Univariate Time Series Forecasting Using M4 Competition Deep Learning Architectures

Yves Greatti, Zian Chen, Sunjoo Park

Due on October 16th - Revised version

Abstract

Forecasting time series data is a challenging but critical task and has many applications in different industries such as finance, media or biomedical. In addition, complex and non-linear interdependencies between time steps and series complicate the task. Traditional methods have mostly be unsuccessful in modeling complex patterns or dependencies. In this project, we aim to try different deep learning models on univariate time series point forecasting. The M4 Competition large data set - 100,000 time series - is our reference dataset. The dataset comes from a wide range of industries including Economics, Finance, Tourism, Trade, Real Estate, and so on, aiming at representing as much real world data as possible (M4-dataset). We will start using a classical statistical models such as SARIMA, or Holt-Winters exponential smoothing to build our benchmark and use deep learning architectures: N-BEATS [2] and ES-RNN [3] to compare how performances improve differently. We will report various accuracy metrics: sMAPE (symmetric Mean Absolute Percentage Error), MASE (Mean Absolute Scaled Error) or OWA (overall weighted average).

1   Related work

Named after the forecasting researcher Spyros Makridakis, the M Competition has been one of the most important events since 1982 in the forecasting community. The competitions compare the accuracy of different time series forecasting methods including the most advanced statistical models. The winner of 2018 competition is called ES-RNN (Exponential Smoothing-Recurrent Neural Network), and is an hybrid approach of attention based dilated LSTM (Long Short-Term Memory) with a classical Holt-Winters model. N-BEATS is a model that purely relies on deep learning, without any combination with statistical models. Specifically, this model challenges ES-RNN model's heavy dependence on the Holt-Winters component. The M4 dataset consists in 100k-series dataset of data frequently encountered in business, financial and economic forecasting and aggregated in sampling frequencies ranging from hourly to yearly. There are six

files in each training and test datasets: Hourly, Daily, Weekly, Monthly, Quarterly and Yearly with anonymized features. The dataset has two significant characteristics: first, the lack of temporal or other connections between observations. Second, each series is variable in length but at times very small, making the forecasting more difficult.

The predictive accuracy metrics in the context of the competition are:

$$\text{sMAPE} = \frac{1}{H}\sum_{i=1}^{H} 2\frac{y_{T+i} - \hat{y}_{T+i}}{|y_{T+i}| + |\hat{y}_{T+i}|}, \ \text{MASE} = \frac{1}{H}\sum_{i=1}^{H} \frac{y_{T+i} - \hat{y}_{T+i}}{\frac{1}{T+H-m}\sum_{j=m+1}^{T+H} |y_j - y_{j-m}|}$$

$$\text{OWA} = \frac{1}{2}\big[\text{sMAPE}/\text{sMAPE}_{\text{Naive2}} + \text{MAPE}/\text{MAPE}_{\text{Naive2}}\big]$$

where T is the length of the history, H is the length of the forecast horizon, and m is the periodicity of the data.

2 Our contribution

We propose to run different experimentations on the M4 dataset with the two main architectures: ES-RNN and N-BEATS, comparing the results obtained with these two models with the benchmark we established using a classical statistical model. The hyper-parameters of these two networks are carefully tuned with hand-crafted validation data sets which we will replicate. In addition, ES-RNN expects to have all the series of equal length, and use a data windowing followed by a one-hot representation. We will report for the two networks the optimal settings of the parameters, and maybe make progress on open tasks reported by each paper. For example, the authors of the ES-RNN mention, that being able to support variable length series will be critical as shorter series could be introduced. Another important factor to consider for these models, is the amount of regularization needed: too much and we prevent the model from learning, and too little might cause to fall into a local optima.

Also these networks consists in complex stacking of layers, one of our goal is to quantify the interpretability of the results provided by each model, by being able to decompose each series into trend and seasonality components. In this regard, we want to investigate a neural network technique called Neural Decomposition (ND) [1]. This sort of neural network decomposes a series into a sum of sinusoids, inspired by Fourier transform, to capture linear trends and other non-periodic components of each series and might perform better than the two previous models.

The work will be organized among us in this respect:

(a) **Zian Chen**: will focus on implementing N-BEATS on M4 datasets, tuning hyper-parameters and comparing the results of N-BEATS with the baseline models.

(b) **Sunjoo Park**: will implement ES-RNN model using GPU on M4 datasets, optimize the model tuning hyper-parameters and differentiate model performances.

(c) **Yves Greatti**: will set up the baseline models using SARIMA and Holt-Winters exponential smoothing, helps if necessary in settings the two networks, and the optimization of the hyper-parameters and will investigate the Neural Decomposition applied to the M4 datasets.

3 Conclusion

Our aim is to reproduce the results achieved by -ES-RNN and N-BEATS, improve upon their performances by a better optimization of their hyper-parameters, or using a Neural Decomposition.

References

[1] Luke B. Godfrey and Michael S. Gashler. Neural decomposition of time-series data for effective generalization. *CoRR*, abs/1705.09137, 2017.

[2] Boris N. Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. N-BEATS: neural basis expansion analysis for interpretable time series forecasting. *CoRR*, abs/1905.10437, 2019.

[3] Andrew Redd, Kaung Khin, and Aldo Marini. Fast ES-RNN: A GPU implementation of the ES-RNN algorithm. *CoRR*, abs/1907.03329, 2019.