Univariate Time Series Forecasting Using Classical Methods and Deep Learning Algorithms

Avesta Ahmadi

McMaster University

ahmada59@mcmaster.ca

December 8, 2019

Overview

Introduction

Time series forecasting:

1. Univariate time series forecasting
   - Single observation recorded sequentially over time
   - Classical methods: SARIMA, ARIMA, Exponential Smoothing, ...
   - Deep learning approaches:
     - ✓ MLP
     - ✓ Memory-based networks (LSTM, RNN, GRU & Bidirectional LSTM)
     - ✓ Convolutional networks (via Causal Convolutions)
     - ✓ SeriesNet (WaveNet based structure)
2. Multivariate time series forecasting
   - Multiple observations recorded sequentially over time that are dependent to each other
   - Classical methods: VAR, VARMA

Descriptive Analysis

- 12 normalized time series datasets, with 75% train/test split

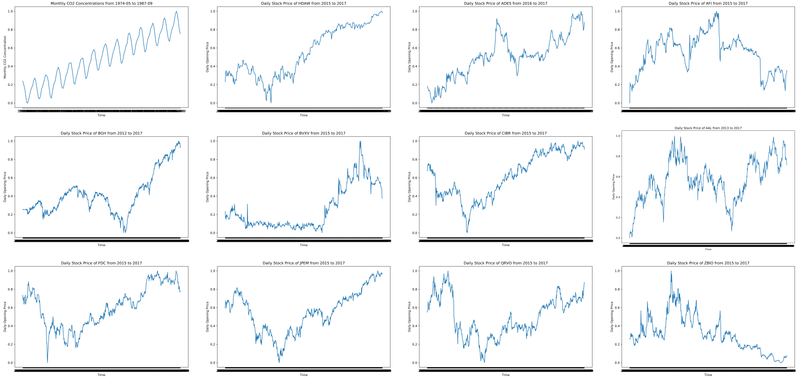


Figure: 12 various time series datasets[1]

[1] Sources: https://www.kaggle.com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs, https://github.com/PacktPublishing/Practical-Time-Series-Analysis/blob/master/Data

SARIMA

SARIMA Model:

- Seasonal Auto-Regressive Integrated Moving Average model.
- Stationary vs. non-stationary data.
- Stationary data is iid noise.
- Detecting stationarity: Correlation plots and unit root tests(ADF test).
- SARIMA(p,d,q)(P,D,Q,S).
- Determining optimal orders of parameters on SARIMA model based on maximum likelihood estimation.

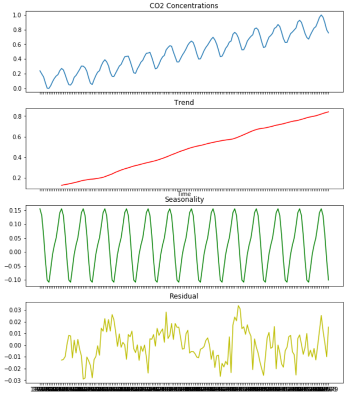
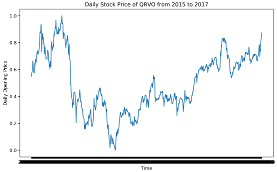


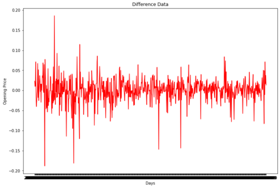
Figure: Monthly CO_2 Concentration data is decomposed into its components

SARIMA

Stationary vs. non-stationary:



Daily opening price of QRVO is shown in the plot. ACF and PACF plots demonstrate that there is significant correlation between lags.



First order differencing of timeseries can take out trend and make it stationary.

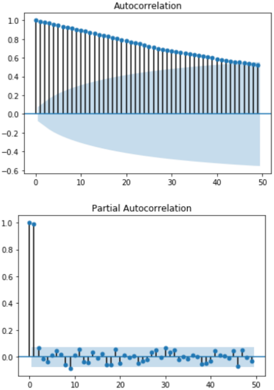


Figure: ACF and PACF plots for QRVO time series up to lag 50

SARIMA model:

- No seasonality.
- ARIMA(2,1,2) model
- Residuals of the model have to be iid noise.

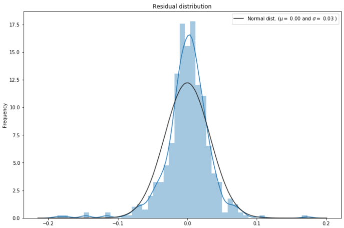


Figure: Distribution of residuals of SARIMA model(blue line) vs. normal distribution(black line)

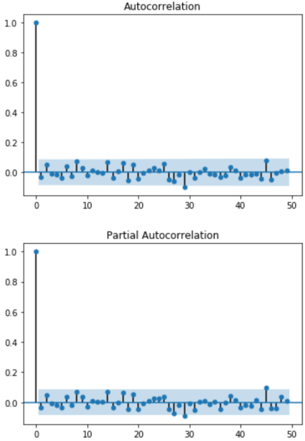


Figure: ACF and PACF plots for residuals of SARIMA model

Memory-based Networks

Methods

1. GRU
2. RNN
3. Deep stacked unidirectional LSTM
4. Deep stacked bidirectional LSTM
    - Extending the memory-based networks to bidirectional ones
    - Can be trained without the limitation of using input information just up to a preset future frame
    - Training the network in a simultaneously positive and negative time directions

Bidirectional LSTM

BD-LSTM

- Contains a forward and a backward LSTM layer
- Forward layer output is calculated using inputs in a positive sequence, while the backward layer uses reversed inputs.
- The σ function combines the two output sequences. It can be concatenating, summation, averaging or multiplying the outputs.

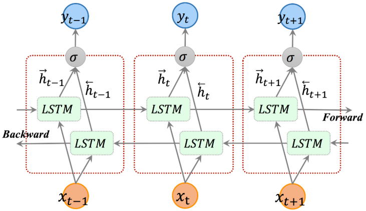


Figure: Structure of a bidirectional layer with three consecutive steps

Convolutional Networks

CNN-LSTM

- Using convolution layers for feature extraction followed by stacked layers of LSTM
- 1D filters are used for univariate time series data.

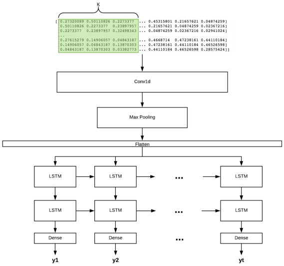


Figure: CNN-LSTM network structure[2]

[2] Source: https://towardsdatascience.com/get-started-with-using-cnn-lstm-for-forecasting-6f0f4dde5826

Convolutional Networks

Causal Convolutions

- To assure the model cannot violate the ordering in which we model the data, causal convolutions are introduced.
- Prediction at timestep t cannot depend on any future timesteps.

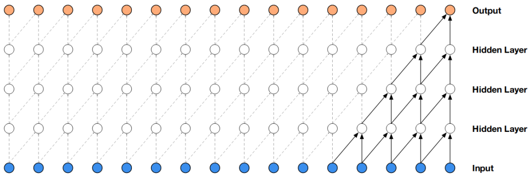


Figure: Causal convolution networks for sequence modelling

Convolutional Networks

Dilated Convolutions

- Dilated convolution is a convolution where the filter is applied over an area greater than its length by skipping some input values with a certain step.
- dilated convolutions are used to increase the receptive field by orders of magnitude, without significantly increasing computational cost.

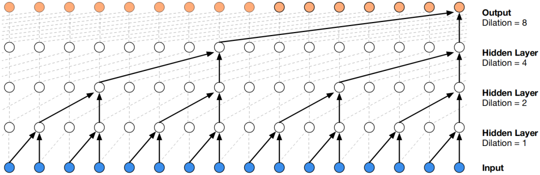


Figure: Dilated causal convolution networks for sequence modelling

Convolutional Networks

WaveNet/SeriesNet

- SeriesNet structure is inspired by WaveNet for time series forecasting.
- Dilated causal convolutions are used in convolution layers.
- The choice of activation function is different in two architectures.

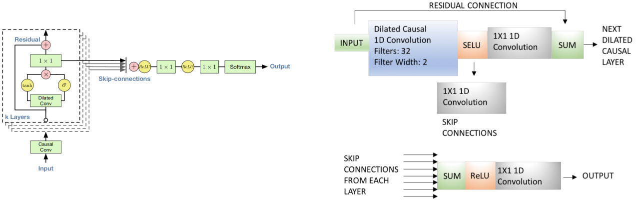


Figure: WaveNet vs. SeriesNet structures for time series prediction

Results & Discussion

Comparing Memory-based Networks

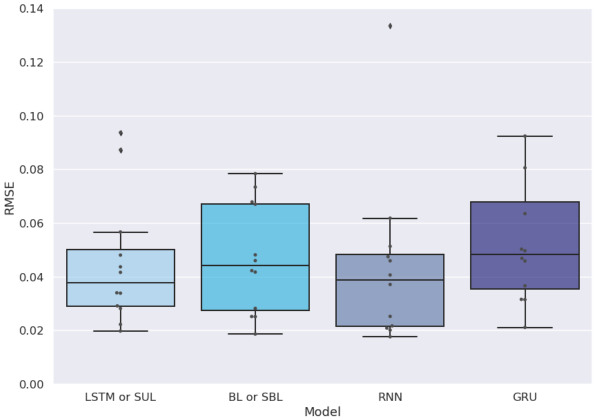


Figure: Performance evaluation of memory-based networks

Results & Discussion

Comparing Memory-based Networks

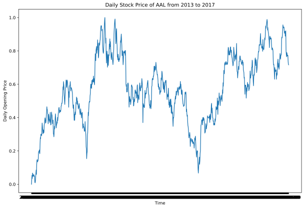


Figure: AAL dataset

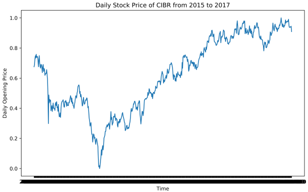


Figure: CIBR dataset

- RMSE of SUL: **0.033**
- RMSE of SBL: 0.043

- RMSE of SUL: 0.035
- RMSE of SBL: **0.025**

Results & Discussion

Comparing classical and deep learning approaches

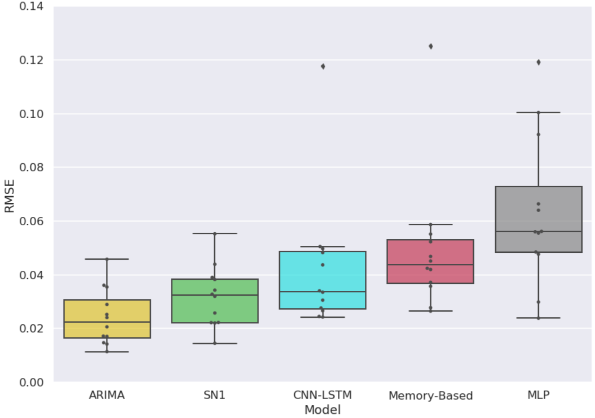


Figure: Performance evaluation of all methods

Results & Discussion

Pros and Cons:

- NNs have no assumptions about the mapping function (linear or nonlinear). They are robust to noise in input data (learning in the presence of missing values).
- NNs can support an arbitrary but fixed number of inputs and outputs in the mapping function, Hence, providing direct support for multivariate prediction and multi-step forecast.
- Data decomposition (prior deseasonalization) step should be employed prior to analysis for NNs. Covolution-based schemes are better at modelling seasonality.
- Dilated Causal Convolutions are designed to capture long-range dependencies effectively along the temporal dimension.
- NN models often suffer from outlier errors for certain time series.

References

Brownlee, Jason. *Long Short-Term Memory Networks with Python: Develop Sequence Prediction Models with Deep Learning*. Jason Brownlee, 2017.

Cui, Zhiyong, Ruimin Ke, and Yinhai Wang. *"Deep stacked bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction."* 6th International Workshop on Urban Computing (UrbComp 2017). 2016.

Christoph, et al. *Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions*. ArXiv.org, 24 Sept. 2019, arxiv.org/abs/1909.00590.

Chen, et al. *Probabilistic Forecasting with Temporal Convolutional Neural Network*. ArXiv.org, 11 June 2019, arxiv.org/abs/1906.04397v1.

Oord, et al. *WaveNet: A Generative Model for Raw Audio*. ArXiv.org, 19 Sept. 2016, arxiv.org/abs/1609.03499.

Borovykh, Anastasia, et al. *Dilated Convolutional Neural Networks for Time Series Forecasting*. Journal of Computational Finance, 2018, doi:10.21314/jcf.2019.358.