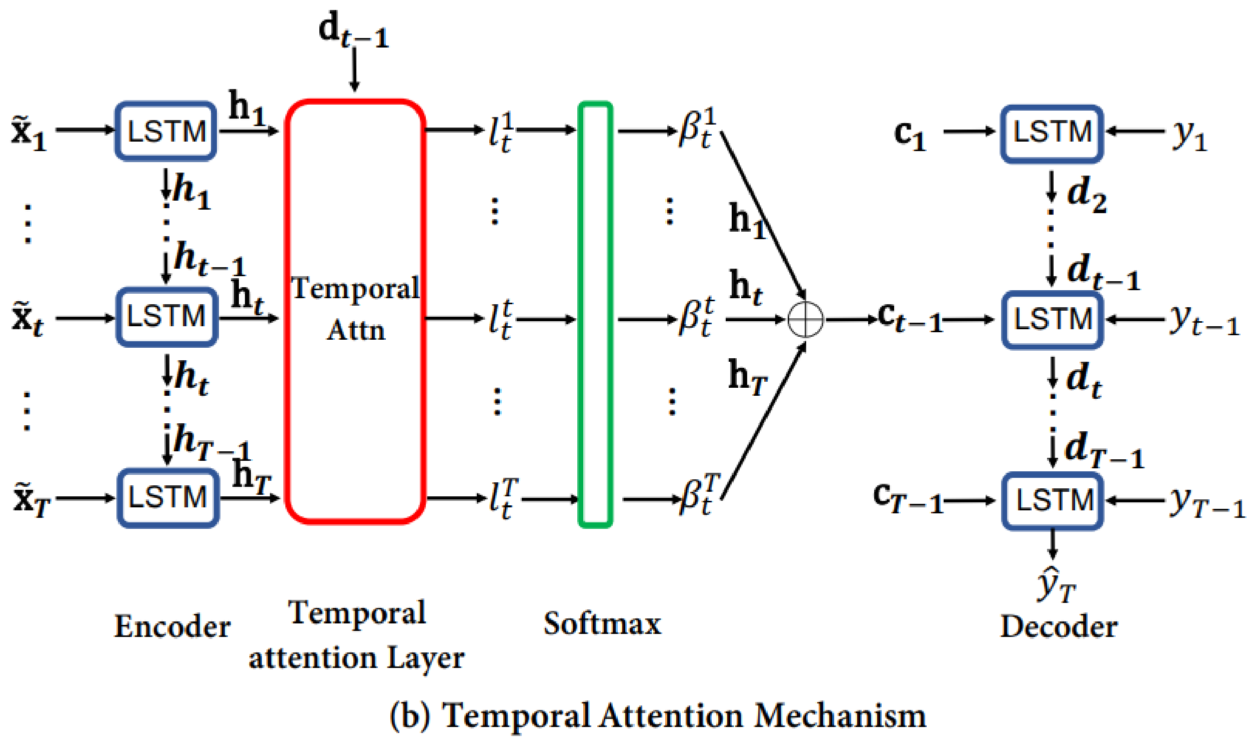


Figure 4: Seconds stage DA-RNN

In the second stage the inputs from stage one are input into an LSTM. They are used in an attention layer in conjunction with the previous decoder state. The output is fed into a softmax layer again. One then gets weights/attention for hidden states. The higher this β is the more important is the hidden timepoint at timepoint t. The weighted sum of attention and hidden states is then used in another LSTM to create predictions.

4.5 Informer

The informer [19] tries to overcome the deficiencies of the transformer. The paper names several improvements that are done in the new model called informer. The transformer is quadratic in memory and complexity, has problems with capacity when handling long inputs and is very slow when predicting long outputs. It solves the first problem by introducing *ProbSparse* as one can see in figure 5. The attention score is a dot-product between query, key and value. But only a few dot products actually contribute to the attention score. They measure the sparsity in the queries and only calculate the dot-product for the queries with high sparsity which reduces the complexity.

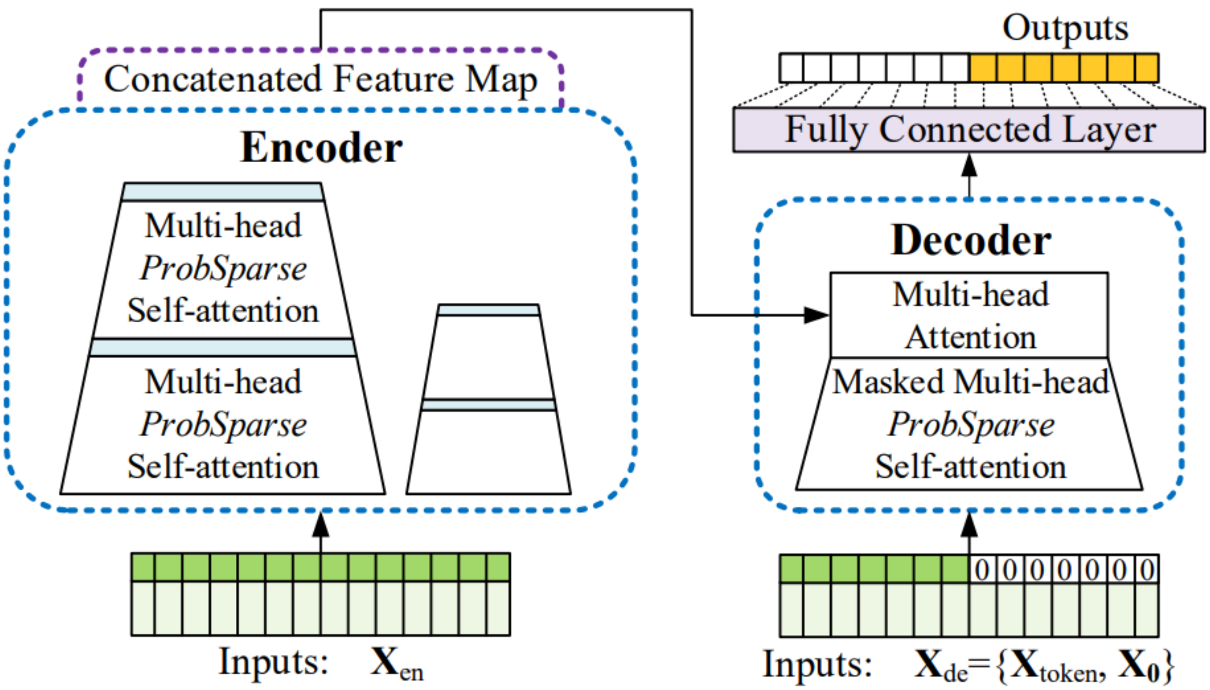


Figure 5: Seconds stage DA-RNN

More relevant for the experiment of my project is actually the second aspect. Long inputs are common in autoregressive models. This is solved by the informer. As you can see in figure 6 a Conv1d with a MaxPool layer is used in the informer module to halve the size of the attention block. This is referred to as self-attention distilling which reduces the size of the model drastically and therefore long inputs are no problem to the informer.

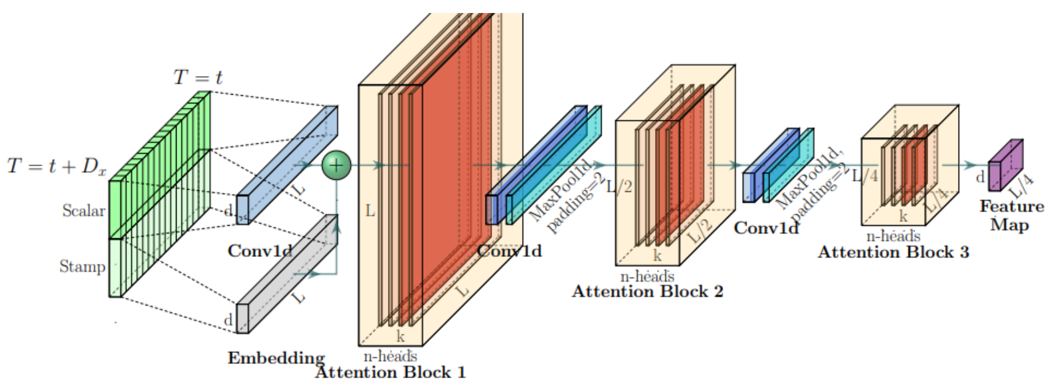


Figure 6: Seconds stage DA-RNN

5 Experiment

All models were used to predict the traffic volume by using the features. 10% of the data is used for validation and 20% of the data is used to calculate test accuracy. I measure RMSE, MAE and correlation in the test set. Results are reported in 6.

6 Results and Conclusion

Method	RMSE	MAE	Correlation
Linear Regression	0.3735	0.2478	0.9274
XGBoost	0.2952	0.1684	0.9553
LSTM	0.2838	0.2415	-0.5520
DA-RNN	0.1962	0.1372	0.9803
Informer	0.1926	0.1302	-

Deep learning methods perform better than Linear Regression and XGBoost. LSTM shows negative correlation. Both DA-RNN and Informer show the best performance. Both show great performance on a new data set.

7 Code and Implementation

The code can be found in the zip attached. Code snippets/models were taken from LSTM [7], DA-RNN [6] and Informer [20].

References

[1] Jinyoung Ahn, Eunjeong Ko, and Eun Yi Kim. "Highway traffic flow prediction using support vector regression and Bayesian classifier". In: *2016 International conference on big data and smart computing (BigComp)*. IEEE. 2016, pp. 239–244.

[2] Faraz Malik Awan, Roberto Minerva, and Noel Crespi. "Improving road traffic forecasting using air pollution and atmospheric data: Experiments based on LSTM recurrent neural networks". In: *Sensors* 20.13 (2020), p. 3749.

[3] Tianqi Chen and Carlos Guestrin. "Xgboost: A scalable tree boosting system". In: *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016, pp. 785–794.

[4] Dheeru Dua and Casey Graff. *UCI Machine Learning Repository*. 2017. URL: http://archive.ics.uci.edu/ml.

[5] Sepp Hochreiter and Jürgen Schmidhuber. "Long short-term memory". In: *Neural computation* 9.8 (1997), pp. 1735–1780.

[6] kaelzhang. *DA-RNN-in-Tensorflow-2-and-PyTorch*. https://github.com/kaelzhang/DA-RNN-in-Tensorflow-2-and-PyTorch. 2021.

[7] *lstm_multivariate_horizon_style.ipynb*. URL: https://colab.research.google.com/github/Apress/hands-on-time-series-analylsis-python/blob/master/Chapter%5C%207/5.%5C%20LSTM%5C%20Multivariate%5C%20Horizon%5C%20Style.ipynb.

[8] Yisheng Lv et al. "Traffic flow prediction with big data: a deep learning approach". In: *IEEE Transactions on Intelligent Transportation Systems* 16.2 (2014), pp. 865–873.

[9] H Zare Moayedi and MA Masnadi-Shirazi. "Arima model for network traffic prediction and anomaly detection". In: *2008 international symposium on information technology*. Vol. 4. IEEE. 2008, pp. 1–6.

[10] Yan Qi and Sherif Ishak. "A Hidden Markov Model for short term prediction of traffic conditions on freeways". In: *Transportation Research Part C: Emerging Technologies* 43 (2014), pp. 95–111.

[11] Yao Qin et al. "A dual-stage attention-based recurrent neural network for time series prediction". In: *arXiv preprint arXiv:1704.02971* (2017).

[12] Alexander Robles. *Deep Learning Time Series Forecasting*. https://github.com/Alro10/deep-learning-time-series. 2021.

[13] WY Szeto et al. "Multivariate traffic forecasting technique using cell transmission model and SARIMA model". In: *Journal of Transportation Engineering* 135.9 (2009), pp. 658–667.

[14] Ashish Vaswani et al. “Attention is all you need”. In: *Advances in neural information processing systems* 30 (2017).

[15] Eleni I Vlahogianni, John C Golias, and Matthew G Karlaftis. “Short-term traffic forecasting: Overview of objectives and methods”. In: *Transport reviews* 24.5 (2004), pp. 533–557.

[16] Eleni I Vlahogianni, Matthew G Karlaftis, and John C Golias. “Short-term traffic forecasting: Where we are and where we’re going”. In: *Transportation Research Part C: Emerging Technologies* 43 (2014), pp. 3–19.

[17] Rose Yu et al. “Deep learning: A generic approach for extreme condition traffic forecasting”. In: *Proceedings of the 2017 SIAM international Conference on Data Mining*. SIAM. 2017, pp. 777–785.

[18] Fan Zhou et al. “Variational graph neural networks for road traffic prediction in intelligent transportation systems”. In: *IEEE Transactions on Industrial Informatics* 17.4 (2020), pp. 2802–2812.

[19] Haoyi Zhou et al. “Informer: Beyond efficient transformer for long sequence time-series forecasting”. In: *Proceedings of AAAI*. 2021.

[20] zhouhaoyi. *Informer2020*. https://github.com/zhouhaoyi/Informer2020. 2020.