J. Appl. Res. Ind. Eng. Vol. 7, No. 2 (2020) 177–189

Journal of Applied Research on Industrial Engineering
www.journal-aprie.com

Improving Demand Forecasting with LSTM by Taking into Account the Seasonality of Data

Hossein Abbasimehr**, Mohammad Khodizadeh Nahari
Faculty of Information Technology and Computer Engineering, Azarbaijan Shahid Madani University, Tabriz, Iran.

P A P E R   I N F O

Chronicle:
Received: 24 January 2020
Revised: 21 March 2020
Accepted: 26 May 2020

Keywords:
LSTM.
Time Series Forecasting.
Demand Prediction.

A B S T R A C T

Demand forecasting is a vital task for firms to manage the optimum quantity of raw material and products. The demand forecasting task can be formulated as a time series forecasting problem by measuring historical demand data at equal intervals. Demand time series usually exhibit a seasonal pattern. The principle idea of this study is to propose a method that predicts the demand for every different season using a specialized forecaster. In this study, we test our proposal using the Long Short-Term Memory (LSTM) which is a deep learning technique for time series forecasting. Specifically, the proposed method instead of learning an LSTM model using the whole demand data builds a specialized LSTM model corresponding to each season. The proposed method is evaluated using different topologies of the LSTM model. The results of experiments indicated that the proposed method outperforms the regular method considering the performance measures. The proposed method can be used in other domains for demand forecasting.

1. Introduction

Strict competition among businesses in any area has made it difficult for them to accurately forecast the customers' demands using traditional demand forecasting approaches [1]. Therefore, companies are moving toward using advanced deep learning techniques to forecast their customer demand. In general, customer demand is represented as a sequential data which is considered as sequences of customer demands over time. Therefore, the demand forecasting task can be articulated as a time series forecasting problem [2]. Time series prediction has been applied in various areas of applications such as electricity load forecasting [3], tourism demand forecasting [4], passenger demand prediction [5], ATM cash demand forecasting [6], forecasting of petroleum production [7], and etc.

Artificial Neural Networks (ANNs) have several valuable characteristics such as universal approximation, being data-driven, and the capability to better model nonlinear patterns in data [8]. A distinct category of ANNs is Recurrent Neural Networks (RNNs) that differ from feed forward neural networks. Contrasting to ANNs, in RNNs, the connections between nodes create a cycle that allows signals to move in different directions [9]. RNNs are better suited for modeling sequence data as they provide a short-term memory by storing the activations from each time step [9]. The major drawback of the RNN model is the vanishing and exploding gradient problem which makes it hard to train [9].

* Corresponding author
E-mail address: abbasimehr@azaruniv.ac.ir
DOI: 10.22105/jarie.2020.232594.1168

The solution to tackle this weakness is using gated architectures, like Long Short-Term Memory (LSTM) [10] which can exploit longer-range timing information [11].

In this study, we aim to forecast the demand for a strategic product of a furniture manufacturer company. The investigation of the demand series used in this study indicated that it contains seasonality patterns. For example, the demand pattern for the first 6-month period of each year differs from the demand pattern of the second 6-month period of the year. Therefore, we propose a method that builds a specialized LSTM model corresponding to each 6-month period. For each period, an LSTM model is trained using training instances whose target values belong to that period. In this study, different topologies of the LSTM network were investigated to assess the proposed approach.

The rest of this study is organized as follows. Section 2 gives a literature review on time series forecasting and demand prediction using LSTM also, this section describes the structure of the LSTM network. In Section 3, we describe the proposed method. Section 4 presents the experimental results and compares the utilized models. In Section 5, we conclude the paper and suggest future works.

2. Literature Review

In this study, we formulate demand forecasting as a time series forecasting problem. Therefore, we provide the related literature in this context. Generally, time series forecasting techniques are classified into three categories including statistical [5], AI-based methods such as ANNs and hybrid techniques [12].

From statistical techniques, the Auto-Regressive Integrated Moving Average (ARIMA) is a widely-used method [12-14]. Statistical methods can model linear patterns existed in time series, but time series contain nonlinear patterns that can suitably be captured using nonlinear models such as ANNs [8]. The popularity of ANNs is attributed to their non-parametric and data-driven nature. The feed forward Multi-Layer Perceptron (MLPs) are the originally-used ANN architectures that have been applied in time series forecasting. However, MLPs are memory-free models and do not consider the temporal dependencies among data [15]. To reflect dependencies among data, new ANN architectures called RNNs have been developed [9]. Although RNNs are suitable for time series forecasting, they suffer from the vanishing and exploding gradient problem which makes them hard to train [9, 16]. To overcome this problem an extended variant of RNNs called LSTM [11, 17] has been introduced that has gained popularity due to its great performance in time series forecasting. Also, various hybrid models have been proposed in the literature to improve forecasting accuracy [12]. In this study, we select LSTM to evaluate the proposed method, so in the following subsection, we present some researches that utilized LSTM for demand forecasting.

2.1. Demand Prediction by LSTM

Prior researches (e.g. [6, 18, 19]) has exhibited the great potential of LSTM network for time series forecasting. The popularity of LSTM is due to its capability in taking into account the long-term dependency among data [11, 17]. As mentioned earlier, demand prediction is considered as a special type of time series forecasting task. In *Table 1*, we provide some recent studies that utilized LSTM for demand forecasting. For each study, the contribution, domain of study, and the utilized dataset are illustrated in *Table 1*.

Table 1. *Some of recent studies in the context of demand prediction by LSTM.*

Study	Contribution	Domain- Dataset
[20]	Proposing a novel deep learning approach based on the fusion of LSTM and convolutional neural networks to forecast short-term passenger demand.	On-demand ride services- dataset from DiDi Chuxing.
[21]	Devising an LSTM-based model for airline demand forecasting.	Airline industry- The dataset contains the information of route from Dalian to Jinan.
[22]	Proposing a hybrid approach based on Empirical Mode Decomposition (EMD) and LSTM to predict electricity demand. Firstly, the load time series is decomposed using EMD, and then each of extracted subseries is predicted using LSTM and finally the predictions are aggregated to obtain the predicted demand.	Electricity industry -Electricity demand of the UT Chandigarh collected in 15 min interval.
[23]	Employing LSTM for prediction of electricity demand data. Genetic algorithm is used to optimize the hyper parameters. The proposed method outperformed the benchmarking models.	Electricity industry- Electricity consumption data collected in 30 minutes interval.
[24]	Proposing a deep learning-based framework using LSTM for electricity demand forecasting.	Electricity industry- Electricity demand of the UT Chandigarh collected in 15 min interval.
[4]	Proposing a deep learning architecture based on LSTM and attention mechanism to forecast monthly tourist arrival.	Tourism industry-tourist arrival data of Macau (a region in China).
[25]	Proposing a hybrid deep learning-based approach using LSTM networks and genetic algorithm for hours-ahead gas demand prediction.	Natural gas industry- data is taken from OpenEI.
[6]	Proposing a framework that initially identifies similar groups of time series using clustering algorithms and then use LSTM to build forecasting models.	Banking, ATM cash demand forecasting- CIF2016, NN5.
[18]	Proposing an optimized LSTM model for demand series prediction. The proposed method outperformed the state-of-the-art methods in time series prediction.	Furniture industry - demand series of a furniture company.
[26]	Developing an LSTM-based ensemble learning method for ultra-short-term power demand forecasting. LSTM is selected as the base learner of the utilized ensemble algorithm.	Electricity industry- The dataset contains half-hourly electricity load data during 2013 in Australia.
[27]	Proposing a method based on bidirectional LSTM to predict tourism demand. The method incorporates a Bayesian Optimization model to choose the best hyper parameters of bidirectional LSTM.	Tourism industry- dataset consists of the number of tourist arrivals from 5 source countries to Singapore.
[28]	Proposing a hybrid method of LSTM and random forest to forecast the demand for products of a multi-channel retailer.	Retailing industry-data is taken from a food retailer.
[29]	Adopting LSTM for power demand forecasting problem. The result of this study indicated that LSTM outperforms the linear regression.	Electricity industry - hourly power consumption data of fifteen apartments in Korea.
[30]	Developing a hybrid method based on Seasonal ARIMA and LSTM to predict tourist arrival. The proposed hybrid method performs the best among the utilized methods in this study.	Tourism industry- the dataset contains information of daily tourist arrival to Macau (a region in China).

As the literature review indicates, LSTM is the popular method in the recent years. Therefore, we use LSTM to evaluate the proposed approach for handling seasonality problem in the utilized demand series.

2.2. LSTM

The LSTM network is an extension of RNN that can memorize long-range time dependency information [31]. As mentioned before, the major weakness of RNNs is the problem of vanishing gradient. LSTM solves this problem by introducing memory cells and gating mechanisms [10]. Each memory cell $c(t)$ preserves the temporal states of the network and consists of three gates, including the input gate $i(t)$, the forget gate $f(t)$ and the output gate $o(t)$ that controls the flow of information. The architecture of LSTM is portrayed in *Fig. 1*. The operations of gates are as follows [32].

- The input gate decides the information that is to be added to the cell state.
- The forget gates determine the information that is to be discarded from the cell state.
- The output gate determines the outgoing information of the cell state.

Considering the following notations, the *Eqs. (1)-(6)* describe the modeling process of LSTM at every timestep t [32].

$x(t)$: The input vector at time step t .

$h(t)$: The output value at time point t .

$c(t)$ and $\tilde{c}(t)$: Cell states and candidate values at time step t .

$b = \{b_i, b_f, b_c, b_o, b_{\tilde{c}}\}$ are biases of $i, f, c, o, \tilde{c}$.

$\vec{W}_1 = \{w_i, w_f, w_c, w_{\tilde{c}}, w_o\}$ are weight matrixes.

$\vec{W}_2 = \{w_{hi}, w_{hf}, w_{hc}, w_{ho}, w_{h\tilde{c}}\}$ are the recurrent weights.

σ and $tanh$ denote the recurrent activation functions, and $\times$ represents the element-wise multiplication.

In the first step, the forget layer decides which information to be discarded from the previous cell states $c(t-1)$. The forget layer utilizes the current input $x(t)$, the output of the memory cells at the previous steps $h(t-1)$ and the bias of the forget gate (b_f) to calculate the values $f(t)$ by applying a sigmoid activation function (σ). Sigmoid function scales all the computed value into $[0,1]$.

$$f(t) = \sigma(w_f x(t) + w_{hf} h(t-1) + b_f). \tag{1}$$

In the second step, the LSTM layer determines which information will be stored in the cell state $C(t)$. This step involves two parts: At the first step, new candidate values $\tilde{c}(t)$ are calculated using a hyperbolic tangent function (tanh) as the activation function. In the second step, the activation values $i(t)$ of the input gates are computed as follows:

$$\tilde{c}(t) = \tanh(w_{\tilde{c}} x(t) + w_{h\tilde{c}} h(t-1) + b_{\tilde{c}}), \tag{2}$$

$$i(t) = \sigma(w_i x(t) + w_{hi} h(t-1) + b_i). \tag{3}$$

In the third step, the new cell state value $c(t)$ is computed using new information.

$$c(t) = f(t) \times c(t-1) + i(t) \times \tilde{c}(t). \tag{4}$$

In the last step, the output of the memory cell at time step t ($h(t)$) is computed using the following equations:

$$o(t) = \sigma(w_o x(t) + w_{ho} h(t-1) + b_o), \tag{5}$$

$$h(t) = o(t) \times \tanh(c(t)). \tag{6}$$

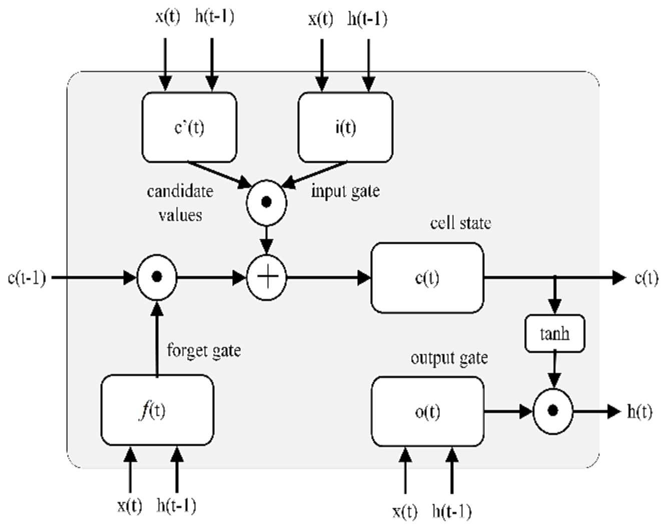

Fig. 1. *The structure of LSTM [32].*

3. Our Proposal

In this section, we describe the steps of the proposed method. Firstly, the data preprocessing steps consisting of the data normalization and instance creation are accomplished. In data normalization step, the given time series is normalized using the min-max normalization [33]. To train an LSTM network, the data must be transformed into a set of instances containing both the inputs and the outputs. In the instance creation step, using a lag L, subsequences of length $L+1$ are derived from the series. The first L points of a sequence are considered as the input and the last point, $L+1$ is labeled as the target. The procedure of construction of instances from a time series with length n is depicted in *Fig. 2.*

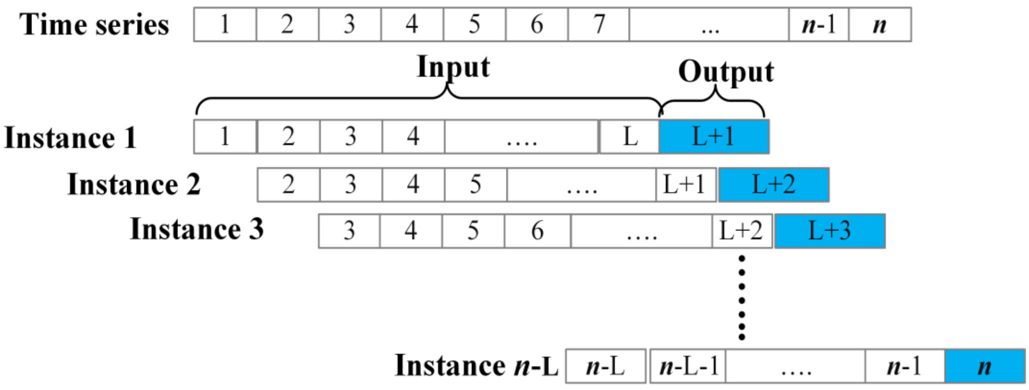

Fig. 2. *Creation of instances for LSTM network.*

3.1. Proposed Model

Usually, when machine learning techniques are applied to forecast demand series, the data of the whole series are utilized for the creation of instances. However, a given demand series often contain different seasonal patterns, and thereby relying on whole data may lead to a forecasting model with poor accuracy. Most of the prior research in time series forecasting and especially in demand forecasting do not consider this issue. In this study, the principle idea is to propose a method that predicts the demand for every different season of demand series using a specialized LSTM.

Investigation of the demand time series used in this study indicated that the pattern of the first 6-month period is not similar to the pattern of the second 6-month period. To improve the forecasting performance, our proposed method trains a specialized LSTM for forecasting each 6-month period of a year. As indicated in [34], including whole time series data for forecasting may decrease the accuracy of a forecasting technique. The LSTM model used to forecast a given 6-month period only consists of instances whose targets belong to that period. The procedure of the proposed approach is described below:

Given a time series of length n , perform the following steps:

- Select a lag L.
- Create instances from the time series using L.
- Select the instances whose target belonging to the first 6-month period and train the LSTM_a using the selected instances.
- Select the instances whose target belonging to the second 6-month period and train the LSTM_b using the selected instances.
- To forecast the demand for a month belonging to the first 6-month period use LSTM_a.
- To forecast the demand for a month belonging to the second 6-month period use LSTM_b.

To describe the proposed method, suppose that we have a monthly time series with a length of 18. Also, assume that the selected lag is 6 ($L=6$). Therefore, the first step is to create instances. The created instances are illustrated in *Table 2.* According to our proposal, the specialized LSTM for the first period (LSTM_a) will be trained using the instances (1)-(6) and the specialized LSTM for the second period (LSTM_b) will be learned using the instances (7)-(12).

Table 2. *Instances for time series with a length of 18 and lag=6.*

Ins #	Input	Output
1	$\{M_1, M_2, M_3, M_4, M_5, M_6\}$	M_7
2	$\{M_2, M_3, M_4, M_5, M_6, M_7\}$	M_8
3	$\{M_3, M_4, M_5, M_6, M_7, M_8\}$	M_9
4	$\{M_4, M_5, M_6, M_7, M_8, M_9\}$	M_{10}
5	$\{M_5, M_6, M_7, M_8, M_9, M_{10}\}$	M_{11}
6	$\{M_6, M_7, M_8, M_9, M_{10}, M_{11}\}$	M_{12}
7	$\{M_7, M_8, M_9, M_{10}, M_{11}, M_{12}\}$	M_{13}
8	$\{M_8, M_9, M_{10}, M_{11}, M_{12}, M_{13}\}$	M_{14}
9	$\{M_9, M_{10}, M_{11}, M_{12}, M_{13}, M_{14}\}$	M_{15}
10	$\{M_{10}, M_{11}, M_{12}, M_{13}, M_{14}, M_{15}\}$	M_{16}
11	$\{M_{11}, M_{12}, M_{13}, M_{14}, M_{15}, M_{16}\}$	M_{17}
12	$\{M_{12}, M_{13}, M_{14}, M_{15}, M_{16}, M_{17}\}$	M_{18}

4. Empirical Study

To assess the proposed method, we carry out a case study using real demand data of a furniture company.

4.1. Data Description

This data is the monthly sale quantity for a specific product from 2007-2017 of a furniture company. In fact, the time series of sale quantity consists of 132 months (time steps) (as seen in *Fig.3*). The aim of the company is to predict the demand for its products accurately to effectively plan to procure raw materials. The prediction of future demand based on this data allows the company to improve its strategies for resource management. As the analyzed data is for a long period, the resulted demand prediction can be reliable.

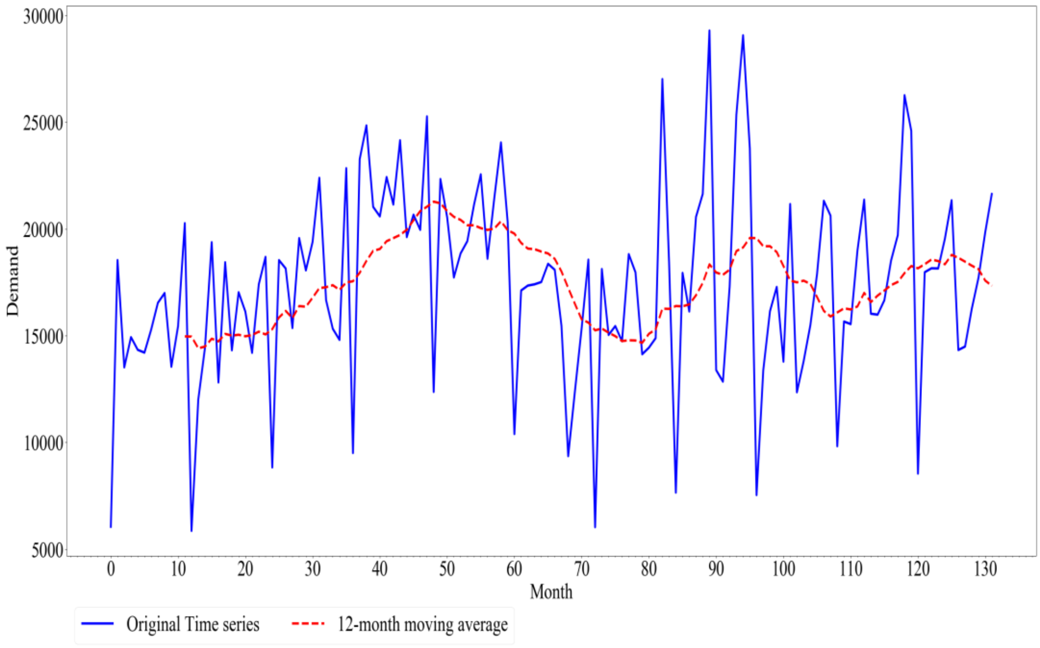

Fig. 3. *The demand time series and its trend plot.*

4.2. Data Preprocessing

The original demand series consists of monthly demand data for 11 years (132 months). We selected three demand series, *D1*, *D2*, and *D3* to conduct the experiments. $D1:\{M_1, M_2 \ldots M_{132}\}$ is the original demand series; $D2:\{M_1, M_2 \ldots M_{120}\}$ and $D3:\{M_1, M_2 \ldots M_{108}\}$ are the demand data for 10 and 9 years. For each demand series, the last $h=12$ points of the series are selected as the test set and the remaining data points are used for model training.

Each series is normalized using the min-max normalization [33]. Also, to create instances from each series, we set the $Lag=12$. As mentioned before, we aim to measure the performance of the LSTM network when utilized with our proposal.

4.3. Model Building and Evaluation

To comprehensively evaluate the performance of our proposal, four different LSTM models have been employed. The topology of each model is portrayed in *Table 3*. These LSTM models are based on the two LSTM architecture illustrated in *Figs. (4)-(5)*. It should be noted that the LSTM models are implemented using Keras [31] in Python.

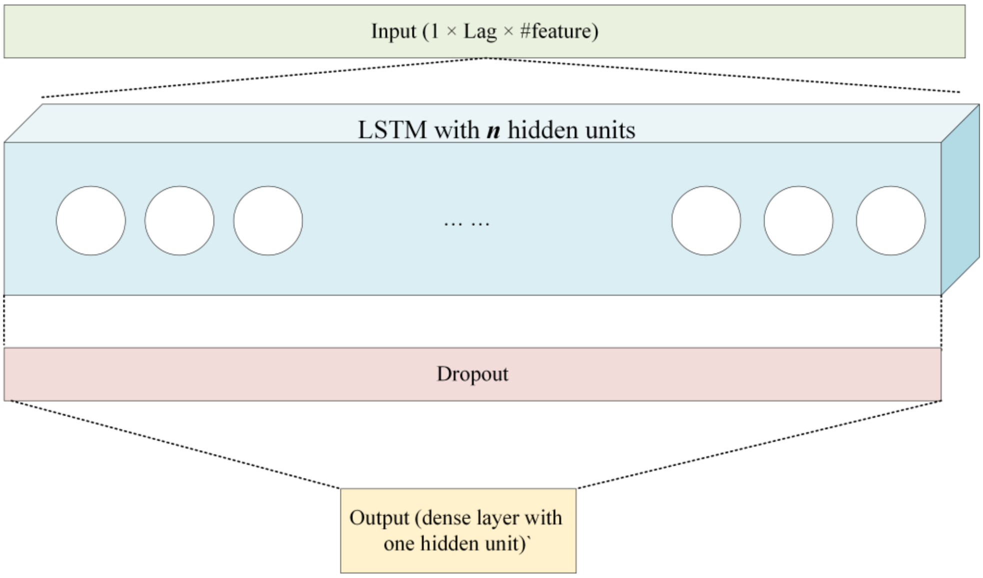

Fig. 4. *The single layer LSTM architecture (This architecture is used in LSTM_1 and LSTM_3).*

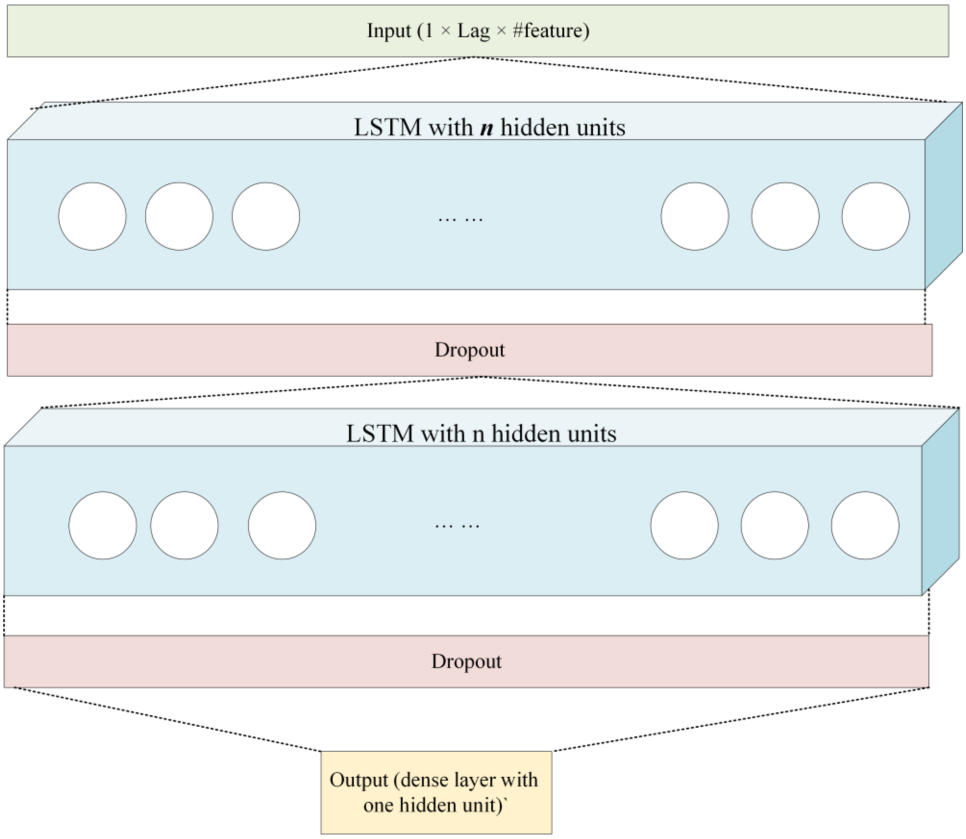

Fig. 5. *The 2-layer LSTM architecture (This architecture is used in LSTM_2 and LSTM_4).*

LSTM networks are prone to overfitting, so to prevent them from overfitting and improving their generalization to new data, we use early stopping [31]. To exploit early stopping, we set the epoch limit to 500. Using early stopping, model training will be stopped considering stopping criteria. In this study, for all the models, we use the Adam optimization algorithm [31]. Furthermore, we use the dropout mechanism to avoid overfitting.

To evaluate the accuracy of the models, we use Symmetric Mean Absolute Percentage (SMAPE) and Root Mean Square Error (RMSE) measures. SMAPE is defined by *Eq. (7)* [18].

$$\mathrm{SMAPE} = \frac{1}{n}\sum_{t=1}^{n}\frac{|\hat{y}_t - y_t|}{\frac{|\hat{y}_t| + |y_t|}{2}}. \quad (7)$$

RMSE is denoted as *Eq. (8).*

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}. \quad (8)$$

In both *Eqs. (7)-(8),* y_t and $\hat{y}_t$ are the actual and predicted values at t , respectively [18].

Table 3. *The topology of the LSTM models.*

Model	**Topology**
LSTM_1	*LSTM layer with 32 units.* *Dropout.* *Output (dense layer with 1 neuron).*
LSTM_2	*LSTM layer with 32 units.* *Dropout.* *LSTM layer with 32 units.* *Dropout.* *Output (dense layer with 1 neuron).*
LSTM_3	*LSTM layer with 64 units.* *Dropout.* *Output (dense layer with 1 neuron).*
LSTM_4	*LSTM layer with 64 units.* *Dropout.* *LSTM layer with 64 units.* *Dropout.* *Output (dense layer with 1 neuron).*

The results of applying all models on *D1*, *D2*, and *D3* are illustrated in *Table 4* and *Table 5*. It should be noted that for each dataset, each model was run 10 times and the results were averaged. As indicated in *Table 4*, all LSTM models with our proposal outperform the regular models in terms of SMAPE that demonstrates the usefulness of using specialized learners for demand prediction. Also, the mean SMAPE values for all models differ slightly. This demonstrated the usefulness of using specialized learners for demand prediction. Also, the mean SMAPE values for all models differ slightly. LSTM_4 achieves the lowest mean SMAPE value which reveals that deep LSTM models can better capture patterns existed in time series. Besides, as illustrated in *Table 5*, the mean RMSE for all models with our proposal is lower than the regular models.

Table 4. *The performance of the LSTM models with our proposal and the regular LSTMs in terms of SMAPE. Regular means that the model is trained using the entire instances.*

	Method	D1	D2	D3	Mean
LSTM _1	*Regular*	*11.85*	*14.45*	*20.23*	*15.51*
	Our proposal	*11.29*	*13.22*	*16.81*	*13.77*
LSTM _2	*Regular*	*12.79*	*15.61*	*16.56*	*14.99*
	Our proposal	*11.79*	*13.40*	*16.76*	*13.98*
LSTM _3	*Regular*	*12.52*	*13.34*	*20.23*	*15.36*
	Our proposal	*11.23*	*13.20*	*17.39*	*13.94*
LSTM _4	*Regular*	*12.15*	*14.16*	*19.59*	*15.3*
	Our proposal	*11.87*	*13.74*	*14.95*	*13.52*

Table 5. *The performance of the LSTM models with our proposal and the regular LSTMs in terms of RMSE. Regular means that the model is trained using the entire instances.*

	Method	D1	D2	D3	Mean
LSTM _1	*Regular*	*2429.91*	*3182.99*	*4329.20*	*3314.03*
	Our proposal	*2205.59*	*3530.47*	*3301.38*	*3012.48*
LSTM _2	*Regular*	*2671.65*	*3343.83*	*4146.58*	*3387.35*
	Our proposal	*2309.39*	*3544.80*	*3400.75*	*3084.98*
LSTM _3	*Regular*	*2565.62*	*3102.76*	*4293.95*	*3320.77*
	Our proposal	*2276.77*	*3504.86*	*3388.68*	*3056.77*
LSTM _4	*Regular*	*2757.78*	*3169.40*	*4268.33*	*3398.50*
	Our proposal	*2347.32*	*3743.16*	*3109.89*	*3066.79*

5. Conclusion

Demand time series usually contain a seasonal pattern. Therefore, considering the entire demand data for building forecaster may result in a model with poor performance. Considering this issue, this study proposed a method that takes into the seasonality of data and trains a specialized learner for every season using instances whose output values are from that season. The proposed method validated using the LSTM model. Different LSTM topologies are selected for this purpose. The experimental study carried out using the demand data of a furniture company revealed that the LSTM models with our proposal achieved better performance than the regular methods. As future work, we aim to apply our proposal for time series forecasting in other domains such as forecasting customer behaviors in banking. Furthermore, we aim to tune the hyper parameters of LSTM using a Bayesian optimization method.

References

[1] Kumar, A., Shankar, R., & Aljohani, N. R. (2020). A big data driven framework for demand-driven forecasting with effects of marketing-mix variables. *Industrial marketing management*, *90*, 493-507.

[2] Villegas, M. A., Pedregal, D. J., & Trapero, J. R. (2018). A support vector machine for model selection in demand forecasting applications. *Computers & industrial engineering*, *121*, 1-7. https://doi.org/10.1016/j.cie.2018.04.042.

[3] Johannesen, N. J., Kolhe, M., & Goodwin, M. (2019). Relative evaluation of regression tools for urban area electrical energy demand forecasting. *Journal of cleaner production, 218*, 555-564.
[4] Law, R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: A deep learning approach. *Annals of tourism research, 75*, 410-423
[5] Olaniyi, A. A., Adedotun, K. O., & Samuel, O. A. (2018). Forecasting methods for domestic air passenger demand in Nigeria. *Journal of applied research on industrial engineering, 5*(2), 146-155. 10.22105/jarie.2018.133561.1038
[6] Bandara, K., Bergmeir, C., & Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. *Expert systems with applications, 140*, 112896. https://doi.org/10.1016/j.eswa.2019.112896
[7] Sagheer, A., & Kotb, M. (2019). Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing*, *323*, 203-213.
[8] Panigrahi, S., & Behera, H. S. (2017). A hybrid ETS–ANN model for time series forecasting. *Engineering applications of artificial intelligence*, *66*, 49-59.
[9] Parmezan, A. R. S., Souza, V. M., & Batista, G. E. (2019). Evaluation of statistical and machine learning models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information sciences*, *484*, 302-337.
[10] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.
[11] Wu, Y., Yuan, M., Dong, S., Lin, L., & Liu, Y. (2018). Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing, 275*, 167-179. https://doi.org/10.1016/j.neucom.2017.05.063.
[12] Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied soft computing, 11*(2), 2664-2675.
[13] Murray, P. W., Agard, B., & Barajas, M. A. (2018). Forecast of individual customer's demand from a large and noisy dataset. *Computers & industrial engineering*, *118*, 33-43.
[14] Abbasimehr, H., & Shabani, M. (2020). A new framework for predicting customer behavior in terms of RFM by considering the temporal aspect based on time series techniques. *Journal of ambient intelligence and humanized computing*. 10.1007/s12652-020-02015-w.
[15] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European journal of operational research, 270*(2), 654-669.
[16] Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks, 5*(2), 157-166.
[17] Xin, W. A. N. G., Ji, W. U., Chao, L. I. U., Haiyan, Y. A. N. G., Yanli, D. U., & Wensheng, N. I. U. (2018). Exploring LSTM based recurrent neural network for failure time series prediction. *Journal of beijing university of aeronautics and astronautics*, *44*(4), 772-784.
[18] Abbasimehr, H., Shabani, M., & Yousefi, M. (2020). An optimized model using LSTM network for demand forecasting. *Computers & industrial engineering*. https://doi.org/10.1016/j.cie.2020.106435.
[19] Shankar, S., Ilavarasan, P. V., Punia, S., & Singh, S. P. (2019). Forecasting container throughput with long short-term memory networks. *Industrial management & data systems. 120*(3), 425-441. 10.1108/IMDS-07-2019-0370.
[20] Ke, J., Zheng, H., Yang, H., & Chen, X. M. (2017). Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation research part C: Emerging technologies*, *85*, 591-608.
[21] Pan, B., Yuan, D., Sun, W., Liang, C., & Li, D. (2018, June). A novel LSTM-Based Daily Airline Demand Forecasting Method Using Vertical and Horizontal Time series. *Pacific-Asia conference on knowledge discovery and data mining* (pp. 168-173). Springer, Cham.
[22] Bedi, J., & Toshniwal, D. (2018). Empirical mode decomposition based deep learning for electricity demand forecasting. *IEEE access*, *6*, 49144-49156.
[23] Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*, *11*(7), 1636.
[24] Bedi, J., & Toshniwal, D. (2019). Deep learning framework to forecast electricity demand. *Applied energy*, *238*, 1312-1326.
[25] Su, H., Zio, E., Zhang, J., Xu, M., Li, X., & Zhang, Z. (2019). A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced Deep-RNN model. *Energy*. 178, 585-597. https://doi.org/10.1016/j.energy.2019.04.167
[26] Tan, M., Yuan, S., Li, S., Su, Y., Li, H., & He, F. (2019). Ultra-short-term industrial power demand forecasting using LSTM based hybrid ensemble learning. *IEEE transactions on power systems*, *35*(4), 2937-2948. 10.1109/TPWRS.2019.2963109

[27] Kulshrestha, A., Krishnaswamy, V., & Sharma, M. (2020). Bayesian BILSTM approach for tourism demand forecasting. *Annals of tourism research*, *83*, 102925. https://doi.org/10.1016/j.annals.2020.102925

[28] Punia, S., Nikolopoulos, K., Singh, S. P., Madaan, J. K., & Litsiou, K. (2020). Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. *International journal of production research*, 1-16. 10.1080/00207543.2020.1735666

[29] Bui, V., Kim, J., & Jang, Y. M. (2020, February). Power demand forecasting using long short-term memory neural network based smart grid. *2020 international conference on artificial intelligence in information and communication (ICAIIC)* (pp. 388-391). IEEE.

[30] Wu, D. C. W., Ji, L., He, K., & Tso, K. F. G. (2020). Forecasting tourist daily arrivals with a hybrid Sarima–Lstm approach. *Journal of hospitality & tourism research*. https://doi.org/10.1177/1096348020934046

[31] Chollet, F. (2015). *Keras*. Retrived January 12, 2020 from https://github.com/fchollet/keras.

[32] Graves, A. (2013). Generating sequences with recurrent neural networks. https://arxiv.org/

[33] Jiawei Han, M. K., & Pei, J. (2011). *Data mining: concepts and techniques: concepts and techniques*. Waltham, USA: Elsevier Science.

[34] Martínez, F., Frías, M. P., Pérez-Godoy, M. D., & Rivera, A. J. (2018). Dealing with seasonality by narrowing the training set in time series forecasting with kNN. *Expert systems with applications*, *103*, 38-48.