A Sequence to Sequence Long Short-Term Memory Network for Footwear Sales Forecasting

Luís Santos[1], Luís Miguel Matos[4], Luís Ferreira[1], Pedro Alves[2], Mário Viana[3], André Pilastri[1], and Paulo Cortez[4](✉)

[1] EPMQ, CCG ZGDV Institute, Guimarães, Portugal
{luis.santos,luis.ferreira,andre.pilastri}@ccg.pt
[2] KYAIA - SOLUÇÕES INFORMÁTICAS, LDA, Guimarães, Portugal
pedro.alves@ksi.pt
[3] OVERCUBE, S.A., Guimarães, Portugal
marioviana@m360.com.br
[4] ALGORITMI Research Centre/LASI, Department of Information Systems, University of Minho, Guimarães, Portugal
{luis.matos,pcortez}@dsi.uminho.pt

Abstract. Footwear sales forecasting is a critical task for supporting product managerial decisions, such as the management of footwear stocks and production levels. In this paper, we explore a recently proposed Sequence to Sequence (Seq2Seq) Long Short-Term Memory (LSTM) deep learning architecture for multi-step ahead footwear sales Time Series Forecasting (TSF). The analyzed Seq2Seq LSTM neural network is compared with two popular TSF methods, namely ARIMA and Prophet. Using real-world data from a Portuguese footwear company, several computational experiments were held. Focusing on daily sales, we analyze data recently collected during a 3-year period (2019–2021) and related with seven types of products (e.g., sandals). The evaluation assumed a robust and realistic rolling window scheme that considers 28 training and testing iterations, each related with one week of multi-step ahead predictions. Overall, competitive predictions were obtained by the proposed LSTM model, resulting in a weekly Normalized Mean Absolute Error (NMAE) that ranges from 5% to 11%.

Keywords: Time series forecasting · ARIMA · Prophet · Deep learning

1 Introduction

The accurate projection of sales is a crucial element to support inventory management systems. Indeed, inventory excesses or shortages are often the result of expectations not being met, which have an immediate detrimental effect on the company's profitability and competitiveness.

This work focuses on a Portuguese footwear company online store that sells several footwear products (e.g., Shoes, Sneakers) across Europe. By adopting

H. Yin et al. (Eds.): IDEAL 2022, LNCS 13756, pp. 465–473, 2022.
https://doi.org/10.1007/978-3-031-21753-1_45

a Time Series Forecasting (TSF) approach, there is a potential to better support the inventory management system of the analyzed company (e.g., reducing stock costs). Moreover, in recent years there has been a growing interest in the usage of deep learning architectures to perform TSF tasks, such as Long Short-Term Memory (LSTM) networks. In this paper, we focus on a recently proposed Sequence (Seq2Seq) LSTM neural network [5], aiming to predict footwear sales. The adopted LSTM is compared with two popular Time Series Forecasting (TSF) methods, namely the Auto-Regressive Integrated Moving Average (ARIMA) and Prophet. Using real-world daily data from the analyzed company, related with a three-year period and seven types of products (e.g., sandals), we execute a realistic rolling window evaluation scheme that considers from 1 to $H = 7$ daily ahead predictions (up to one week) and several training and testing evaluations.

2 Related Work

TSF is widely adopted in several application domains (e.g., Finance, Production, Sales). Due to its importance, there is a wide range of methods that can perform TSF tasks. While proposed in the 1970s, the AutoRegressive Integrated Moving Average (ARIMA) methodology [1] is still a popular approach, including its Seasonal ARIMA (SARIMA) variant, which is capable of modeling trend and seasonal effect. Another popular method is Prophet, which was introduced by the Facebook company in 2018 [6]. Prophet is based on a additive regression approach that is capable of modeling trends, seasonal patterns and even outliers associated with weekends or specific events (e.g., holidays) [4]. In recent years, there has also been a growing interest in using deep learning methods, including LSTM recurrent networks, to perform TSF tasks [5,8,13].

In terms of the sales application domain, several studies have adopted TSF approaches. For instance, ARIMA was used to perform retail forecasting projections related with five categories of women's footwear [10]. Also, the SARIMA was adopted in [7] to forecast the monthly number of car sales in South Africa. In [3], ARIMA and Support Vector Machines were explored to predict the foot traffic of a retail store. As for the LSTM model, it was proposed in [13] to forecast sales of 66 different products over 45 weeks. In another study, the LSTM neural architecture obtained the best predictive results when predicting pharmaceutical sales, outperforming the Prophet model [8].

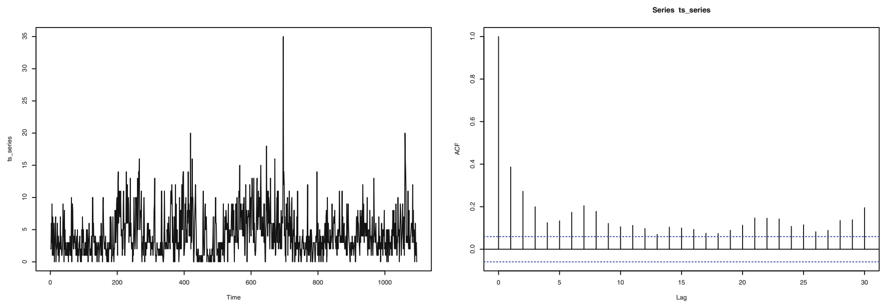
Recently, we have compared three distinct LSTM architectures to perform multi-step ahead predictions of a different application domain: movements of industrial workers [5]. The best results were obtained by a Seq2Seq LSTM, which assumes a encoder-decoder architecture. Following on these good results, in this paper we explore the Seq2Seq LSTM architecture for footwear sales prediction, comparing it with two other TSF method (ARIMA and Prophet).

3 Materials and Methods

3.1 Footwear Sales Data

The data was extracted from a database concerning the business management and sales of fashion products made available at the *Overcube* online platform over a three-year time span, from 2019 to 2021. The raw dataset was comprised of thirty-one product features (e.g., identification, size). The online platform also contains sales data related with 78 countries (e.g., Germany, Spain, Portugal).

In this work, we analyze seven footwear product categories (accessories, ankle boots, boots, sandals, shoe care, shoes and sneakers) that were sold at Portugal. The sales were aggregated into daily values. Then, the empty entries (corresponding to no sales) were replaced by zero values, resulting in a total of 1,095 daily records for all seven time series. Figure 1 shows an example of the accessories sales time series (left graph) and its respective autocorrelation values (right plot).



(a) Accessories daily sales (b) Autocorrelation of Accessories time series

Fig. 1. Daily accessories sales time series (right, x-axis denotes the time period, y-axis the number of daily sales) and its autocorrelation values (left, x-axis denotes the time lags and y-axis the autocorrelations).

3.2 Modeling

An autoregressive time series model predicts a value for current time t based on past observations $\hat{y}_t = f(y_{t-k_1}, ..., y_{t-k_I})$, where f is the forecasting function and the k_i values denote the past time lags (assuming a total of I inputs). Daily series often present a weekly seasonal period ($K = 7$), which is the case of the analyzed footwear sales series. For example, the right of Fig. 1 shows higher autocorrelation values for the multiples of $K = 7$, confirming a weekly seasonal pattern. Therefore, the three time series methods (LSTM, ARIMA and Prophet) are set to built weekly forecasts, by computing from $h = 1$ up to $h = H = 7$ multi-step ahead daily predictions, where h denotes the ahead time in which the forecast is executed and $H = 7$ denote the maximum horizon value (up to one week).

The LSTM is a popular deep learning model to process temporal data. In effect, the LSTM is a special type of Recurrent Neural Network (RNN) that resolves the "short-term memory" problem by using a mechanism of gates that regulate the flow of information [2]. This type of RNN is capable of learning the order dependence in a sequential prediction problem, including time series. The

LSTM architecture is composed of a set of cells, where each cell includes three control gates: the "forget gate" that defines whether the information is relevant (1) or not (0); the "memory gate" that decides the new data that should be stored and modified in the cell; and the "output gate" that controls what is produced in each cell [11].

There are several LSTM variants. In this work, we explore a recently proposed Seq2Seq LSTM architecture that outperformed other two LSTM variants (standard LSTM and a stacked LSTM with two hidden layers) when predicting the shoulder angular movements of industrial workers [5]. Since the model is capable of memorizing temporal sequences, the Seq2Seq LSTM is only fed with one time lag y_{t-1}. The Seq2Seq LSTM model assumes a sequence to sequence architecture (encoder-decoder) that includes one model for reading and encoding the input sequence and a second model for decoding and performing predictions (Fig. 2). The Seq2Seq LSTM model includes two LSTM layers (each one with $L = 100$ cells), one repeat vector layer with $H = 7$ nodes, to repeat the incoming inputs for up to H times, and one time distributed layer, to process the output from the LSTM hidden layer and generate the H sequential output values. Thus, the Seq2Seq LSTM model is capable of performing from $H = 1$ to $H = 7$ multi-step ahead predictions when fed with the last known series value. The deep learning model was trained with the Adam optimizer, using the Mean Squared Error (MSE) loss function and assumed the ReLU activation function, since the sales data can not contain negative values.

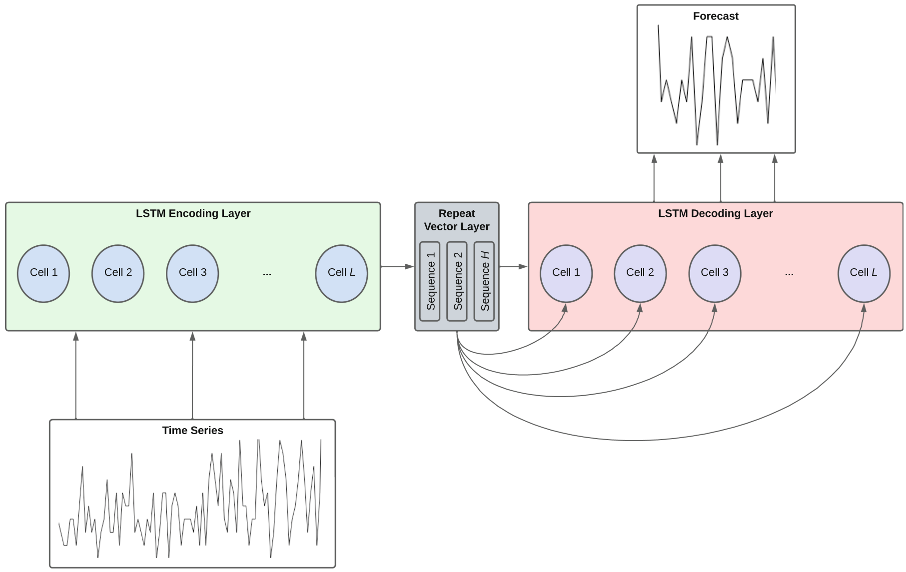


Fig. 2. Architecture of the Seq2Seq LSTM model.

As a baseline comparison, the proposed Seq2Seq LSTM model is compared with two popular seasonal TSF methods [6]: SARIMA and Prophet. Both methods were set using a weekly time period ($K = 7$).

3.3 Evaluation

The TSF methods are evaluated using a robust rolling window validation [9,12], which simulates a real usage of a forecasting model through time, with several training and test updates (Fig. 3). The training set assumes a fixed window length with W examples. In the first iteration ($u = 1$), the model is adjusted to a training window with the W oldest values, and then predicts up to T test ahead predictions (in this paper, $T = H = 7$). Next, in the second iteration ($u = 2$), the training data is updated with S newer examples, allowing to fit a TSF model with W values and perform newer T predictions, and so on. In total, this produces $U = \frac{D-(W+T)}{S}$ model updates, where D is the data length (number of time series examples). After consulting with the company experts, we opted to use the realistic values of $W = 881$ and $H = 7$ with a $S = 7$, thus resulting in $U = 28$ model fitting and testing updates for each TSF method (LSTM, SARIMA and Prophet).

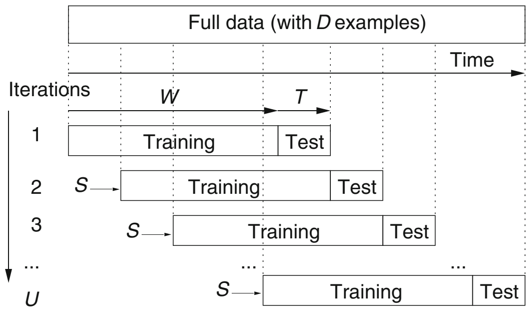


Fig. 3. Schematic of the rolling window procedure.

To measure the quality of the forecasts, we adopt the Normalized Mean Absolute Error (NMAE). The NMAE measure normalizes the popular Mean Absolute Error (MAE) by the output target range on the test set, thus resulting in a scale-independent percentage that is easier to understand and is represented by the following formula [9]:

$$NMAE = \frac{MAE}{(y_{max} - y_{min})} \tag{1}$$

where the y_{max} and the y_{min} represent the highest and the lowest values of the target series. Note that the lower NMAE values, the more accurate is the model (the perfect value is 0%).

For each footwear product, the adopted rolling window scheme produces $U = 28$ sets of predictions, each set with H =7 multi-step ahead forecasts (from $h = 1$ to $h = 7$). Following the procedure used in [3], we compute vertical and horizontal forecasting errors. The former assumes the NMAE value for a fixed h value, thus computed with $U = 28$ observations. The latter, termed here as Multi-Step Ahead Aggregation (MSAA), works by first computing the NMAE value using all multi-step ahead forecasts (7 values, from $h = 1$ to $h = 7$ for a targeted week). Then, the distinct U =28 NMAE values (one for each tested week) are aggregated by computing the median values, which is less sensitive to outliers when compared with the average.

4 Results

All computational experiments were conducted using code written in the Python programming language. The Seq2Seq LSTM neural network implementation is based on the `tensorflow` API structure[1]. Each rolling window iteration assumes an initial Seq2Seq LSTM network with random generated connection weights. Then, the Adam optimizer is run, assuming an early stopping procedure (10% of the most recent training data is used as the validation set) and a maximum of 100 learning epochs. Regarding the SARIMA, we adopted the `auto.arima` function of the `pmdarima` python module[2], which executes an automatic SARIMA model identification and fit for each u-th iteration of the rolling window scheme. As for the Prophet, the `prophet` python package[3] was adopted. Similarly to the previous methods, the Prophet model is refit using the training data available in each iteration of the rolling window procedure.

Table 1 summarizes the obtained predictive results obtained by the ARIMA, Prophet and LSTM models. In terms of the vertical NMAE values (for a fixed h), the proposed Seq2Seq LSTM model produces competitive results. In effect, for almost all $h \in \{1, 2, ..., 7\}$ ahead ranges and footwear products, the LSTM model obtains the lowest vertical NMAE values. In 49 NMAE comparisons, there are only four cases in which Prophet produces similar (sneakers and $h = 7$) or better results (e.g., ankle boots and $h = 3$). Moreover, the average vertical NMAE values, considering all products (last row of Table 1) favors the deep learning method when compared with the two baseline methods (SARIMA and Prophet). Turning to the horizontal multi-step ahead forecasting results (column **MSAA**), the computed median NMAE values also position the Seq2Seq LSTM model at the first place, producing the lowest values for all seven footwear products, ranging from 4.8% (accessories) to 10.6% (sneakers). On average (considering all products), the NMAE median value is 7.07%, which is around 2% points better when compared with SARIMA and Prophet.

For demonstration purposes, Fig. 4 shows the last rolling window iteration weekly ahead forecasts ($U = 28$, h from 1 to 7) for the sandals product sales.

[1] https://www.tensorflow.org/.
[2] https://alkaline-ml.com/pmdarima/0.9.0/index.html.
[3] https://github.com/facebook/prophet.

Table 1. Predictive results (NMAE values for a fixed h, in %; median NMAE values for MSAA, in %; the best values are highlighted by using a **boldface** font).

Model	Product	$h = 1$	$h = 2$	$h = 3$	$h = 4$	$h = 5$	$h = 6$	$h = 7$	MSAA
SARIMA	Accessories	5.94	4.85	6.19	6.04	6.35	6.09	4.11	6.32
	Ankle Boots	5.71	3.78	3.87	6.13	4.62	4.04	4.35	6.33
	Boots	5.23	3.36	5.50	5.21	4.71	6.85	4.27	7.34
	Sandals	9.59	9.27	10.75	12.22	12.68	13.24	13.49	10.75
	Shoe Care	9.94	6.76	7.77	10.36	11.03	8.52	7.94	10.42
	Shoes	10.70	7.44	10.67	9.56	8.41	8.85	8.34	8.95
	Sneakers	10.54	12.04	11.46	10.09	16.45	12.98	14.04	13.71
Average		8.24	6.79	8.03	8.52	9.18	8.65	8.08	9.12
Prophet	Accessories	6.88	5.84	5.93	6.88	7.29	6.64	5.49	7.09
	Ankle Boots	5.79	3.89	**3.76**	5.87	4.78	3.39	4.31	5.82
	Boots	5.46	3.87	6.05	5.61	5.52	6.43	4.70	7.45
	Sandals	9.69	10.40	9.41	10.75	10.22	10.26	12.42	8.90
	Shoe Care	9.54	8.09	7.59	11.53	12.23	8.75	**6.82**	10.70
	Shoes	11.14	8.03	10.71	9.61	8.67	**6.39**	8.08	9.31
	Sneakers	11.87	13.24	11.18	11.37	17.35	13.67	**13.90**	14.11
Average		8.62	7.62	7.80	8.80	9.44	7.93	7.96	9.05
Seq2Seq LSTM	Accessories	**4.54**	**3.97**	**4.91**	**5.02**	**4.42**	**4.63**	**3.36**	**4.80**
	Ankle Boots	**4.94**	**3.49**	4.02	**5.37**	**4.31**	**3.10**	**2.60**	**4.75**
	Boots	**3.70**	**2.43**	**3.30**	**2.95**	**2.58**	**4.32**	**2.96**	**5.40**
	Sandals	**9.04**	**8.11**	**7.63**	**8.72**	**9.92**	**9.05**	**10.53**	**7.79**
	Shoe Care	**8.46**	**6.06**	**6.80**	**9.66**	**9.93**	**7.67**	6.92	**9.18**
	Shoes	**8.34**	**5.61**	**8.08**	**7.33**	**6.44**	7.08	**6.49**	**6.95**
	Sneakers	**9.61**	**6.55**	**10.20**	**6.88**	**11.40**	**12.10**	**13.90**	**10.64**
Average		**6.95**	**5.17**	**6.42**	**6.56**	**7.00**	**6.85**	**6.68**	**7.07**

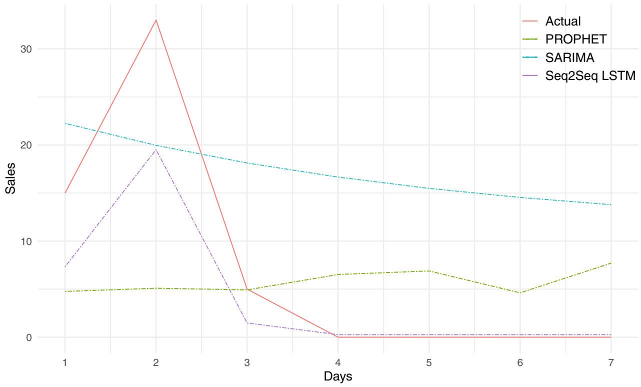


Fig. 4. Last rolling window weekly forecasts for the three TSF methods (x-axis denotes the horizon, y-axis the number of daily sales) for the sandals product.

In this example, the best forecasts are provided by the Seq2Seq LSTM model, which correctly identify a peak for $h = 2$ and also detect a stagnation "no sale" period from $h = 4$ to $h = 7$.

5 Conclusions

In this paper, we explored a recently proposed Seq2Seq LSTM deep learning method to forecast the number of daily sales of seven products from a Portuguese footwear company. Using real-world data, collected during a three-year time period and a robust rolling windows evaluation scheme, several computational experiments were conducted, comparing the proposed deep learning model with two popular TSF methods (ARIMA and Prophet). Overall, when considering both a fixed ahead forecast (vertical analysis) and multi-step ahead weekly forecasts (horizontal analysis), competitive results were obtained by the proposed Seq2Seq LSTM method. In particular, interestingly low multi-step ahead NMAE values were achieved by the deep learning method, ranging from 5% (accessories sales) to 11% (sneakers sales).

The obtained results were shown to the footwear company, which found them interesting and valuable to support the management of its footwear stocks. Indeed, in future work, we intend to deploy the proposed Seq2Seq LSTM model into a friendly decision support system, aiming to provide forecasting insights for the inventory management system currently adopted by the company. We also plan to adapt the forecasting methods to other time scales (e.g., monthly), explore multivariate forecasting methods and explore other recent forecasting methods (e.g., Temporal Convolutional Network).

Acknowledgments. This work was financed by the project "GreenShoes 4.0 - Calçado, Marroquinaria e Tecnologias Avançadas de Materiais, Equipamentos e Software" (N° POCI-01-0247-FEDER-046082), supported by COMPETE 2020, under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF).

References

1. Box, G., Jenkins, G.: Time Series Analysis: Forecasting and Control. Holden Day, San Francisco (1976)
2. Chopra, S., Meindl, P.: Supply chain management. strategy, planning & operation. In: Boersch, C., Elschen, R. (eds.) Das Summa Summarum Des Management, pp. 265–275. Springer, Cham (2007). https://doi.org/10.1007/978-3-8349-9320-5_22
3. Cortez, P., Matos, L.M., Pereira, P.J., Santos, N., Duque, D.: Forecasting store foot traffic using facial recognition, time series and support vector machines. In: Graña, M., López-Guede, J.M., Etxaniz, O., Herrero, Á., Quintián, H., Corchado, E. (eds.) SOCO/CISIS/ICEUTE -2016. AISC, vol. 527, pp. 267–276. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-47364-2_26
4. Ensafi, Y., Amin, S.H., Zhang, G., Shah, B.: Time-series forecasting of seasonal items sales using machine learning-a comparative analysis. Int. J. Inf. Manag. Data Insights **2**(1):100058 (2022)

5. Fernandes, C., et al.: A deep learning approach to prevent problematic movements of industrial workers based on inertial sensors. In International Joint Conference on Neural Networks, IJCNN 2022, Padua, Italy, 18–23 July 2022. IEEE (2022)
6. Hyndman, R.J., Athanasopoulos, G.: Forecasting: Principles and Practice, 3rd edn. O Texts (2021)
7. Makatjane, K., Moroke, N.: Comparative study of holt-winters triple exponential smoothing and seasonal arima: forecasting short term seasonal car sales in south africa. Risk Gov. Control Financ. Markets Institutions **6** (2016)
8. Meng, J., Yang, X., Yang, C., Liu, Y.: Comparative analysis of prophet and LSTM model in drug sales forecasting. **1910** (2021). IOP Publishing
9. Oliveira, N., Cortez, P., Areal, N.: The impact of microblogging data for stock market prediction: Using twitter to predict returns, volatility, trading volume and survey sentiment indices. Expert Syst. Appl. **73**, 125–144 (2017)
10. Ramos, P., Santos, N., Rebelo, R.: Performance of state space and ARIMa models for consumer retail sales forecasting. Rob. Comput.-Integrat. Manuf. **34**, 151–163 (2015)
11. Siami-Namini, S., Tavakoli, N., Namin, A.S.: A comparison of ARIMA and LSTM in forecasting time series. In: Arif Wani, M., Kantardzic, M.M., Mouchaweh, M.S., Gama, J., Lughofer, E. (eds.) 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018, Orlando, FL, USA, 17–20 December 2018, pp. 1394–1401. IEEE (2018)
12. Tashman, L.J.: Out-of-sample tests of forecasting accuracy: an analysis and review. Int. Forecast. J. **16**(4), 437–450 (2000)
13. Yu, Q., Wang, K., Strandhagen, J.O., Wang, Y.: Application of long short-term memory neural network to sales forecasting in retail—a case study. In: Wang, K., Wang, Y., Strandhagen, J.O., Yu, T. (eds.) IWAMA 2017. LNEE, vol. 451, pp. 11–17. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-5768-7_2