

Article

Tourism Demand Forecasting Based on an LSTM Network and Its Variants

Shun-Chieh Hsieh 

Department of Land Management and Development, Chang Jung Christian University,
Tainan City 71101, Taiwan; sch@mail.cjcu.edu.tw

Abstract: The need for accurate tourism demand forecasting is widely recognized. The unreliability of traditional methods makes tourism demand forecasting still challenging. Using deep learning approaches, this study aims to adapt Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), and Gated Recurrent Unit networks (GRU), which are straightforward and efficient, to improve Taiwan's tourism demand forecasting. The networks are able to seize the dependence of visitor arrival time series data. The Adam optimization algorithm with adaptive learning rate is used to optimize the basic setup of the models. The results show that the proposed models outperform previous studies undertaken during the Severe Acute Respiratory Syndrome (SARS) events of 2002–2003. This article also examines the effects of the current COVID-19 outbreak to tourist arrivals to Taiwan. The results show that the use of the LSTM network and its variants can perform satisfactorily for tourism demand forecasting.

Keywords: deep learning; long short-term memory; tourism demand; Taiwan; gated recurrent unit



Citation: Hsieh, S.-C. Tourism Demand Forecasting Based on an LSTM Network and Its Variants. *Algorithms* **2021**, *14*, 243. <https://doi.org/10.3390/a14080243>

Academic Editor: Antonio Della Cioppa

Received: 28 July 2021

Accepted: 16 August 2021

Published: 18 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Perishability is one of the most important characteristics of the tourism industry, making the need for accurate tourism demand (TD) forecasting crucial [1]. Governments and organizations always need to accurately estimate the expected TD in order to make valid policy planning, tactical and operational decisions [2,3]. Accurate TD forecasting can effectively boost economic development and employment [4]; hence, the need for accurate TD forecasting is widely recognized [5]. Quantitative TD forecasting techniques can be divided into time series models, econometric approach and artificial intelligence (AI) techniques [6]. However, no single technique outperforms others on all scenarios in terms of accuracy.

Time series models have been very popular for TD forecasting, and their advantage lies in the validity and efficiency of autoregressive integrated moving average (ARIMA) and its variants [7]. Most ARIMA variants are subject to some limitations, such as the assumption of a linear relation between future and past time step values, and the number of observations [8]. Therefore, when solving complex nonlinear problems, the estimates obtained may be inaccurate.

Econometric methods can determine the cause-and-effect relation between TD dependent variables and independent variables [9]. However, most econometric models have several limitations. For instance, the independent variables are either exogenous or endogenous, and are decided in advance before the modelling process [10].

AI techniques including machine learning and deep learning are becoming increasingly popular in TD forecasting [11,12]. Among the AI techniques, artificial neural networks (ANN) provide a potential alternative for solving complex nonlinear problems. Numerous studies showed that ANNs generally outperformed other methods [3,5,13–16]. In general, AI technology can approximate arbitrarily complex nonlinear dynamic systems without any initial or extra information about data such as distribution. This brings considerable benefits and simplifications to modelling, but on the other hand, AI techniques hardly

provide any information about potential determinism or even process understanding. However, since we are interested in TD patterns rather than physical processes, this model property is not the main disadvantage.

With the advancement of ANNs, researchers find that deep learning methods, especially recurrent neural network (RNN) architectures, are more suitable than feedforward neural networks in dealing with the complexity of time series [17]. However, RNN training has the problem of vanishing gradients, so various variants of RNN models have been proposed, such as long short-term memory (LSTM), bidirectional long short-term memory (Bi-LSTM) and gated recurrent unit (GRU). The aforementioned networks are different from other methods in that they back-propagate through immediate historical data and current data, and are more suitable for detecting development trends. The architecture of these networks overcame the weaknesses of traditional RNNs in capturing long-term dependencies, as shown by Bengiot et al. [18]. With this feature, these networks have been widely used to solve time series forecasting problems [11,19–25].

According to Annual Survey Report on Visitors Expenditure and Trends in Taiwan, the total revenue of international tourism increased from US\$5936 million in 2008 to US\$14,411 million in 2019. In recent years, international tourism has become a key service industry in Taiwan. Since 2008, Taiwan's international tourism revenue has exceeded domestic tourism revenue. If this momentum can be maintained, it will contribute to the development of the tourism industry and economic growth in the future.

The outstanding application results of LSTM networks and its variants in different fields show that they can not only seize the changing data trend, but also describe the dependence of time series data. Therefore, this research tries to use an LSTM network and its variants to predict Taiwan's TD.

The number of passengers is still the most popular TD measure over last decades. Since the time series model only requires historical observation of a variable, the cost of data collection and model estimation is low. Hence, this study adapts an LSTM network and its variants to forecast Taiwan's TD. In order to validate the model, a data set including the severe acute respiratory syndrome (SARS) outbreak threatening tourism demand from November 2002 to June 2003 was used to compare the prediction results of the models reported in the other papers. In view of the strong autoregressive pattern of the number of tourists [26], data from the SARS outbreak was used to train the network to predict the impact of the current COVID-19 epidemic on the number of tourists in Taiwan.

The remainder of this paper is organized as follows. Section 2 describes the LSTM, Bi-LSTM, and GRU networks. Section 3 presents data description. Section 4 describes the results and discussion for TD forecasting before our conclusions are provided in Section 5.

2. Methods

2.1. LSTM Network

In RNN, the output can be given back to the network as input, thereby creating a loop structure. RNNs are trained through backpropagation. In the process of backpropagation, RNN will encounter the problem of vanishing gradient. We use the gradient to update the weight of the neural network. The problem of vanishing gradient is when the gradient shrinks as it propagates backwards in time. Therefore, the layers that obtain small gradients will not learn, but will instead cause the network to have short-term memory.

The LSTM network was introduced by Hochreiter and Schmidhuber [27] to alleviate the problem of vanishing gradients. LSTMs can use a mechanism called gates to learn long-term dependencies. These gates can learn which information in the sequence is important to keep or discard. LSTMs have three gates: input, forget and output. Figure 1a shows the architecture of LSTM cell [22,28,29]. The horizontal line between C_{t-1} and C_t is called the cell state. This is the core of the LSTM model, where pointwise addition and multiplication are performed to add or delete information from the memory. These operations are performed using the input and forget gate of the LSTM block, which also

contains the output “tanh” activation function. The computations inside the LSTM neurons are shown as follows [27]:

Forget gate:

$$f_t = \sigma(\mathbf{W}_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

Input gate:

$$i_t = \sigma(\mathbf{W}_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

Output gate:

$$o_t = \sigma(\mathbf{W}_o \cdot [h_{t-1}, x_t] + b_o) \quad (3)$$

Process input:

$$\tilde{C}_t = \tanh(\mathbf{W}_C \cdot [h_{t-1}, x_t] + b_C) \quad (4)$$

Cell update:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (5)$$

Output:

$$h_t = o_t * \tanh(C_t) \quad (6)$$

where σ refers to sigmoid function, h_{t-1} represents the output of pervious cell state, x_t represents the input of current cell state, $\mathbf{W}_f, \mathbf{W}_i, \mathbf{W}_o$ and b_f, b_i, b_o are the weight matrices and bias of the forget, input and output gates, respectively. \mathbf{W}_C and b_C are the weights and bias of the cell state, and “.” means point-wise multiplication. o_t is used to evaluate which part of the cell state to be exported, and h_t calculates the final outputs.

2.2. Bi-LSTM Network

Figure 1b shows the general structure of Bi-LSTM network. One input sequence is processed from right to left, and the other is processed from left to right. This structure allows the model to learn the input sequence in both directions. The interpretations of the forward and backward LSTM network output are combined to generate predictions at the next time step. By using time series data and its reverse copy to make predictions, it can provide supplementary context for the model to learn problems faster and more effectively [30].

Hai et al. [21] surveyed different variants of LSTM (Vanilla, Stacked, Bi-directional), which were applied to the stock prices of 20 companies on the VN Index Stock Exchange during the five-year period from 2015 to 2020. The results show that the most accurate model is Bi-LSTM.

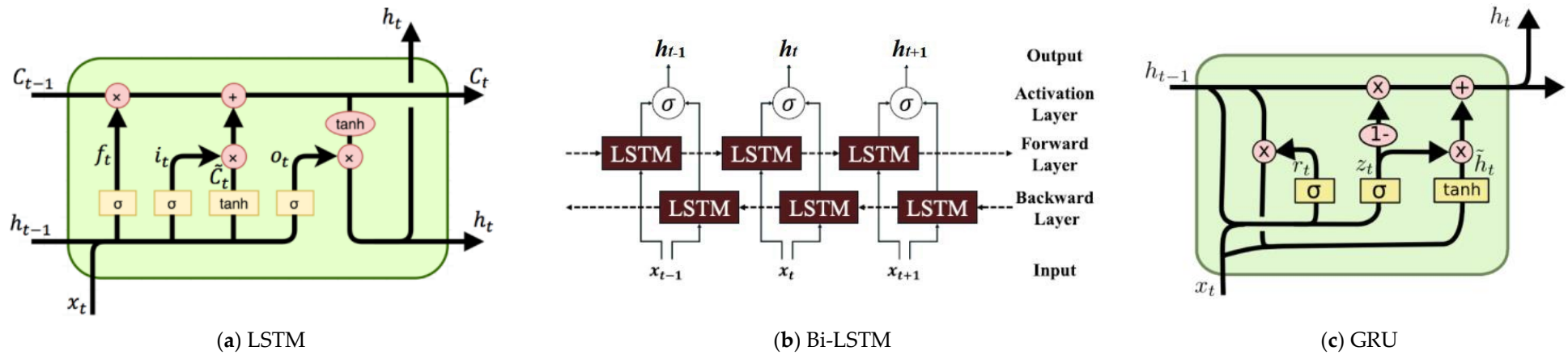


Figure 1. Structure of network according to Ko et al. [22], Graves [28] and Olah [29].

2.3. GRU Network

The GRU network proposed by Cho et al. [31] is a modified LSTM model with two gates, so that each cyclic unit can adaptively seize the dependencies of different time scales. Different from the LSTM network, the GRU structure is not so uncomplicated, but its usefulness has not been reduced, and sometimes even a little better than LSTM [32].

GRU eradicates the cell state and applies the hidden state to transmit information. Another distinction between GRU and LSTM is that the forget gate and input gate in LSTM are combined into an update gate. Figure 1c shows the architecture of GRU cell. The mathematical operations inside the GRU neurons are shown as follows [31]:

Reset gate:

$$r_t = \sigma(\mathbf{W}_r \cdot [h_{t-1}, x_t] + b_r) \quad (7)$$

Update gate:

$$z_t = \sigma(\mathbf{W}_z \cdot [h_{t-1}, x_t] + b_z) \quad (8)$$

Process input:

$$\tilde{h}_t = \tanh(\mathbf{W}_{\tilde{h}} \cdot [r_t \times h_{t-1}, x_t] + b_{\tilde{h}}) \quad (9)$$

Output:

$$h_t = (1 - z_t) \times h_{t-1} + z_t \times \tilde{h}_t \quad (10)$$

where b_r is the bias vector of the reset gate. Its function is the same as that in the LSTM, that is, the smaller r_t is, the less information passes. b_z and $b_{\tilde{h}}$ are the bias of the update gate and cell state, respectively. \mathbf{W}_r , \mathbf{W}_z and $\mathbf{W}_{\tilde{h}}$ are the weight matrices of the reset gate, update gate, and cell state, respectively.

Mean squared error was used as the loss function and “Adam” optimizer [33] was used to find the optimum weights for the networks. All the models were implemented using Keras in Python [34], and a Tensorflow backend [35]. In this study, during the training stage, all possible configurations of manually defined parameter subsets were tried to choose parameters for the network.

The data was normalized to be between 0 and 1. After using LSTM, Bi-LSTM and GRU models to make predictions, the predicted data was inverted and restored to the original state. Equation (11) describes the function used in this study to normalize the dataset:

$$x'_t = \frac{x_t - x_{\min}}{x_{\max} - x_{\min}} \quad (11)$$

where x_t is the input time series, x'_t is the normalized time series, and x_{\min} and x_{\max} are the minimum and maximum values of the time series respectively.

In order to evaluate the forecasting performances of the model, the root mean squared error (RMSE) was used:

$$\text{RMSE} = \sqrt{\frac{1}{M} \sum_{m=1}^M (y_m - y_m^*)^2} \quad (12)$$

where y_m and y_m^* are the observed and predicted values, respectively; M is the number of data samples.

3. Data

The forecast target is the number of tourists visiting Taiwan each month. The data obtained from the official website (<https://stat.taiwan.net.tw/inboundSearch> (accessed on 17 July 2021)) of Taiwan Tourism Bureau, Ministry of Transportation and Communications extend from January 1984 to May 2021. This study uses two series datasets to verify the feasibility and effectiveness of the proposed forecasting models. Series 1 is split into a training dataset, covering the period from January 1984 to August 1998, and testing dataset, for the period from September 1998 to September 2005, as shown in Figure 2. The training and testing ratio is a ratio of 70:30. Series 2 is divided into training dataset, covering the

period from January 1984 to March 2010, and testing dataset, for the period from April 2010 to May 2021 (Please see in Table S1), as shown in Figure 3.

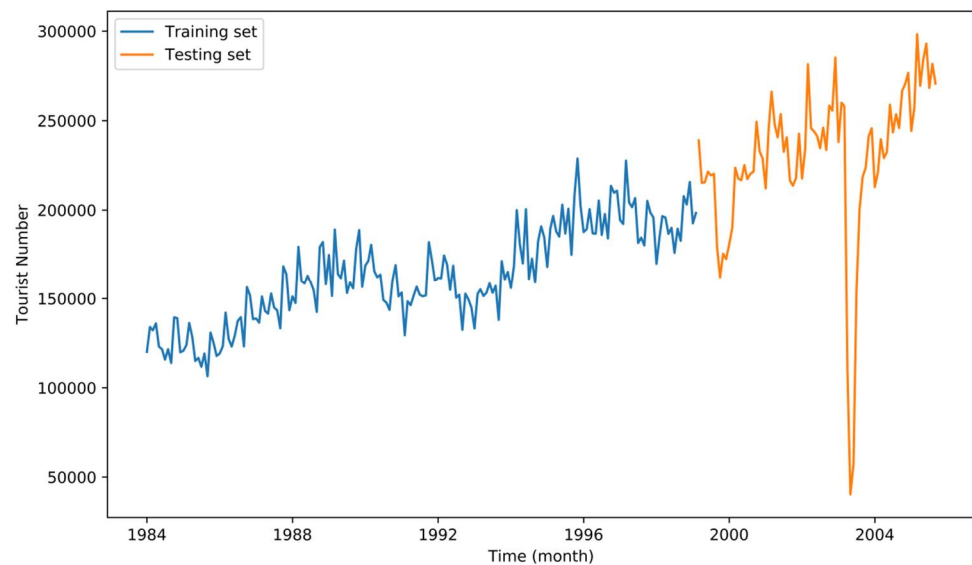


Figure 2. Monthly tourist number of Taiwan from January 1984 to September 2005.

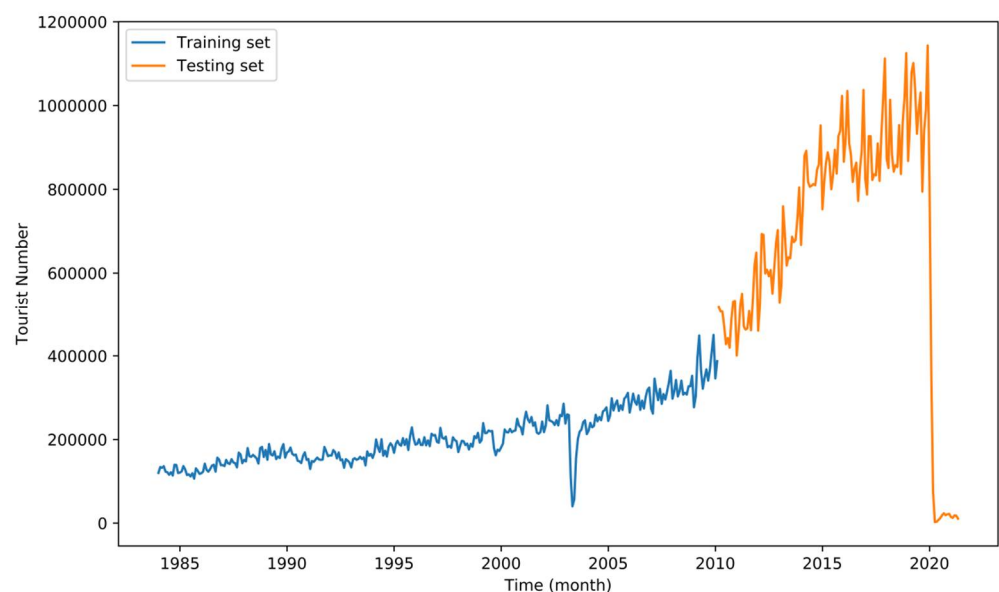


Figure 3. Monthly tourist number of Taiwan from January 1984 to May 2021.

The period of testing dataset of data series 1 covered the SARS outbreak, which had a great impact on Taiwan's TD [36]. The period of testing dataset of data series 2 covers the outbreak of COVID-19. The significance of this research lies in the model's ability to predict time series of catastrophic events, such as the SARS and COVID-19 outbreaks.

4. Results and Discussion

4.1. Series 1

The LSTM, Bi-LSTM, and GRU networks yielded RMSEs of 29,537, 30,264, and 30,531 for the testing dataset, respectively. The results are compared with those of the previous fuzzy time series studies [37–39], as shown in Table 1. RMSE of Huarng et al. [39] is smaller than those of Chen [37] and Huarng et al. [38]. The RMSEs for the period from May 2000 to September 2005 obtained by Huarng et al. [39], LSTM, Bi-LSTM, and GRU are 30,789,

31,182, 34,872, and 34,770 respectively. The actual and various predicted tourist numbers of Taiwan from May 2000 to September 2005 are depicted in Figure 4. Considering the SARS period only (November 2002 to June 2003), the RMSEs for Huarng et al. [39], LSTM, Bi-LSTM, and GRU are 61,863, 59,276, 59,480 and 59,369 respectively. These RMSEs are almost twice the corresponding RMSEs of the entire period, indicating that TD forecasting during this period is difficult. However, the RMSEs of this study achieved 4% better error rates than that of the aforementioned research, indicating that the prediction model can be used to predict time series with catastrophic events. The comparison results show that the LSTM model is slightly better than that of the Bi-LSTM, and GRU in terms of RMSE. The RMSEs for the entire period, including the SARS period, are compared in Table 1.

The reported fuzzy time series models have achieved successful prediction results under the modelling frameworks. However, the models have a potential vulnerability (without long-term dependency). Due to the lack of “memory” functions in the structure, the model is more sensitive to short-term relationships than long-term dependencies, and cannot capture some important recurring features. Furthermore, deep learning approaches are non-parametric and are more generalizable without fuzzification and de-fuzzification.

Table 1. Comparison of forecast.

Month	Actual	Huarng et al. [39]	LSTM	Bi-LSTM	GRU
May-00	216,692	219,138	209,198	207,674	206,979
Jun-00	225,069	217,519	208,465	206,980	206,283
Jul-00	217,302	224,546	215,417	213,552	212,875
Aug-00	220,227	219,122	208,971	207,459	206,763
Sep-00	221,504	220,453	211,399	209,754	209,065
Oct-00	249,352	221,995	212,459	210,756	210,070
Nov-00	232,810	247,299	235,545	232,562	231,973
Dec-00	228,821	235,712	221,837	219,618	218,966
Jan-01	199,800	230,085	218,529	216,493	215,828
Feb-01	234,386	222,144	204,655	203,378	202,671
Mar-01	251,111	232,287	233,190	230,340	229,739
Apr-01	235,251	249,710	249,466	245,695	245,187
May-01	227,021	238,079	234,424	231,505	230,910
Jun-01	239,878	228,914	228,374	225,793	225,169
Jul-01	218,673	238,869	239,059	235,879	235,309
Aug-01	224,208	240,966	221,670	219,460	218,807
Sep-01	193,254	224,076	228,381	225,799	225,175
Oct-01	192,452	215,560	208,305	206,830	206,132
Nov-01	190,500	193,244	205,900	204,556	203,851
Dec-01	210,603	191,470	209,408	207,872	207,177
Jan-02	217,600	208,926	230,045	227,370	226,754
Feb-02	233,896	217,268	209,218	207,693	206,997
Mar-02	281,522	232,541	222,738	220,469	219,820
Apr-02	245,759	279,376	262,144	257,641	257,220
May-02	243,941	267,961	232,569	229,753	229,149
Jun-02	241,378	244,875	231,063	228,331	227,720
Jul-02	234,596	242,421	228,939	226,326	225,705
Aug-02	246,079	236,270	223,318	221,017	220,371
Sep-02	233,613	245,205	232,834	230,003	229,401
Oct-02	258,360	236,077	222,503	220,247	219,598
Nov-02	255,645	256,345	243,001	239,598	239,051
Dec-02	285,303	256,724	240,755	237,478	236,918
Jan-03	238,031	283,235	265,265	260,579	260,181
Feb-03	259,966	240,587	226,166	223,707	223,073
Mar-03	258,128	258,138	244,330	240,851	240,312
Apr-03	110,640	259,062	242,809	239,417	238,868
May-03	40,256	111,762	120,256	123,504	122,983

Table 1. Cont.

Month	Actual	Huarng et al. [39]	LSTM	Bi-LSTM	GRU
Jun-03	57,131	41,693	61,741	68,211	68,356
Jul-03	154,174	55,717	75,754	81,435	81,375
Aug-03	200,614	155,234	156,490	157,801	157,099
Sep-03	218,594	198,470	195,112	194,353	193,627
Oct-03	223,552	217,083	210,043	208,473	207,780
Nov-03	241,349	223,489	214,158	212,362	211,682
Dec-03	245,682	239,859	228,915	226,304	225,682
Jan-04	212,854	245,725	232,505	229,693	229,089
Feb-04	221,020	235,124	205,278	203,968	203,262
Mar-04	239,575	220,528	212,057	210,376	209,689
Apr-04	229,061	238,021	227,445	224,915	224,287
May-04	232,293	231,267	218,728	216,681	216,017
Jun-04	258,861	232,482	221,409	219,213	218,559
Jul-04	243,396	256,818	243,416	239,989	239,444
Aug-04	253,544	246,198	230,611	227,905	227,292
Sep-04	245,915	252,812	239,016	235,838	235,268
Oct-04	266,590	247,735	232,698	229,875	229,272
Nov-04	270,553	264,855	249,808	246,017	245,512
Dec-04	276,680	270,632	253,084	249,105	248,621
Jan-05	244,252	276,447	258,146	253,875	253,425
Feb-05	257,340	266,528	231,321	228,575	227,965
Mar-05	298,282	256,305	242,157	238,802	238,249
Apr-05	269,513	296,152	275,967	270,648	270,336
May-05	284,049	291,862	252,225	248,295	247,805
Jun-05	293,044	282,861	264,230	259,605	259,199
Jul-05	268,269	292,460	271,650	266,588	266,240
Aug-05	281,693	290,673	251,196	247,326	246,829
Sep-05	270,700	280,606	262,286	257,774	257,354
RMSE (May-00~Sep-05)		30,789	31,182	34,872	34,770
RMSE (Nov-02~Jun-03)		61,863	59,276	59,480	59,369

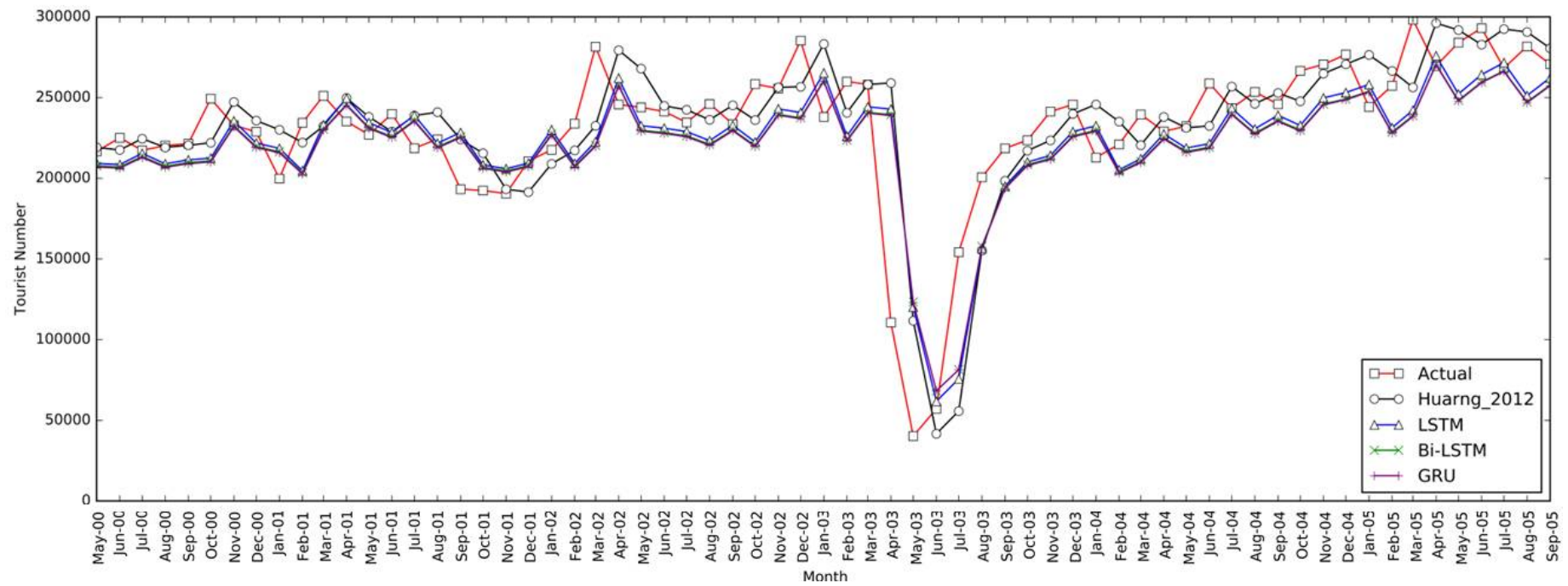


Figure 4. The actual and predicted tourist number of Taiwan from May 2000 to September 2005.

4.2. Series 2

The RMSEs for the testing dataset of the LSTM, Bi-LSTM and GRU networks are 100,410, 102,754 and 105,768, respectively, showing that they have similar performance. Even so, the LSTM model has slightly higher accuracy compared to Bi-LSTM and GRU. The model training vs. validation loss and actual data vs. prediction are depicted in Figures 5–10. The aforementioned networks successfully identify the future trends and can emulate instances of extreme arrival dips.

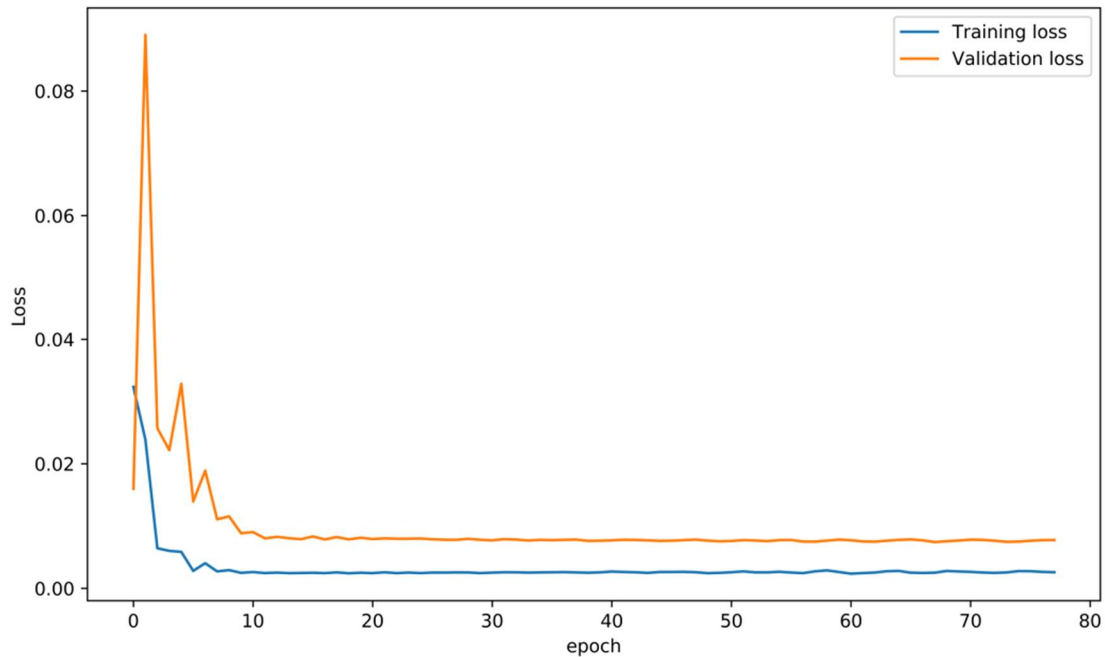


Figure 5. Model training vs. validation loss for LSTM.

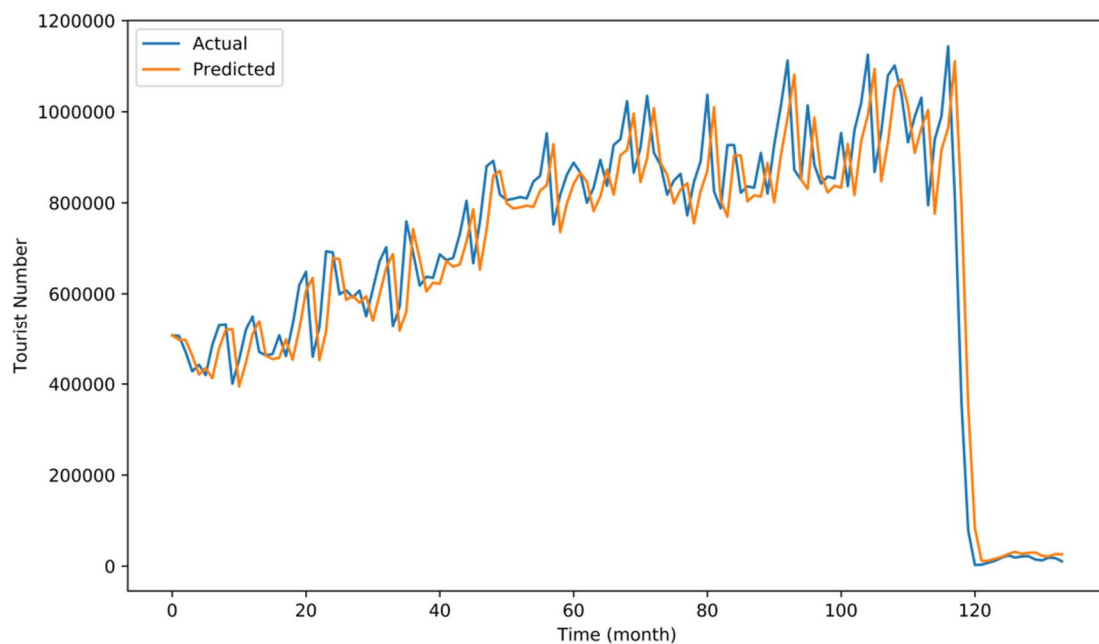


Figure 6. Actual data vs. prediction for LSTM.

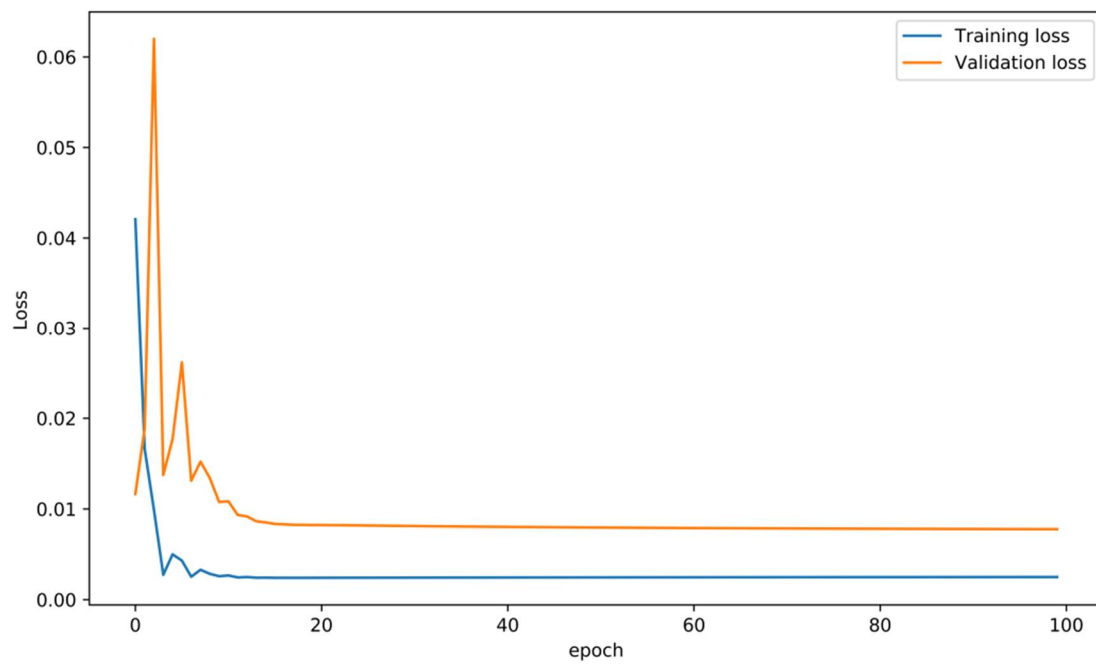


Figure 7. Model training vs. validation loss for Bi-LSTM.

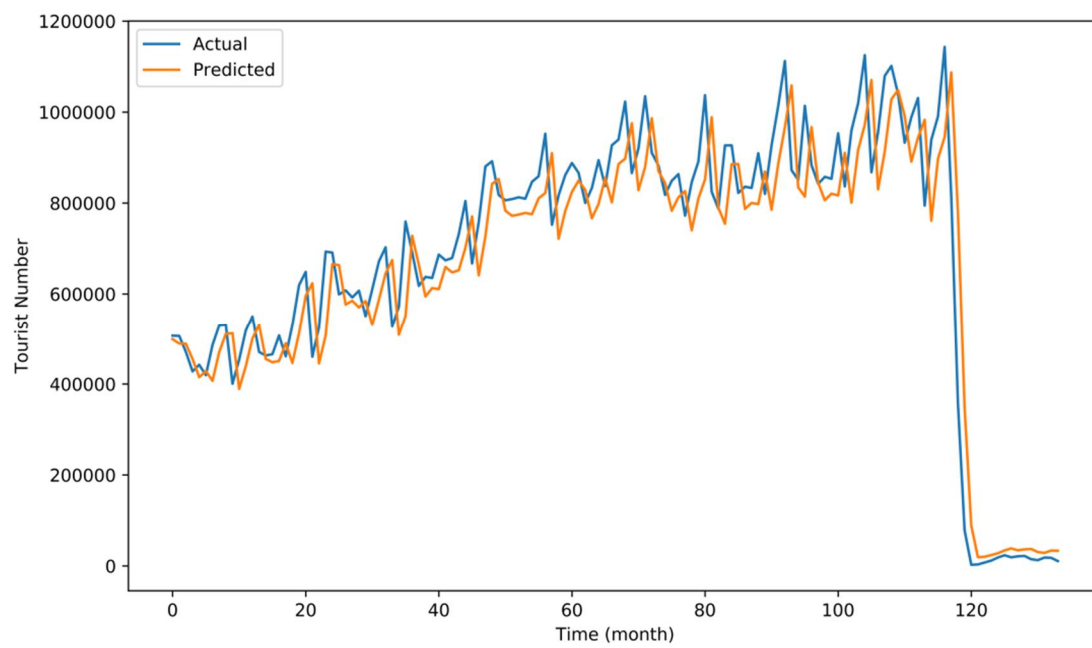


Figure 8. Actual data vs. prediction for Bi-LSTM.

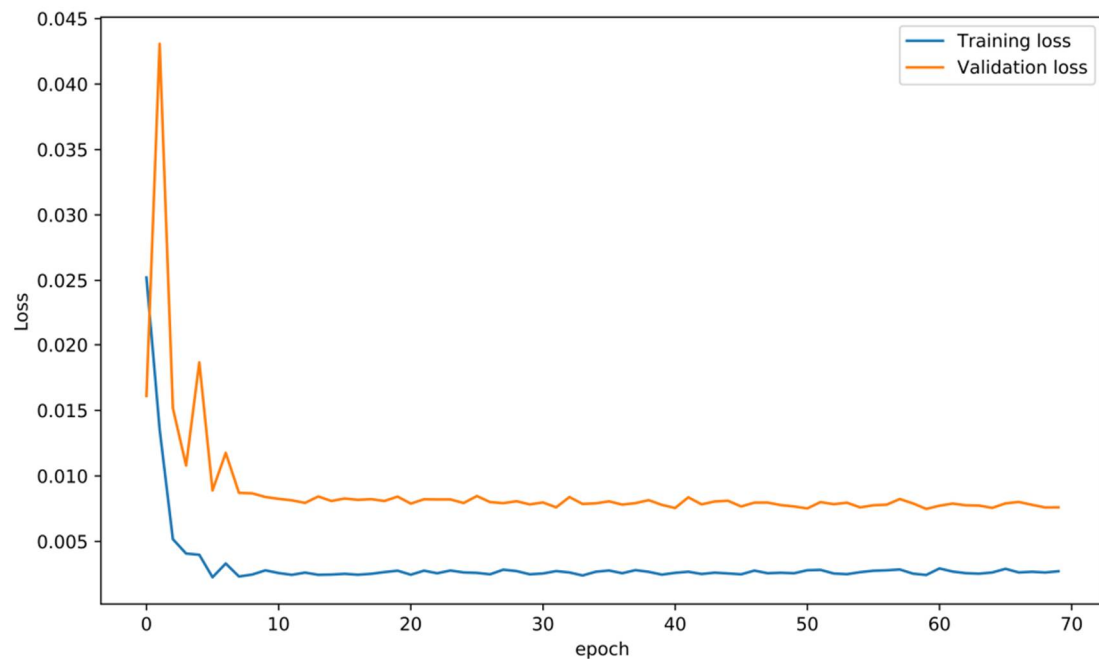


Figure 9. Model training vs. validation loss for GRU.

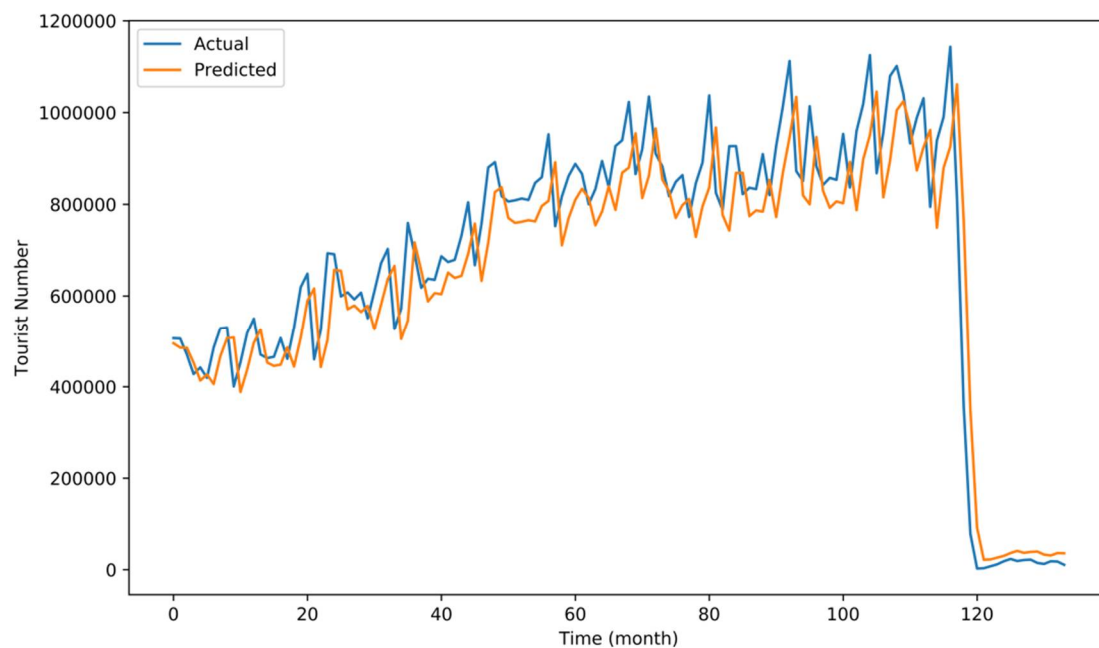


Figure 10. Actual data vs. prediction for GRU.

SARS and COVID-19 are two catastrophic events that profoundly affect the world's TD [40]. Polyzos et al. [41] used the data from the SARS epidemic outbreak to train an LSTM network, similar to the approach of Law et al. [11]. In the first training phase, the error is returned to the network to calibrate the model. In addition, errors will continue to be used in the gates of the network. Moreover, the LSTM network does not react to the lags between events in the time series. Therefore, when we try to derive unknown prediction models, the LSTM algorithm works better than other ANNs (such as hidden Markov, support vector regression, etc.) or other prediction techniques (such as ARIMA) [11].

5. Conclusions

The purpose of this study is to adapt an LSTM network and its variants to improve Taiwan's TD forecasting. The results show that the proposed models are more simple and effective than others to forecast nonlinear data with shocks. These techniques reveal adequate to other catastrophic situations that can affect the tourism industry.

To overcome statistical complexities through analysing time series, this study empirically analyses the accuracy of LSTM, Bi-LSTM and GRU models applied in Taiwan's TD forecasting with shocks—namely, the SARS epidemic and the COVID-19 pandemic. The forecasting models of deep learning perform better than the other three fuzzy time series models when considering the period of catastrophic events. The results show that the use of the LSTM network and its variants can be applied to the arrival time series, given its strong autoregressive nature, using a calibration network with training data from a similar a past event—namely, the SARS epidemic. From the global error perspective, the performance of the LSTM model is slightly better than those of the Bi-LSTM and GRU in terms of the RMSE value.

In other destinations with similar techniques, Polyzos et al. [41] employed an LSTM network to forecast the effect of the current pandemic COVID-19 outbreak on the arrivals of Chinese tourists to the USA and Australia. Kulshrestha et al. [42] ascertained the validity of the Bayesian Bi-LSTM model using the TD data of Singapore. Therefore, the robustness of TD forecasting using an LSTM network and its variants is not country-specific.

However, the proposed models are unable to interpret TD from the economic perspective, and therefore provide little help in policy evaluation. Incorporating additional explanatory variables such as weather data and search engine data [11,43] is promising to increase the accuracy of TD forecasting. Therefore, establishing a comprehensive ability to summarize variables selection should be the research direction going forward.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/a14080243/s1>, Table S1: Dataset.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used to support the findings of this study are provided as supplementary materials.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Frechtling, D.C. *Forecasting Tourism Demand: Methods and Strategies*; Butterworth-Heinemann: Oxford, UK, 2001.
2. Lim, C. Review of International Tourism Demand Models. *Ann. Tour. Res.* **1997**, *24*, 835–849. [\[CrossRef\]](#)
3. Song, H.; Li, G. Tourism Demand Modelling and Forecasting: A Review of Recent Research. *Tour. Manag.* **2008**, *29*, 203–220. [\[CrossRef\]](#)
4. Hassani, H.; Silva, E.S.; Antonakakis, N.; Filis, G.; Gupta, R. Forecasting Accuracy Evaluation of Tourist Arrivals. *Ann. Tour. Res.* **2017**, *63*, 112–127. [\[CrossRef\]](#)
5. Witt, S.F.; Witt, C.A. Forecasting Tourism Demand: A Review of Empirical Research. *Int. J. Forecast.* **1995**, *11*, 447–475. [\[CrossRef\]](#)
6. Peng, B.; Song, H.; Crouch, G.I. A Meta-analysis of International Tourism Demand Forecasting and Implications for Practice. *Tour. Manag.* **2014**, *45*, 181–193. [\[CrossRef\]](#)
7. Zhang, G.P. Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing* **2003**, *50*, 159–175. [\[CrossRef\]](#)
8. Tseng, F.M.; Tzeng, G.H.; Yu, H.C.; Yuan, B.J.C. Fuzzy ARIMA Model for Forecasting the Foreign Exchange Market. *Fuzzy Sets Syst.* **2001**, *118*, 9–19. [\[CrossRef\]](#)
9. Li, G.; Song, H.; Witt, S. Recent Development in Econometric Modelling and Forecasting. *J. Travel Res.* **2005**, *44*, 82–99. [\[CrossRef\]](#)
10. Zhang, Y.; Li, G.; Muskat, B.; Law, R. Tourism Demand Forecasting: A Decomposed Deep Learning Approach. *J. Travel Res.* **2021**, *60*, 981–997. [\[CrossRef\]](#)
11. Law, R.; Li, G.; Fong, D.K.C.; Han, X. Tourism Demand Forecasting: A Deep Learning Approach. *Ann. Tour. Res.* **2019**, *75*, 410–423. [\[CrossRef\]](#)

12. Sun, S.; Wei, Y.; Tsui, K.L.; Wang, S. Forecasting Tourist Arrivals with Machine Learning and Internet Search Index. *Tour. Manag.* **2019**, *70*, 1–10. [\[CrossRef\]](#)
13. Burger, C.J.S.C.; Dohnal, M.; Kathrada, M.; Law, R. A Practitioners Guide to Time-series Methods for Tourism Demand Forecasting—A Case Study of Durban, South Africa. *Tour. Manag.* **2001**, *22*, 403–409. [\[CrossRef\]](#)
14. Cho, V. A Comparison of Three Different Approaches to Tourist Arrival Forecasting. *Tour. Manag.* **2003**, *24*, 323–330. [\[CrossRef\]](#)
15. Kon, S.C.; Turner, W.L. Neural Network Forecasting of Tourism Demand. *Tour. Econ.* **2005**, *11*, 301–328. [\[CrossRef\]](#)
16. Wong, K. The Relevance of Business Cycles in Forecasting International Tourist Arrivals. *Tour. Manag.* **1997**, *18*, 581–586. [\[CrossRef\]](#)
17. Hüsken, M.; Stagge, P. Recurrent Neural Networks for Time Series Classification. *Neurocomputing* **2003**, *50*, 223–235. [\[CrossRef\]](#)
18. Bengiot, Y.; Simard, P.; Frasconi, P. Learning Long-term Dependencies with Gradient Descent Is Difficult. *IEEE Trans. Neural Netw.* **1994**, *5*, 157–166. [\[CrossRef\]](#)
19. Bai, Y.; Xie, J.; Liu, C.; Tao, Y.; Zeng, B.; Li, C. Regression Modelling for Enterprise Electricity Consumption: A Comparison of Recurrent Neural Network and Its Variants. *Int. J. Electr. Power Energy Syst.* **2021**, *126*, 106612. [\[CrossRef\]](#)
20. Fischer, T.; Krauss, C. Deep Learning with Long Short-term Memory Networks for Financial Market Predictions. *Eur. J. Oper. Res.* **2018**, *270*, 654–669. [\[CrossRef\]](#)
21. Hai, P.N.; Tien, N.M.; Hieu, H.T.; Chung, P.Q.; Son, N.T.; Ha, P.N.; Son, N.T. An Empirical Research on the Effectiveness of Different LSTM Architectures on Vietnamese Stock Market. *ACM Int. Conf. Proc. Ser.* **2020**, *2020*, 144–149. [\[CrossRef\]](#)
22. Ko, M.S.; Lee, K.; Kim, J.K.; Hong, C.W.; Dong, Z.Y.; Hur, K. Deep Concatenated Residual Network with Bidirectional LSTM for One-hour-ahead Wind Power Forecasting. *IEEE Trans. Sustain. Energy* **2021**, *12*, 1321–1335. [\[CrossRef\]](#)
23. Song, X.; Liu, Y.; Xue, L.; Wang, J.; Zhang, J.; Wang, J.; Jiang, L.; Cheng, Z. Time-series Well Performance Prediction Based on Long Short-term Memory (LSTM) Neural Network Model. *J. Pet. Sci. Eng.* **2020**, *186*, 106682. [\[CrossRef\]](#)
24. Yuan, S.; Wang, C.; Mu, B.; Zhou, F.; Duan, W. Typhoon Intensity Forecasting Based on LSTM Using the Rolling Forecast Method. *Algorithms* **2021**, *14*, 83. [\[CrossRef\]](#)
25. Zhang, J.; Zhu, Y.; Zhang, X.; Ye, M.; Yang, J. Developing a Long Short-term Memory (LSTM) Based Model for Predicting Water Table Depth in Agricultural Areas. *J. Hydrol.* **2018**, *561*, 918–929. [\[CrossRef\]](#)
26. Athanasopoulos, G.; de Silva, A. Multivariate Exponential Smoothing for Forecasting Tourist Arrivals. *J. Travel Res.* **2012**, *51*, 640–652. [\[CrossRef\]](#)
27. Hochreiter, S.; Schmidhuber, J. Long Short-term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#)
28. Graves, A.; Liwicki, M.; Fernández, S.; Bertolami, R.; Bunke, H.; Schmidhuber, J. A Novel Connectionist System for Unconstrained Handwriting Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 855–868. [\[CrossRef\]](#)
29. Olah, C. Understanding LSTM Networks. 2015. Available online: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/> (accessed on 7 July 2021).
30. Yildirim, Ö. A Novel Wavelet Sequences Based on Deep Bidirectional LSTM Network Model for ECG Signal Classification. *Comput. Biol. Med.* **2018**, *96*, 189–202. [\[CrossRef\]](#)
31. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations Using RNN Encoder-decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; Moschitti, A., Pang, B., Daelemans, W., Eds.; Association for Computational Linguistics: Doha, Qatar, 2014; pp. 1724–1734. [\[CrossRef\]](#)
32. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modelling. In *NIPS 2014 Deep Learning and Representation Learning Workshop*; WikiCFP: Montreal, QC, Canada, 2014; pp. 1–12. Available online: <https://arxiv.org/abs/1412.3555> (accessed on 17 July 2021).
33. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations*, 3rd ed.; ICLR: San Diego, CA, USA, 2015; Available online: <https://arxiv.org/abs/1412.6980> (accessed on 17 July 2021).
34. Python Software Foundation. Python 3.9 Documentation. 2021. Available online: <https://docs.python.org/3.9/> (accessed on 7 July 2021).
35. Ketkar, N. *Deep Learning with Python*; Apress: New York, NY, USA, 2017.
36. Min, J.C.H. The Effect of the SARS Illness on Tourism in Taiwan: An Empirical Study. *Int. J. Manag.* **2005**, *22*, 497–506.
37. Chen, S.M. Forecasting Enrolments Based on Fuzzy Time Series. *Fuzzy Sets Syst.* **1996**, *81*, 311–319. [\[CrossRef\]](#)
38. Huarng, K.H.; Moutinho, L.; Yu, H.K. An Advanced Approach to Forecasting Tourism Demand in Taiwan. *J. Travel Tour. Mark.* **2007**, *21*, 15–24. [\[CrossRef\]](#)
39. Huarng, K.H.; Yu, H.K.; Moutinho, L.; Wang, Y.C. Forecasting Tourism Demand by Fuzzy Time Series Models. *Int. J. Cult. Tour. Hosp. Res.* **2012**, *6*, 377–388. [\[CrossRef\]](#)
40. Ying, T.; Wang, K.; Liu, X.; Wen, J.; Goh, E. Rethinking Game Consumption in Tourism: A Case of the 2019 Novel Coronavirus Pneumonia Outbreak in China. *Tour. Recreat. Res.* **2020**, *46*, 304–309. [\[CrossRef\]](#)
41. Polyzos, S.; Samitas, A.; Spyridou, A.E. Tourism Demand and the COVID-19 Pandemic: An LSTM Approach. *Tour. Recreat. Res.* **2021**, *46*, 175–187. [\[CrossRef\]](#)
42. Kulshrestha, A.; Krishnaswamy, V.; Sharma, M. Bayesian BiLSTM Approach for Tourism Demand Forecasting. *Ann. Tour. Res.* **2020**, *93*, 102925. [\[CrossRef\]](#)
43. Bi, J.W.; Liu, Y.; Li, H. Daily Tourism Volume Forecasting for Tourist Attractions. *Ann. Tour. Res.* **2020**, *83*, 102923. [\[CrossRef\]](#)