Literature Review - Univariate Time Series: Air Traffic Forecasting using Motifs and LSTM

Student ID:	20211232, 20210360
Student name:	Gayathri Sridhar, Meenakshi Srinivasan
Student email	gayathri.sridhar2@mail.dcu.ie, meenakshi.srinivasan2@mail.dcu.ie
Chosen major:	Data Analytics
Supervisor	Dr. Martin Crane
Date of Submission	06-02-2021

Introduction

Statistical models have been used as a conventional method in various domains for forecasting. Time series analysis considers the past behaviour of data over time. Previous research works on time series analysis were majorly focused on financial domain such as portfolio management [10]. In this paper we aim to do time series analysis on business level, in aviation industry. As air travel is becoming more popular mode of transport, forecasting the passenger counts could benefit the passengers as well as the operating airlines to meet the raising demands. Due to the seasonal and cyclic nature of this industry, many research works were based on the Holt Winters Exponential Smoothing algorithm (HWES) [2]. As an alternative, Long Short Term Memory (LSTM) [7] model have been used in our paper. Motifs [3] are sub sequences /patterns occurring in a time series which could be used to enhance the performance of the model and hence we have chosen to implement it alongside LSTM.

Related work

Tularam, A. and Saeed, T. (2016) [1], implemented not only a number of components of ARIMA but also Exponential Smoothing for forecasting. ARIMA model depends on regressing a variable on values over time. Exponential Smoothing is used in cases where the data pattern shows no particular trend or seasonal variation.

Deetchiga, S. et al., (2018) [2] explored the future trends of air travel using Holt-Winters Exponential Smoothing (HWES) model. As the air traffic industry involves seasonality and trends, HWES was one of an appropriate model.

Matrix profile algorithm [3] is one of the most efficient algorithms to discover motifs. Eamonn Keogh, the pioneer of matrix profile developed various algorithms such as STAMP [3], STOMP [3] and SCRIMP++ [4]. Scalable Time Series Anytime Matrix Profile (STAMP) is an anytime algorithm which assesses the distance profiles in random order, whereas Scalable Time Series Ordered-Search Matrix Profile (STOMP) performs an ordered search. SCRIMP++ combines the properties of both STAMP and STOMP algorithm and produces the correct output in interactive sessions.

As the above mentioned algorithms require a lot of computational time, Law, S. (2019) [5] came up with STUMPY algorithm

which is a powerful and scalable library that computes matrix profile efficiently to discover motifs. STUMPY makes use of parallel processing which does ordered search for sub sequences in a time series to minimise the runtime.

Long Short Term Model (LSTM) is a form of the Recurrent Neural Network used to retrieve information from sequential data. It plays a major role in forecasting real time data ranging from IOT systems, cyber-networks, to industrial systems and healthcare. As the real data is often complicated with anomalies and change points, which can lead the learned models deviating from the underlying patterns of the time series, Guo,T., et al.,(2016) [6] explored the local features of time series to automatically weight the gradients of the loss of the newly available observations with distributional properties of the data in real time.

Liu, Y., et al., (2019) [7] implemented LSTM on time series and found that it outperformed traditional Auto-Regression (AR) models. Yadav, A., et al., (2020) [8] emphasises the importance of fine tuning the hyper parameters of LSTM model. They modified the number of hidden layers and also provided the performance of the model by comparing the stateless and stateful models.

Barry, B. and Crane, M. (2019) [9] have incorporated motifs in the LSTM neural network and compared the forecast accuracy of the LSTM model with motifs to standard LSTM models. They have proved that the LSTM model with motifs have reduced the RSME value by 8% than that of the standard one.

Conclusion

Since we are doing a Univariate time series analysis, various papers involving different time series models were considered. The distinguishing features of ARIMA, HWES and Exponential Smoothing models were explored. In [2], implemented HWES method and used seasonal parameters to handle the trend and seasonality in air passengers' data. In order to obtain seasonal index of all airports for accurate capacity planning and maximizing productivity of airports, Holts Linear Exponential Smoothing method with two parameters was used for passenger traffic prediction in [11]. As previous papers related to air traffic industry focused on forecasting via smoothing, LSTM approach is implemented in our paper. In addition, papers describing the fine tuning of hyper parameters of LSTM model were analysed. As an alternative approach, we are using motifs to enhance the performance of LSTM model. STUMPY, a novel algorithm for time series mining using matrix profile with minimal runtime, compared to STAMP and STOMP algorithms, is used in our paper.

References

[1] Tularam, A. and Saeed, T. (2016) "Oil-Price Forecasting Based on Various Univariate Time-Series Models", American Journal of Operations Research, 06, pp. 226-235.

[2] Deetchiga, S., Harini, U. K., Marimuthu, M. and Radhika, J. (2018) "Prediction of Passenger Traffic for Global Airport using Holt's Winter Method in Time Series Analysis," 2018 International Conference on Intelligent Computing and Communication for Smart World (I2C2SW), Erode, India, 2018, pp. 165-169

[3] Y. Zhu et al., "Matrix Profile II: Exploiting a Novel Algorithm and GPUs to Break the One Hundred Million Barrier for Time Series Motifs and Joins," 2016 IEEE 16th International Conference on Data Mining (ICDM), Barcelona, 2016, pp. 739-748, doi: 10.1109/ICDM.2016.0085

[4] Zhu,Y.,Yeh,C.M.,Zimmerman,Z.,Kamgar, K. and Keogh,E. "Matrix Profile XI: SCRIMP++: Time Series Motif Discovery at Interactive Speeds," 2018 IEEE International Conference on Data Mining (ICDM), Singapore, 2018, pp. 837-846.

[5] Law, S. (2019) "STUMPY: A Powerful and Scalable Python Library for Time Series Data Mining", Journal of Open Source Software, 4(39), 1504.

[6] Guo, T., Xu, Z., Yao, X., Chen, H., Aberer, K. and Funaya, K., "Robust Online Time Series Prediction with Recurrent Neural Networks," 2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA), Montreal, QC, 2016, pp. 816-825.

[7] Liu, Y., Su, Z., Li, H. and Zhang, Y. (2019) "An LSTM based classification method for time series trend forecasting," *2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, Xi'an, China, 2019, pp. 402-406.

[8] Yadav, A., Jha, C K., Sharan, A.,"Optimizing LSTM for time series prediction in Indian stock market,Procedia Computer Science", Volume 167,2020, Pages 2091-2100,ISSN 1877-0509.

[9] Barry, B. and Crane, M. (2019) "Analysis of Cryptocurrency Commodities with Motifs and LSTM", AICS 2019:27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science, National University of Ireland, Galway, 5-6 December 2019. pp. 28-39.

[10] Luo, C. and Wu,D., "Environment and economic risk: An analysis of carbon emission market and portfolio management", Environmental Research, Volume 149, 2016,Pages 297-301,ISSN 0013-9351.

[11] Önder, E. and Kuzu, S., "Forecasting Air Traffic Volumes Using Smoothing Techniques", Journal of Aeronautics and Space Technologies, January 2014, Volume 7, Number 1 (65-85).