CS230

Predicting Retail Petrol Prices in Australian Restorative Markets (Time Series Forecasting)

Miles Perry
Department of Computer Science
Stanford University
mperry2@stanford.edu

Abstract

Major cities in Australia experience restorative cycles in retail petrol prices due to the competitive nature of these markets. This report presents a simple LSTM, a sequence-to-sequence LSTM and a WaveNet-inspired Temporal CNN (TCN) for time series forecasting of petrol prices in Australia. The hope is this will provide better visibility to consumers and businesses about upcoming changes in price, allowing them to plan ahead more efficiently. The TCN model performs best with a 1.38 % MAPE over a 60-day prediction window using the test data.

1 Introduction

Restorative price cycles for petrol have existed in Australian major cities for many years now. Price competition in petrol markets is high and petrol retailers compete aggressively on price until margins have been eroded. Below you can see the cyclical price patterns at a site in Sydney and how prices drop close to an average wholesale cost before jumping back up again.

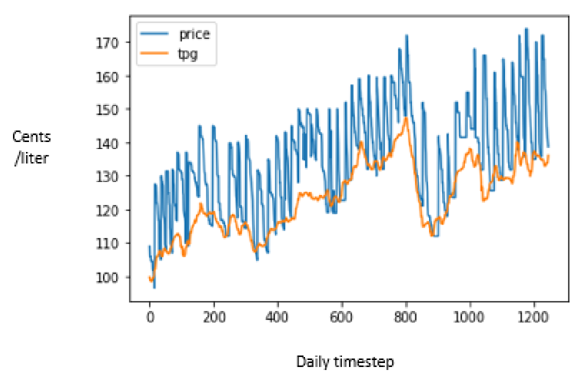

Figure 1: Prices and wholesale costs (tpg) for site in Sydney.

When margins are low the local market will wait for a retailer with a strongest brand image (usually a major oil brand) to raise prices significantly. When this happens, the other brands will follow. This is known as a restoration and the price cycle begins again. Price cycles are confusing for consumers and complex for fuel retailers to predict. An algorithm that can accurately predict petrol prices would

CS230: Deep Learning, Spring 2019, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

allow both consumers and fuel retailers to plan ahead. For example, the Australian Competition and Consumer Commission estimates the average motorist in Sydney could save $175 per year if they avoid buying at the peak price [1].

This project takes a simple LSTM as a baseline and compares it with a more complex WaveNet-inspired Temporal CNN (TCN) and a Sequence-to-Sequence (Seq2Seq) LSTM for predicting petrol prices in New South Wales, Australia. Initially I explored the TCN versus the baseline only. However, when I saw the prediction capability of the TCN I wanted to cross compare with a more complex LSTM architecture to better understand the strengths and weakness of the TCN. The baseline algorithm consists of a single layer LSTM with 8 units, connected to a single dense layer with 1 unit. The Seq2Seq2 LSTM uses a two-layer encoder-decoder network with 256 units per LSTM cell and Teacher Forcing. The TCN consists of multiple dilated causal convolutional layers with gated activations, residual and skip connections. All algorithms input a 600-day history of daily price data and use it to predict the next 60-day interval of daily prices.

The algorithms were written using Tensorflow 2.2.0 and executed in Google Colab. Code for this project can be found at https://github.com/milesperry81/CS230.

2 Related work

An article on Linked-In by Jacob Bourne [5] presents a simple LSTM to do next day petrol price predictions in the US using multi-variate inputs. Whilst a good inspiration for my baseline model I do not believe it represents state-of-the-art technology. Another article related to predicting petrol prices was made available by Claudio Stamile on Medium.com [6]. This article uses publicly available data in Italy to predict the next 14 days of petrol prices. The model used was a 1D dilated causal convolutional neural network. The results were interesting and there was a link to a GitHub repo for a similar project by Joseph Eddy on web traffic data using a WaveNet-inspired algorithm [12][16] and Seq2Seq LSTM. This repo formed the base code for this project.

An academic paper by Alberto Gasparin at al. called Deep Learning for Time Series Forecasting: The Electric Load Case [7] provides a good summary of different deep learning techniques for time series, including a WaveNet-inspired TCN and Sequence to Sequence LSTM. The conclusion was that TCN's are very promising for sequence modelling, and Seq2Seq LSTM's perform significantly better than the standard LSTM. A paper by Krist Papadopoulos [13] also suggests using a WaveNet-inspired TCN algorithm to perform time series forecasting with promising results. Deep Neural Networks Approach for Multivariate Time Series Forecasting by Renzhuo Wan et al. [14] explores the use of Multivariate TCN's and concludes promising results versus LSTMs. Conditional Time Series Forecasting with Convolutional Neural Networks by Anastasia Borovykh at al. [17], compares a WaveNet-inspired TCN architecture to a LSTM, concluding that the TCN is well suited to time series forecasting and outperforms the LSTM on financial index data.

There was sufficient evidence to show that I should explore TCN's and LSTM's for this project. TCN's are relatively unexplored for time series prediction. They promise faster computation and good, if not better, performance than traditional LSTM's for time series forecasting. LSTM models are well known for addressing time series forecasting and the Seq2Seq LSTM is a sophisticated LSTM variant that can be should be explored for cross comparison with the TCN. There is limited work where these approaches have been applied to petrol price prediction. Hence, for this project I chose to explore the Seq2Seq LSTM and TCN algorithms versus a simple LSTM baseline.

3 Dataset and Features

Fuel Check [2] is a regulatory body in the state of New South Wales. They collect price change data from all fuel retailers in the state. The price change data is available by site, by product, by date and time. Data from August 2016 to January 2019 was available in an immediately useable format. The data for one site can be seen in Figure 1 above alongside wholesale cost data from the AIP [3]. There is on average 1.5 price changes per site per day in New South Wales. As such the data was manipulated into daily timesteps using SQL. Sites with zero prices on their first timestep were removed due to a general lack of price data at later time steps. Data from September 2018 was excluded due to a big drop in wholesale cost affecting the price cycles.

The time series data was segmented into smaller example windows of 600-day history and 60-day prediction intervals. Examples were then split 80/10/10 percent into train/validation/test data. In all there were 10456/1960/1961 examples for train/validation/test based on 1,307 sites in New South Wales, including Sydney. All input data was normalized using mean and standard deviation.

4 Methods

Baseline LSTM

The baseline LSTM consists of one LSTM layer (8 units) and one dense layer (1 unit). It predicts one timestep ahead. A walk forward approach is then used to feed the predictions back into the model to generate a 60-day prediction interval.

Sequence-to-Sequence (Seq2Seq) LSTM

The Seq2Seq algorithm has two LSTM layers with 256 units in each cell. Figure 3 below shows a similar RNN architecture as that used in this project.

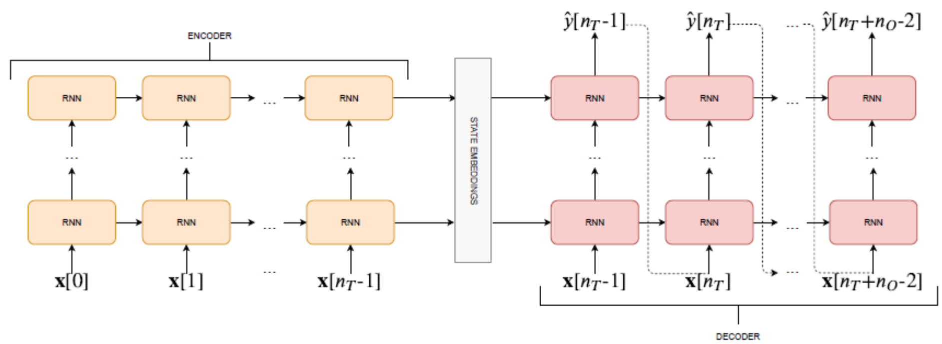

Figure 3. Seq2Seq RNN architecture [7].

During training Teacher-Forcing was used, such that the input to the first decoder timestep is the input to the last encoder timestep and for subsequent decoder timesteps the actual target value from the previous timestep is input, instead of the decoder output from previous timestep. Teacher-Forcing assists training and speeds up optimization convergence. When making predictions Teacher Forcing is not used and the output from each decoder timestep is passed as input to the next decoder timestep as demonstrated in Figure 3. The final LSTM states from the encoder are used to initialise the decoder states.

Temporal Convolutional Network (TCN)

The algorithm uses dilated causal convolutions in combination with WaveNet-inspired gated activations, skip connections and residual connections. Dilated convolutions skip a number of input values at each layer of the network. Thus, providing a coarse exposure to a larger than usual receptive field for later layers of the network. The causal component ensures the current time step cannot depend on future time steps. A dilated causal convolutional network is illustrated in Figure 4.

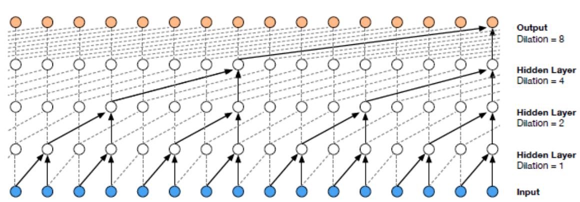

Figure 4. Dilated causal CNN layers [16].

There is a repeated residual block that runs the convolutions with different dilation rates. A residual connection combines the input of the block with the output, which is then passed to the next residual block. Within the block there is a gated activation unit consisting of a ReLU filter and sigmoid gate. This algorithm uses a ReLU activation for the filter, unlike the WaveNet paper which use tanh. This was found to yield better performance for this project. The gate and filter are multiplied together to form the output and the skip connection output. After the residual blocks, the skip connections are combined and fed into the final convolutional layers with dropout. The WaveNet residual block architecture can be seen in Figure 5. The last 60 timesteps of the final 1D convolutional layer are extracted and used as the prediction interval.

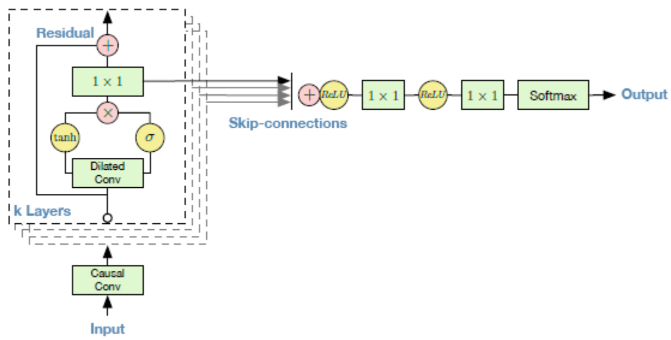

Figure 5. WaveNet residual block architecture [16].

Loss Function

An Adam optimizer was used with Mean Absolute Error (MAE) [18] for the loss function to reduce the loss between the numerical prediction and actual target.

5 Experiments/Results/Discussion

A learning rate scheduler was used to find a good starting value for the learning rate (1e-3) before training. A batch size of 128 was found to be suitably fast at converging to an optimum. MAE and Mean Absolute Percentage Error (MAPE) [19] were used to evaluate the model in percentage and absolute price terms. The historic input time frame was 600 days with a 60 day prediction interval.

Some notable hyperparameter choices made during experimentation can be seen in Figure 6.

TCN (100 epochs)

Index	Dilated filter size	Dilated Conv1D channels	Gate filter activation	Dilations Rates	Validation MAE/MAPE
1	2	32	tanh	[1,2,4,8,16,32,64,1,2,4,8,16,32,64]	2.86 / 2.02%
2	2	64	tanh	[1,2,4,8,16,32,64,1,2,4,8,16,32,64]	2.64 / 1.86%
3	5	64	tanh	[1,2,4,8,16,32,64,1,2,4,8,16,32,64]	2.10 / 1.47%
4	5	64	ReLU	[1,2,4,8,16,32,64,1,2,4,8,16,32,64]	2.05 / 1.44%
5	5	64	ReLU	[1,2,4,8,16,32,64,1,2,4,8,16,32,64,1,2,4,8,16,32,64]	1.99 / 1.39%

Seq2Seq LSTM

Index	Epochs	LSTM units	LSTM layers	Validation MAE/MAPE
1	100	50	1	4.68 / 3.18%
2	100	256	1	4.36 / 3.01%
3	100	256	2	5.26 / 3.61%
4	400	256	1	3.59 / 2.48%
5	400	256	2	3.43 / 2.44%

Figure 6. Notable hyperparameters and results.

The final results for all models on the validation and test data can be seen in Figure 7.

	100 Epochs				400 Epochs			
Architecture	Validation, MAE (cents/litre)	Validation, MAPE (%)	Test, MAE (cents/litre)	Test, MAPE (%)	Validation, MAE (cents/litre)	Validation, MAPE (%)	Test, MAE (cents/litre)	Test, MAPE (%)
Baseline LSTM	8.26	5.84	8.18	5.80	6.37	4.48	6.37	4.48
Seq2Seq LSTM	5.26	3.61	5.21	3.58	3.43	2.44	3.33	2.38
TCN	1.99	1.39	2.02	1.41	1.91	1.34	1.96	1.38

Figure 7. Table of results.

The baseline LSTM shows limited ability to model the petrol price data beyond the first timestep. Predictions tend to drift off in a single direction. This is reflected in the higher MAE and MAPE.

The Seq2Seq LSTM shows improved MAE and MAPE results over the baseline. Additional layers and longer training were required to see improvements in the results (see Figure 6 and 7). At 400 epochs it is good at modelling price cycles, but less so where the cycles are not well defined. See Figure 8.

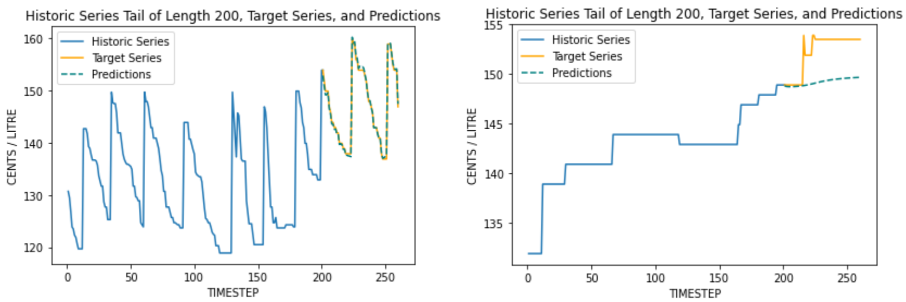

Figure 8. Seq2Seq LSTM 60-day prediction, 400 epochs.

The TCN models the price cycles very well and has the best MAE and MAPE results. It trained and learned faster than the Seq2Seq LSTM. More dilated layers with 64 channels and a filter size of 5 yielded best results during training (see Figure 6). Predictions on less well-defined cycles are better than the Seq2Seq LSTM. See Figure 8.

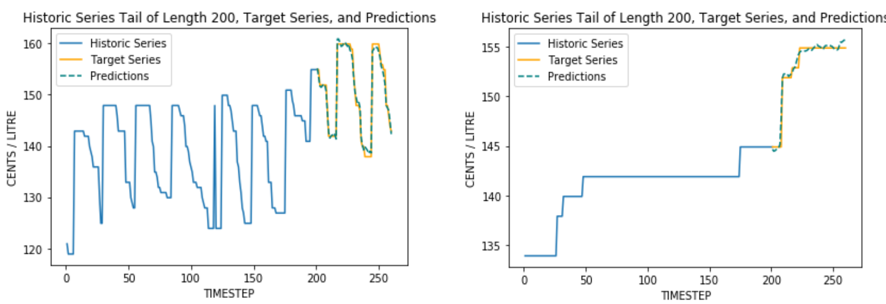

Figure 9. TCN 60-day prediction, 400 epochs.

6 Conclusion/Future Work

The Seq2Seq LSTM predictions were good, but showed larger error on less well-defined cycles. The TCN model shows a superior forecasting ability and faster training times when compared to the other models. Its predictions closely follows price cycles of varying wavelengths and amplitudes. It is likely that the deep architecture and dilated convolutions allows it to learn more complex patterns and take longer time frames into account. The relatively fewer number of parameters in the TCN make it computation more efficient than the Seq2Seq2 LSTM.

Some considerations for future work:

- Train the Seq2Seq LSTM longer. Can it reach the TCN performance?
- Try shorter historic time windows. What is the accuracy with less historic input data?
- Make predictions on petrol prices from a different distribution? E.g. Queensland.
- How good is the TCN at predicting non-restorative fuel prices, such as diesel?

7 Contributions

This project was the work of a single contributor, Miles Perry.

References

[1] Australian Competition and Consumer Commission. “Petrol Price Cycles in Australia”. https://www.accc.gov.au/system/files/ACCC%20Petrol%20price%20cycles%20in%20Australia%E2%80%94 December%202018.pdf

[2] Fuel Check. https://data.gov.au/dataset/ds-nsw-a97a46fc-2bdd-4b90-ac7f-0cb1e8d7ac3b/details?q=.

[3] Australian Institute of Petroleum. https://www.aip.com.au/historical-ulp-and-diesel-tgp-data.

[4] David P Byrne et al. "Australia Has the World’s Best Petrol Price Data: FuelWatch and FuelCheck". The Australian Economic Review, vol. 51, no. 4, pp. 564–577.

[5] Jacob Bourne. "Using Machine Learning to Predict Retail Gasoline Prices". https://www.linkedin.com/pulse/using-machine-learning-predict-retail-gasoline-prices-jacob-bourne/

[6] Claudio Stamile. "Seq2Seq in Keras for Petrol Price Prediction using Italian Open Data". https://medium.com/isiway-tech/seq2seq-in-keras-for-petrol-price-prediction-using-italian-open-data-767cb1016af3.

[7] Alberto Gasparin at al. "Deep Learning for Time Series Forecasting The Electric Load Case". arXiv:1907.09207v1 [cs.LG] 22 Jul 2019.

[8] Yongli Zhu et al. "Power Market Price Forecasting via Deep Learning". arXiv:1809.08092v2.

[9] Tensorflow. https://www.tensorflow.org/tutorials/structured_data/time_series#the_weather_dataset.

[10] Feng Xu et al. "Time Series Forecasting Models for Gasoline Prices in China". https://doi.org/10.5539/ijef.v10n12p43.

[11] Jason Brownlee. "How to Develop LSTM Models for Time Series Forecasting". https://machinelearningmastery.com/how-to-develop-lstm-models-for-time-series-forecasting/

[12] Joseph Eddy. “High Dimensional Time Series Forecasting with Convolutional Neural Networks”. https://github.com/JEddy92/TimeSeries_Seq2Seq/blob/master/notebooks/TS_Seq2Seq_Conv_Intro.ipynb.

[13] Krist Papadopoulos. “SeriesNet: A Dilated Causal Convolutional Neural network for Forecasting”. https://github.com/kristpapadopoulos/seriesnet.

[14] Renzhuo Wan et al. “A Deep Neural Networks Approach for Multivariate Time Series Forecasting”. https://www.mdpi.com/2079-9292/8/8/876.

[15] Coursera TensorFlow in Practice Specialisation. https://www.coursera.org/specializations/tensorflow-in-practice.

[16] Aaron van den Oord et al. “WaveNet: A Generative Model For Raw Audio”. arXiv:1609.03499v2 [cs.SD] 19 Sep 2016.

[17] Anastasia Borovykh at al., “Conditional Time Series Forecasting with Convolutional Neural Networks”. arXiv:1703.04691v5 [stat.ML] 17 Sep 2018.

[18] Mean Absolute Error. https://en.wikipedia.org/wiki/Mean_absolute_error.

[19] Mean Absolute Percentage Error. https://en.wikipedia.org/wiki/Mean_absolute_percentage_error.