Annals of Tourism Research 75 (2019) 410–423

Contents lists available at ScienceDirect

Annals of Tourism Research

journal homepage: www.elsevier.com/locate/annals

Tourism demand forecasting: A deep learning approach

Rob Law[a,*], Gang Li[b], Davis Ka Chio Fong[c], Xin Han[d]

[a] School of Hotel & Tourism Management, The Hong Kong Polytechnic University, 17 Science Museum Rd, TST East, Kowloon, Hong Kong
[b] School of Information Technology, Deakin University, Geelong, VIC 3216, Australia
[c] Faculty of Business Administration, University of Macau, E22, Avenida da Universidade, Taipa, Macau, China
[d] School of Computer Science, Xi'an Shiyou University, Shaanxi 710065, China

ARTICLE INFO

Keywords:
Tourism demand forecasting
Deep learning
Long-short-term-memory
Attention mechanism
Feature engineering
Lag order

ABSTRACT

Traditional tourism demand forecasting models may face challenges when massive amounts of search intensity indices are adopted as tourism demand indicators. Using a deep learning approach, this research studied the framework in forecasting monthly Macau tourist arrival volumes. The empirical results demonstrated that the deep learning approach significantly outperforms support vector regression and artificial neural network models. Moreover, the construction and identification of highly relevant features from the proposed deep network architecture provide practitioners with a means of understanding the relationships between various tourist demand forecasting factors and tourist arrival volumes.

This article also launches the Annals of Tourism Research Curated Collection on Tourism Demand Forecasting, a special selection of research in this field

Introduction

Unoccupied hotel rooms, unsold event tickets and unconsumed food items represent unnecessary costs as well as unrealized revenue, a combination that poses a potential threat to financial sustainability. In short, many tourism and hospitality products cannot be stockpiled for future use, making the need for accurate tourism demand forecasting crucial (Frechtling & Frechtling, 2001). As such, accurate tourism demand forecasts provide valuable aid for strategic, tactical and operational decision making (Lim, 1997; Song & Li, 2008). For example, governments need accurate tourism demand forecasts for informed decision making on issues such as infrastructure development, and accommodation site planning (Chan, Hui, & Yuen, 1999); organizations need the forecasts to make tactical decisions related to tourism promotion brochures, and tourism and hospitality practitioners need accurate forecasts for operational decisions such as staffing and scheduling. Accordingly, accurate tourism demand forecasting is an essential element that provides crucial information for tourism-related decision making.

The majority of tourism demand forecasting studies fall under the well-established category of quantitative approach, which constructs the model from training data on past tourist arrival volumes and various tourism demand forecasting factors (Song & Li, 2008; Wu, Song, & Shen, 2017). With the advances in Web technology, search engines have become essential for tourists in planning their trips by obtaining destination information on hotels, attractions, and weather. SII data have been acknowledged as a potential indicator of tourism demand in the destination market (Dergiades, Mavragani, & Pan, 2018; Fesenmaier, Xiang, Pan, & Law, 2011; Yang, Pan, & Song, 2014), and researchers have examined Search Intensity Indices (SII) data for tourism demand forecasting (Volchek, Liu, Song, & Buhalis, 2018; Xiang & Pan, 2011).

Although incorporating SII data is promising for accurate tourism demand forecasting, some practical challenges have arisen for

* Corresponding author.
E-mail addresses: rob.law@polyu.edu.hk (R. Law), gang.li@deakin.edu.au (G. Li), davisfong@umac.mo (D.K.C. Fong), xin.han@tulip.org.au (X. Han).

https://doi.org/10.1016/j.annals.2019.01.014
Received 9 October 2018; Received in revised form 27 January 2019; Accepted 29 January 2019
0160-7383/

practitioners attempting to use them with traditional forecasting models. More specifically, the following two practical barriers exist.

The first barrier is related to feature engineering. As mentioned by Song and Li (2008), a large number of factors have been considered as potential tourism demand forecasting determinants or indicators, examples include exchange rate, tourism prices, travel costs and various SII data. As the number of potentially influential factors increases, available training data in the feature space become increasingly sparse. In tourism demand forecasting, this means insufficient data for the construction of a reliable model. Many forecasting models have difficulties in learning from the training data with too many explanatory factors (Guyon & Elisseeff, 2003). Therefore, feature engineering has been an important step in forecasting model construction because it focuses on extracting the best set of relevant features from a large variety of potential factors (Zhang, Zhang, & Yang, 2003).

Even though the meaning of factors, such as search engine keywords, are largely known, in the real world scenario, thousands of potential keywords may be related to a destination tourism market. Currently, feature engineering for tourism demand forecasting depends largely on domain knowledge on the destination tourism market and requires significant human efforts in selecting effective features (Xiang & Pan, 2011; Yang, Pan, Evans, & Lv, 2015).

Lag order selection is the second barrier. Despite an increasing number of tourism demand forecasting methods adopting the SII data, only a small number of studies detect the lead or lag relationships between time series data. Most existing works examined the hypothesis of no predictability through Pearson correlation coefficients or the Granger causality test (Dergiades et al., 2018; Li, Pan, Law, & Huang, 2017), in which the null hypothesis is investigated by testing whether the lagged values of a factor are strongly related or contributing significantly to tourist arrival volume. However, neither Pearson correlation coefficients or Granger causality test works reliably when the underlying relationship is nonlinear (Reshef et al., 2011). Hence, the capability of selecting all potentially interesting relationships in a dataset will allow tremendous versatility in the construction of more accurate forecasting models.

Time series, econometric and artificial intelligence models provide excellent forecasting performance, and they break the barrier of feature engineering based on the domain knowledge of the destination market. However, adopting existing forecasting methods for every destination market is inconvenient because of the effect from complicated real-world situations. This is especially true when massive amounts of SII data are adopted as tourism demand indicators during which they may require significant domain expertise to confine the uncertainty. Moreover, for each SII feature, the number of effective lags may be different as well. The difficulty increases when combined with other issues, such as the language bias and platform bias (Dergiades et al., 2018).

Recent advances in artificial intelligence, especially the deep learning techniques, have provided methods of breaking the above barriers and enabling more accurate tourism demand forecasting (Pouyanfar et al., 2018). Deep network architectures extend the artificial neural network models with more than two nonlinear processing layers, and have been proven effective for various applications. Their success is attributed usually to their built-in feature engineering capability, which motivates us to break those two barriers simultaneously within the machine learning process. With regard to the contextual information for time series analysis, deep network architectures also have certain advantages in flexible yet discriminative non-linear relationships. Specifically, Recurrent Neural Network (RNN), Long-Short-Term-Memory (LSTM) and Attention Mechanism are capable of handling and learning long-term dependencies. These properties make deep learning an alternative solution to tourism demand forecasting. In this paper, we aim to fill the void by proposing a deep learning approach to tourism demand forecasting and address the previously mentioned two practical barriers simultaneously.

The rest of this paper is organized as follows. The Literature review section reviews the related literature on tourism demand forecasting, and introduces the deep learning technique. The Methodology section describes the conceptual framework of tourism demand forecasting with deep learning. The Empirical study section provides a case study on Macau tourist arrivals, and analyzes the comparison results with baseline methods. Finally, conclusions and implications for future works are summarized in the Conclusions section.

Literature review

Tourism is an important source of economic growth as well as foreign exchange earnings and jobs creation. Accurate tourism demand forecasts are paramount because they provide vital information to tourism practitioners and researchers for making decisions on activities such as earmarking resources, as well as identifying priorities and potential risks. This section provides a brief literature review on tourism demand forecasting studies and the deep learning technique that motivates this work.

Tourism demand forecasting

Tourism demand forecasting studies can be categorized broadly into qualitative and quantitative approaches. Among them, the qualitative approach, such as the symposium Delphi and consensus methods, is usually dependent on the qualitative intuition, experience and insight on a specific tourism market. However, these methods are often considered as “artistic in nature”, and with poor generalization capability (Moutinho & Witt, 1995; Witt & Witt, 1995). Accordingly, researchers have been working on the quantitative approach that estimates quantitative relationships among different observations in tourism data. Based on the past data of factors and tourism volume, the constructed model can then be used to predict future tourist arrival volumes. Generally speaking, two strategies have been adopted in the quantitative approach to improve the performance. The first strategy attempts to incorporate more relevant factors that potentially affect tourists' travel motivations, while the second strategy is to adopt more sophisticated models with better generalization capability on future trends.

In terms of model construction, tourism demand forecasting studies rely heavily on input factors, which are expected to be highly related to tourism demand, without missing or incorrect values. Using different criteria, tourism demand forecasting factors can be categorized into different ways. Based on whether they reflect directly or indirectly tourism demand, they can be categorized into determinants and indicators.

Determinants are the primary factors for forecasting. Conventional economic theories, such as consumption behavior theory and utility theory, suggest that both quantitative and qualitative economic factors, such as price, income, and advertising influence tourism demand (Goh & Law, 2003). However, qualitative economic factors are rarely incorporated into forecasting models because their quantification is usually arduous. In contrast, quantitative economic factors are commonly used because they are measurable and can be used easily as features for most forecasting models. Given the nature of tourism demand, the inclusion of only economic factors is insufficient. Prior studies have focused on how non-economic determinants could reflect travel motivations, and how travel motivation could further affect the choices of destinations. For example, Goh, Law, and Mok (2008) introduced qualitative non-economic determinants including special events, climate index, and leisure time index.

Based on the connection with the source market, those determinants can also be divided into push, pull, and resistance factors (Frechtling & Frechtling, 2001). Among them, the pull factors are attributes of the destination tourism market, such as the quality of the natural resources and Foreign Direct Investments for social ties (Meleddu & Pulina, 2016). Push factors are attributes related to the source market, such as leisure time, per capita income, consumers sentiment, and mood (Martins, Gan, & Ferreira-Lopes, 2017). In contrast, resistance factors include those that constrain travel from the source market to the destination. Examples of these factors include perceived corruption and relative prices (Poprawe, 2015; Saha & Yap, 2015).

Tourism demand is determined by the determinants from the economic theory. To further improve forecasting accuracy, some leading indicators which are considered as the secondary factors can also be included into the model (Volchek et al., 2018). With the advances in Web technology, most tourists resort to search engines for information on all aspects of a trip, ranging from selecting destinations, booking flights, to reserving accommodations, and planning activities. SII data reflect the attention of tourists, so they are considered effective indicators for tourism demand and have been introduced into various tourism demand forecasting models. Xiang and Pan (2011) analyzed the relationship between tourists' search queries of US cities and the attractiveness of the city. The researchers concluded that SII data represent important indicators on the scale of tourism in the destination market. Choi and Varian (2012) used Google Trends index to forecast tourist demand of Hong Kong from nine source countries, and confirmed the usefulness of SII data. According to Yang et al. (2014), SII data reveal the preferences of tourists, provide more prompt data, and depict the timely changes in tourists' preferences. Pan and Yang (2017) also showed the effectiveness of SII data in the hotel occupancy forecasting. Those characteristics make them rich information to traditional univariate time-series models because they can help address the inherent specification problem when encountering sudden changes in econometric patterns (Bangwayo-Skeete & Skeete, 2015).

The selection of search engines depends largely on their popularity in the source tourism markets. Two commonly used SII data in the existing literature are provided by Google (Önder & Gunter, 2016a, 2016b; Rivera, 2016) and Baidu (Yang et al., 2015). Among them, Google Trends are weekly or monthly normalized intensity index, while Baidu Index are daily search volume.

In terms of forecasting models, a wide range of methods have been introduced into tourism demand forecasting. According to Song and Li (2008), these methods can be categorized into time-series models, econometric models and AI models.

Time-series and econometric models are well-adopted in tourism demand forecasting (Andrew, Cranage, & Lee, 1990; Frechtling & Frechtling, 2001). In particular, many popular methods are variants of autoregressive moving average model (Gunter & Onder, 2015), whereas sophisticated models, such as Markov-switching model, Bayesian model, generalized dynamic factor model, and time-varying parameter models have been proposed as improvements (Akin, 2015; Athanasopoulos & Hyndman, 2008; Guizzardi & Stacchini, 2015). From the perspective of methodology, methods in this category utilize past time-series patterns and explore the relationships between various tourism demand factors and tourist arrival volumes. The primary task of forecasting model construction is to incorporate suitable factors to reduce forecasting errors as measured by some performance indicators such as root mean squared errors (RMSE), mean absolute errors (MAE), or mean absolute percentage error (MAPE).

Artificial Intelligence models are related to machine learning and soft computing methods that have been adopted to tourism demand forecasting. For example, based on multivariate regression analyses, Law and Au (1999) have developed methods to capture the nonlinear relationship using the neural networks that mimic the process of human brain. Goh et al. (2008) further adopted the rough set approach to improve comprehensibility of the constructed tourism demand forecasting model. Alvarez-Diaz, Mateu-Sbert, and Rossello-Nadal (2009) indicated that an evolutionary computing method, Genetic Programming, is robust and can easily interpret forecasting of monthly tourist arrivals to the Balearic Islands of Spain. Machine learning techniques such as support vector regression (SVR) model has also been found effective in modeling the nonlinear data series (Pai, Hong, Chang, & Chen, 2006; Zhang, Huang, Li, & Law, 2017). Some recent studies have also shown that ensemble methods which integrate results from different models generate better results (Chan, Witt, Lee, & Song, 2010; Coshall & Charlesworth, 2011; Shen, Li, & Song, 2011; Zhou, 2012).

Because of the "No Free Lunch" theorem (Wolpert, 1996; Wolpert & Macready, 1997), it is well recognized that no single method outperforms others on all scenarios in terms of accuracy, and all methods have their own limitations. Typically, time-series and econometrics models rely on the stability of historical patterns and economic structure, while artificial intelligence models are dependent on the quality and size of available training data.

Deep learning

Artificial intelligence models such as neural network and SVR have found successful applications in tourism demand forecasting. Hinton, Osindero, and Teh (2006) made a breakthrough in efficient training of deep network architectures via greedy layer-wise pre-training, which enables a wide range of practical implementations of deep learning. Compared with traditional artificial intelligence

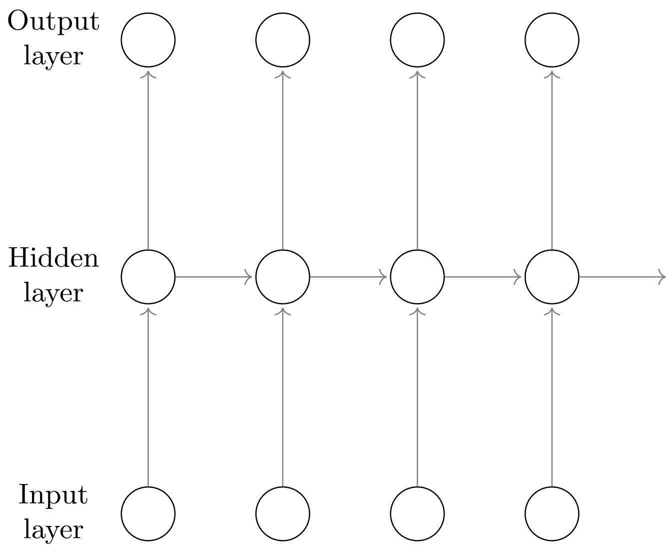

Fig. 1. RNN architecture.

models, deep learning technique provides a mechanism of feature engineering that extracts discriminative features with minimal domain knowledge and human effort (Pouyanfar et al., 2018).

After decades of development, deep learning has experienced phenomenal success in a wide range of challenging artificial intelligence applications that range from pattern recognition tasks such as image captioning (Le Cun, Bengio, & Hinton, 2015) and natural language processing (Socher, Bengio, & Manning, 2012), to forecasting problems for sequential data such as finance prediction and forecasting of changing directions in trades (Bao, Yue, & Rao, 2017).

In this section, two popular deep network architectures are discussed. These architectures have been shown with great success on time series forecasting These architectures are RNN and LSTM with attention mechanism.

RNN is a widely used deep network architecture that utilizes the sequential information (Cho et al., 2014). RNN works by selectively passing information across the time steps while processing data elements. This property is essential for applications including tourism demand forecasting where the embedded structure in the sequential time-series data conveys useful context information. As illustrated in Fig. 1b, both input x and output y of an RNN are time-series data, though either can be a single data point. RNN preserves its memory in the fixed sized hidden layer neuron, which captures all previously processed information. The output of the neuron is then generated based on the current input and the previous hidden layer neuron state through the feedback loop in the network. RNN can model the dependency relationship between sequences of elements through loops, and has found successful applications on non-time series data, including genetic data (Baldi & Pollastri, 2003).

LSTM is an extension of RNN which has not only the recurrent learning unit inside the network but also several gates to capture the longer states from the beginning unit and the shorter states from the last unit. By having this feature, LSTM has been broadly used to solve time series forecasting problems.

Attention mechanism is a feature engineering method which works along with various deep network architectures. By assigning different weight percentages to different inputs, the model could learn the importance of the input without having to do it before fitting into the model. Attention mechanism on LSTM could perfectly fit into tourism demand forecasting and provides an end-to-end solution to both feature selection and prediction.

Rationale of this work

Significant advances have been made in tourism demand forecasting. However, this development has not been matched by parallel improvement in feature engineering for tourism demand forecasting, despite the fact that the performance of a demand forecasting model is driven mainly by the superiority of features in the training data. More specifically, there are two practical limitations related to tourism demand forecasting.

The first limitation is related to feature engineering. For secondary tourism demand forecasting factors, the purpose of feature engineering is mainly query selection that aims to collect tourism related search engine keywords. However, the presence of long tail in SII data implies that a large number of search queries with small search intensities exist, which reflect unique and heterogeneous travelling experiences (Yang et al., 2014). While some redundant and irrelevant queries can be eliminated by using common senses or domain knowledge, determining the optimal subset of relevant queries requires significant efforts from human experts.

Lag order selection is the second limitation. The relationship between time series data is dependent critically on the lag order. An important preliminary step in tourism demand forecasting is to select the lag-order of time series variables using methods such as Granger causality test (Li et al., 2017). However, most tests do not account for the latent confounding effects or capture non-linear relationships (Dergiades et al., 2018). Inaccurate lag order selection will render the subsequent forecasting model construction less effective.

With the heightened development in artificial intelligence, deep learning technique is regarded as a promising alternative to tourism demand forecasting models. This is due to two unique properties compared with traditional neural network models. First, the essence of forecasting models is to construct the model that approximate a non-linear relationship between the input and outputs. Through learning the non-linear combinations of features in deeper layers of the network, a deep network architecture can naturally learn the highly non-linear correlations. Second, one of the most exciting properties of deep learning is that it can automatically construct suitable features at different network layers, and then comes with a built-in mechanism of feature engineering. Moreover, the temporally local correlation between various factors and tourism demand can be exploited to reduce lag selection because it can carefully craft related features from all raw input data. With these properties, deep learning provides the potential to alleviate the dependency on domain expertise. In view of this trend, this study aims to establish a deep learning approach to tourism demand forecasting by adopting a deep network architecture to extract automatically the influential features from various potential factors with suitable lag orders.

Methodology

This study proposes a conceptual framework of tourism demand forecasting with deep learning, and then describes the deep network architecture that addresses the above mentioned challenges.

Conceptual framework for tourism demand forecasting

For tourism demand forecasting, we propose the conceptual framework to construct the tourism demand forecasting model with the following four main steps.

The first step is for search engine platforms identification. During the travel planning process, tourists usually utilize search engines for information on various facets of the destination, such as selection of attractions, choices of clothing, and planning of transportation. Different source markets may have different search engine platforms, and thus, in this step, the set of search engine platforms needs to be identified. For example, Google is a dominant search engine in most English-speaking countries, while Baidu and Yandex are more popular in other source markets, such as China and Russia (Dergiades et al., 2018; Yang et al., 2015).

Data collection is the second step. The tourist arrival volume, typically noted as monthly data, needs to be obtained from reliable data providers. Various tourism demand forecasting factors could be collected from different resources depending on the availability of data. For secondary factors such as SII data, the data collection involves the following stages:

(1) We identify the seed search keywords for the destination market and maximize the possible set of search keywords to represent all aspects of tourists' interests on the destination through Google Trends related queries.
(2) For the set of potentially tourism related keywords, we use Google Translate to convert them into languages for other search engine platforms, such as Chinese for Baidu.
(3) For each specified search engine, we extract the monthly SII data corresponding to the keywords from the last stage. Google Trends provides monthly popularity data, while other search engines such as Baidu Index may only provide daily data, which need to be converted into monthly data. SII data of some keywords may not be available on all search engine platforms.

The collected training data usually contain hundreds of factors. Although the deep network architecture applied in the next step does not require feature selection, several obviously irrelevant factors that have low association with tourist arrival volume can be automatically removed. With regard to the linear limitation of Pearson correlation coefficients, the maximal information coefficient (MIC) can be utilized to pre-filter factors with minor associations. MIC focuses on the idea that if two features are related, then a grid that partitions the data can be drawn on their scatter plot to encapsulate the relationship between the two features. Given that MIC is general, $MIC - \rho^2$, where ρ is the Person correlation coefficient, can be adopted as a natural measure of nonlinearity. For a high MIC value, a large $MIC - \rho^2$ denotes a nonlinear relationship, and a small $MIC - \rho^2$ denotes a linear relationship (Reshef et al., 2011).

The third step is for deep learning model training. Considering that deep learning technique has the built-in mechanism of feature engineering, we propose a deep network architecture that can automatically select a set of influential factors and determine the lag order of time-series sequences.

The last step is for model interpretation. The trained deep learning model comprehensively represents the temporal relationship between various tourism demand forecasting factors and tourist arrival volume. The weights of the neuron links and the attention scores can be applied to determine which original factors have the most influential lag orders. Notably, no manual feature selection or extraction is required in this framework, and all feature engineering tasks are automatically performed by the deep learning model.

Deep network architecture

In this section, we articulate the task and introduce the deep network architecture with historical time-series tourism demand data. We then demonstrate the integration of the attention mechanism, which provides attention scores for various factors, into the LSTM recurrent neural network.

The objective of tourism demand forecasting is to predict tourist arrival volume according to multivariate factors from the past.

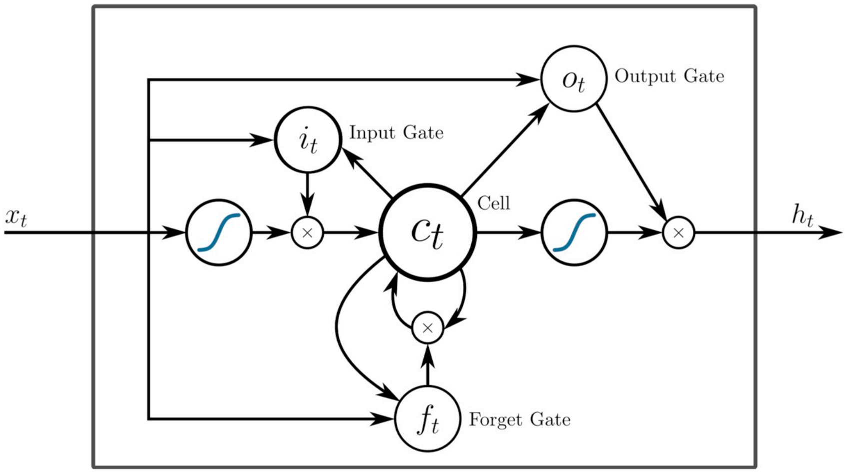

Fig. 2. LSTM architecture.

Formally, the input is represented as the fully observed feature vector set $\{x_t\}_{t=1}{}^T = x_1, x_2, \ldots, x_3$ and the corresponding tourist arrival volume $\{y_t\}_{t=1}{}^T = \{y_1, y_2, \ldots, y_T\}$. T is the length of total time steps, such as the number of months in the collected dataset. At time step t, y_t is the tourist arrival volume and x_t is typically the vector of multivariate factors, such as determinants, and SII data correspond to related keywords.

The tourism demand forecasting problem uses the time series of multivariate factors $\{x_t\}_{t=1}{}^T$ and the real tourist arrival volume $\{y_t\}_{t=1}{}^T$ as inputs and constructs a model F to forecast y at future time steps.

$$\{\hat{y}_t\}_{t=T+1}^{T+\Delta} = \mathcal{F}(\{x_t\}_{t=1}^{T}, \{y_t\}_{t=1}^{T})$$

This formulation is different from that in autoregressive models, which usually assume that $\{x_t\}_{t=T+1}{}^{T+\Delta}$ is available when predicting $\{\hat{y}_t\}_{t=T+1}^{T+\Delta}$ because they are designed to model the mapping between conditions and consequences (Qin et al., 2017).

Theoretically, RNN can handle long-term dependencies. However, in training, RNN is sensitive to vanishing and exploding gradients. The LSTM recurrent neural network can handle this issue by providing memory blocks in its recurrent connections. Each block contains a memory cell that stores the network temporal states and three gates, namely, remember, forget, and output. These gates control the information flow so that weak signals can be blocked. The architecture of LSTM is shown in Fig. 2.

With time series $\{x_t\}_{t=1}{}^T$ as an input, LSTM encodes it into a sequence of hidden states $\{h_t\}_{t=1}{}^T$. The main idea behind LSTM is that at each time step, a few gates are implemented to regulate the information passing along the sequences, thereby capturing any long-range dependencies accurately. To capture long-range dependencies, at each time step t in LSTM, hidden state h_t is updated by existing data at the same time step x_t, the hidden state at the previous time step h_{t-1}, input gate i_t, forget gate f_t, output gate o_t, and memory cell c_t (Zhao, Yan, Wang, & Mao, 2017). The following equations are used (Hochreiter & Schmidhuber, 1997):

$$\begin{aligned}
i_t &= \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i)\\
f_t &= \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f)\\
o_t &= \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o)\\
c_t &= f_t \times c_{t-1} + i_t \times \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)\\
h_t &= o_t \times \tanh(c_t)
\end{aligned}$$

where σ and tanh are recurrent activation functions, $\times$ denotes element-wise multiplication, W and b are the LSTM parameters learned during model training. The output in the final step is used to predict the output of a linear regression layer $\hat{y}_i = W_r h_i^T$, where W_r is the weight of the linear regression layer.

The attention mechanism has become an integral part of sequence modeling, and it allows the modeling of dependencies without regard to their distance in sequences (Vaswani et al., 2017). When the sequence proceeds to the output, it generates an attention range to highlight the part of the sequence that should receive much attention from LSTM. Thus, it selects a subset from a sequence of inputs by generating attention scores for each element of the sequence (Kim, Denton, Hoang, & Rush, 2017). Having an attention mechanism makes the model highly interpretable and allows the model to ignore irrelevant information. This feature is highly effective in tourism demand forecasting, in which the model learns to attend to factors that are related to tourism demand.

The proposed deep network architecture for tourism demand forecasting is shown in Fig. 3. The model is based on LSTM augmented with the attention mechanism. Identifying the lead or lag relationships between time series data is crucial in tourism demand forecasting because the influence of a feature differs with different lags. LSTM is utilized to model long-term dependencies in time-series data, and the attention mechanism is used to learn which subsets of the sequential units in the model are influential. This architecture allows us to capture two critical pieces of information in tourism demand forecasting, namely, (1) the temporal relationship between various factors and demand and (2) the importance of factors according to their impacts on tourism demand. Thus, the long-term temporal dependencies between various factors and tourist arrival volumes can be automatically detected by

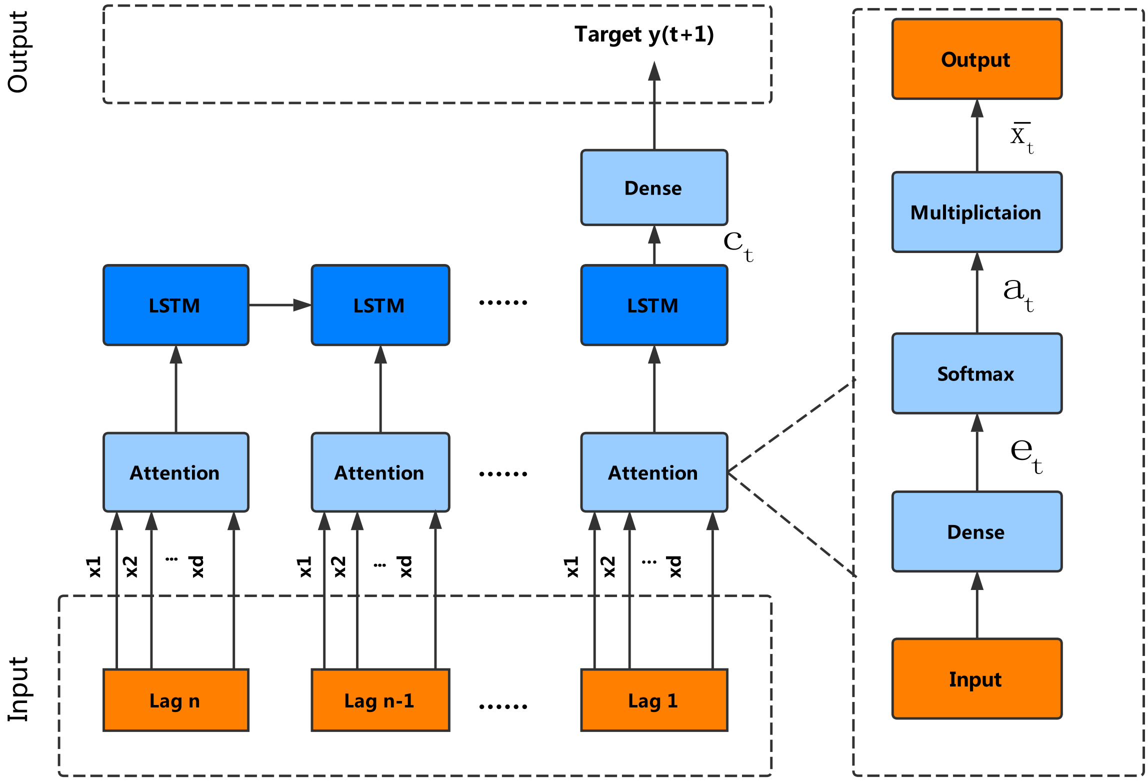

Fig. 3. A deep network architecture for tourism demand forecasting.

using LSTM with the attention mechanism.

The input is of size $m \times d \times n$, where $m \times d$ is the size of collected training data and n is the maximum lag order specified by the user. For each attention, a fully connected layer (dense) is constructed in accordance with the attention mechanism, which then selects the most relevant information in the driving series. The perception accepts the concatenation of h_{t-1} (hidden state in the last time step) and s_{t-1} (cell state in the last time step).

$$e_t = W_e^1 \tanh W_e^2 [h_{t-1}; s_{t-1}] + W_e^3 x_t,$$

where $W_e^{\ 1}$, $W_e^{\ 2}$, and $W_e^{\ 3}$ are the weights to be learned by the model. e_t is the vector of weights that measures the importance of features in the driving series at time $t(x_t)$, and a_t is the normalized e_t. Afterward, driving series x_t is multiplied by attention weight a_t: $\bar{x}_t = x_t \times a_t$. The LSTM component uses $\bar{x}_t$ and h_{t-1} as its input and updates the hidden state at time h_t. Context vector c_t is introduced by summing up the multiplication $c_t = \sum_{i=1}^{T} h_i w_i$. Then, the linear layer is formed to generate the final result

$$y_t = w_y^1 (W_y^2 c_t + b_1) + b_2.$$

LSTM and the perceptron can be trained simultaneously, and the model automatically focuses its attention on certain important features in the time series.

Model training

MAE is used as the loss function for model construction, and it is calculated based on the predicted results and actual targets. The deep learning parameters are updated by back propagation.

Given that the deep network architecture is highly complex, the scale of training data should be large to guarantee robust model performance. The Internet restricts the collection of large-scale training data. For example, we only have less than 100 monthly observations for the case study reported in the next section. Hence, the regularization technique is applied to the proposed model. Dropout is presented in the stage of model training. The process of dropout randomly masks several parts of the hidden outputs so that these neurons would not affect the forward propagation in the training process (Zhao et al., 2017). As soon as the dropout arrives in the testing phases, it is turned off, and the outputs of all hidden neurons exert their effects on the model testing. In other words, dropout helps considerably enlarge the size of the training data. The use of random masking in each training phase generates new variations in the data samples. In the current model, we adopt one dropout layer between LSTM and the first fully connected layer and another dropout layer between the first and second fully connected layers (Zhao et al., 2017). The masking probabilities are set to

0.8.

Empirical study

To empirically investigate the prediction performance of the proposed conceptual framework, we conduct an empirical study on the forecasting of monthly tourist arrival volume in Macau. Macau is an autonomous region of China, and it is across the Pearl River Delta from Hong Kong. Gaming and tourism make up the pillar industry of Macau, and they contribute remarkably to the economic growth of the city. Thus, maintaining a timely and accurate forecasting of tourist arrival volumes is essential to the prosperity of the economy. In this empirical study, due to the lack of expertise and reliable sources of determinants, Macau tourist arrival volumes are predicted based on the secondary indicators, namely, SII data, by using the proposed conceptual framework with the deep learning model.

Specification of search engines

According to Statistics and Census Services (DSEC) of the Macau government, the major source market of Macau's tourism industry is Mainland China, which contributes to more than 60% of tourist arrival volumes. The most popular search engine platform in China is Baidu, which accounts for 69% of the total market share (Yang et al., 2015). For the rest of the world, Google is the largest search engine, with more than 90% market share since 2010 (Kim et al., 2017). Both search engine platforms provide historical SII data. Specifically, Google Trends (https://trends.google.com) provides information on how frequently a certain keyword has been searched compared with the overall intensity over a certain period on a weekly or monthly basis. Baidu Index (https://zhishu.baidu.com) provides search intensity data on the keyword in the absolute number of volumes on a daily basis.

Data collection

Monthly Macau tourist arrival volumes are available from DSEC of the Macau government. In order to indicate the generalization of our algorithm, two types of tourist arrival volumes, corresponding to the global market and mainland China, are collected from DSEC website.[1] Given that the Baidu Index only has become available after January 2011, the range of our data is from January 2011 to August 2018 (92 observations). Figs. 4 and 5 show the Macau tourism arrival volumes from the global market and mainland China, respectively. It is clear that those volumes present similar cyclic fluctuations.

To capture SII data that are potentially related to Macau tourism, we start with several seed keywords in seven major categories, namely, dining, lodging, transportation, tour, clothing, shopping, and recreation. Table 1 lists the seed keywords used in this study.

Then, a set of potentially tourism-related keywords is obtained by using Google Trends' related queries section. After removing duplicates, a list of 211 related Google search keywords obtained, which are then translated into Chinese for Baidu via Google Translate service. Baidu does not provide search intensity data for a certain keyword if the keyword's volume is too low. After eliminating keywords without Baidu search intensity data, 45 Baidu keywords are obtained. A Python program is developed to collect monthly SII data from Google, “crawl” the daily intensity from Baidu, and summarize the data on a monthly basis. Finally, 92 monthly data (representing monthly SII data on both search engine platforms starting from January 2011 to August 2018) for those 256 keywords were obtained. For tourism demand forecasting, the training data are in the form of ${(x_t, y_t)_{t=1}}^T$, where $T = 92$ and x_t is a vector with 256 dimensions. The data set has been publicly released as "Macau2018" at: http://github.com/tulip-lab/open-data.

Performance evaluations

To investigate the performance of the proposed deep learning model (DLM) in tourism demand forecasting, we use a naive method, a support vector regression (SVR) model (Zhang et al., 2017), an artificial neural network (ANN) (Law, 2000), an ARIMAX model (Box, Jenkins, Reinsel, & Ljung, 2015) and an ARIMA model (Goh & Law, 2002) as baseline models. The naive method uses the tourist arrival volume dated 12 months back as the estimate of y_{k+1}, namely, $\hat{y}_{k+1} = y_{k-11}$. The SVR and ANN models use the past 12 months' data ${(x_t, y_t)_{t=k-11}}^k$ as the input and to predict $\hat{y}_{k+1}$. The ANN model has one hidden layer with the sigmoid activation function and is trained with the back-propagation algorithm. Both ARIMA and ARIMAX are using AR order of (p, d) times of difference to get stationary series, and the MA order of q to train the past 12-month tourist data and do the prediction series for the next 12-month tourist arrival volumes which is incrementally increased during the walk through validation. The difference between them is that ARIMA only uses the tourist arrival volume for prediction while ARIMAX also includes exogenous variables ${x_{t=1}}^T$.MIC is adopted to filter out features with low association with tourist arrival volumes because SVR and ANN are ineffective in handling data with large numbers of features.

To mimic the real-world scenario where new tourist arrival observations become available each month and are used in forecasting of the following month, *walk-forward* model validation is used. In each step, the training data “walk” by one month, and the forecasting model is trained and makes a forecast for the next month. Then, the actual demand value for the next month is obtained from the test set and made available to the forecasting model for the next month.

In this study, three measures of forecasting accuracy are calculated using the acquired predicted values. They are RMSE, MAE, and

[1] https://www.dsec.gov.mo/Statistic.aspx?NodeGuid=251baebb-6e5b-4452-8ad1-7768eafc99ed.

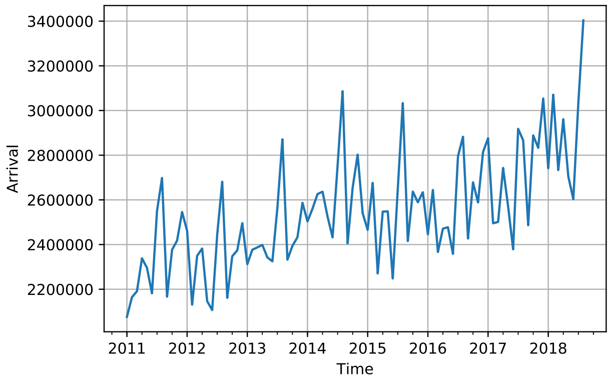

Fig. 4. Monthly Macau Tourist Arrival Volumes (Global Market).

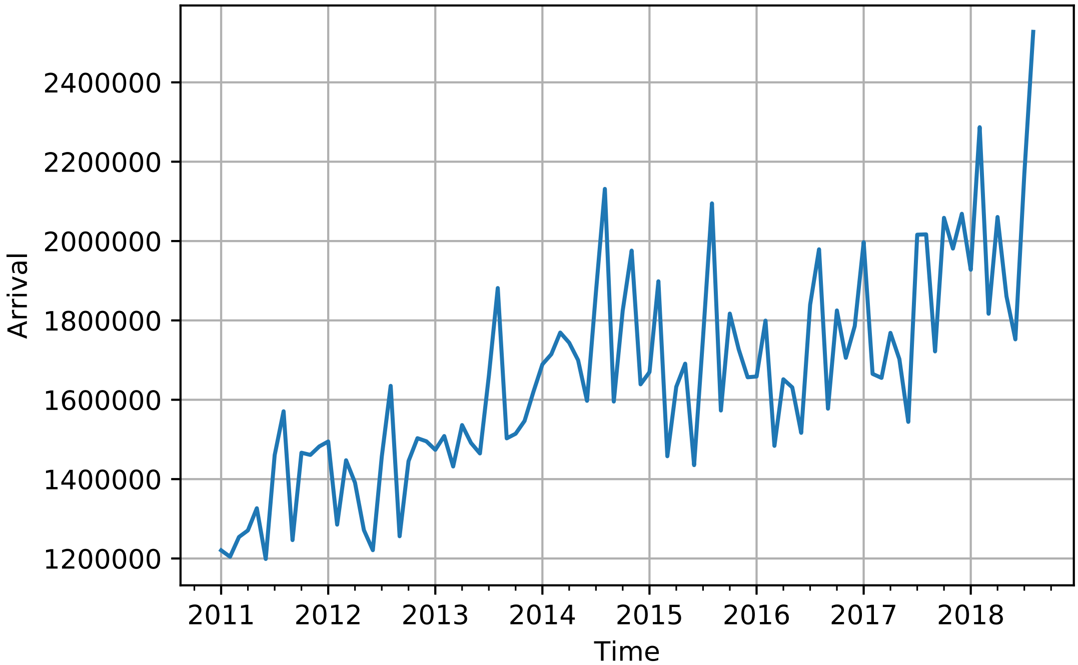

Fig. 5. Monthly Macau Tourist Arrival Volumes (Mainland China).

Table 1
Seed Search Keywords for Macau Tourism.

Category	Seed keywords
Dining	Macau food, Macau restaurant
Lodging	Macau hotel, Macau accommodation
Transportation	Macau ferry, Macau flights
Tour	Macau travel, Macau map, Macau travel agency, Macau tourism
Clothing	Macau weather
Shopping	Macau shopping, Macau shopping mall
Recreation	Macau bar, Macau show, Macau night life, Macau casino

Table 2
MAPE comparison for 5 years forecast (Global).

Year	DLM	SVR + F.E.	ANN + F.E.	ANN	SVR	ARIMA	ARIMAX	Naive
2013	1.186	3.996	6.027	15.516	5.164	7.31	8.10	5.278
2014	0.922	7.342	5.578	7.663	8.716	5.30	6.19	7.187
2015	1.132	4.994	7.301	5.092	5.725	6.30	7.35	4.388
2016	2.425	4.463	4.882	31.050	4.392	6.92	6.89	2.966
2017	1.671	4.638	5.827	12.219	8.433	6.51	7.12	5.964
Mean	1.467	5.086	5.923	14.307	6.482	6.468	7.13	5.156

MAPE, which are defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i}$$

The model with the lowest values in these measures is considered the best forecasting model. To ensure the robustness of the proposed deep learning tourism demand forecasting model, we repeat the *walk-forward* validation for all compared algorithms by five times to forecast 12 monthly tourist arrivals from 2013 to 2017 (five-year period).

Notably, the compared baseline models, SVR and ANN, do not have features crafted by domain experts, and their input features are simply automated from MIC filtering. In the experiment, we further implemented both models using the feature engineering results identified by deep learning model, and these two models are labelled as SVR + F.E. and ANN + F.E., respectively.

For the forecasting of the global tourism arrival volumes, MAPE, MAE and RMSE of all compared algorithms in these five years are summarized in Tables 2, 3, and 4. As indicated by the results, the deep learning model achieves minimal errors on all these measurements for the five consecutive years compared with the benchmark models. The forecast accuracy of MAPE decreases from 5.156% to 1.467%, which is a significant reduction.

Table 5 presents the results of one-tailed *t*-test results for MAPE, MAE and RMSE of DLM compared with the baseline models. The t-test results are presented with $\alpha = 0.05$. The null hypothesis, which states that the mean of DLM equals the mean of the compared model, is rejected. Thus, DLM achieves a considerably lower MAPE, MAE and RMSE than the baseline models.

Notably, the results from SVR and ANN without human effort in identifying the influential factors did not outperform the naïve method. On the contrary, the proposed DLM does not require any hand-crafted features but still produces significantly better forecasts than the other models. Moreover, using those features identified by DLM, both SVR and ANN have improved in their performance, with the MAPE reduced from 6.482% to 5.086%, and from 14.307% to 5.923%, respectively. This further confirms the capability of the proposed approach in feature engineering.

When comparing DLM with ARIMA and ARIMAX, though all methods do not require any hand-crafted features. The DLM can generate the effective relevant factors automatically from the raw SII data, which outperform the ARIMA/ARIMAX.

Similarly, we completed the comparison using only the Baidu keywords and Macau tourist arrival from mainland China. MAPE, MAE and RMSE for all compared algorithms in these five years are summarized in Tables 6, 7 and 8 respectively. We can see the same significant performance boosting in all three tables, for example, MAPE decreases from 7%–11% down to 1.95%.

Table 9 presents the t-test results for DLM compared with the baseline models. From those results, it further confirms the good performance of the deep learning algorithm.

Model interpretation

The attention mechanism in deep learning can automatically identify the factors that contribute considerably to tourism demand

Table 3
MAE comparison for 5 years forecast (Global).

Year	DLM	SVR + F.E.	ANN + F.E.	ANN	SVR	ARIMA	ARIMAX	Naive
2013	27,911	105,379	154,665	323,342	120,669	172,118	183,911	128,430
2014	22,351	201,483	154,195	222,612	210,313	118,140	131,011	191,194
2015	29,572	123,687	178,772	128,181	148,684	141,281	167,811	108,088
2016	64,596	115,744	122,547	607,327	113,710	181,659	197,721	77,968
2017	42,556	128,305	163,969	295,976	217,232	193,382	217,991	165,889
Mean	37,397	134,919	154,829	315,487	162,121	161,316	179,689	134,313

Table 4
RMSE comparison for 5 years forecast (Global).

Year	DLM	SVR + F.E.	ANN + F.E.	ANN	SVR	ARIMA	ARIMAX	Naive
2013	35,976	179,665	211,317	370,106	184,262	199,316	210,219	148,386
2014	27,100	268,097	210,417	278,145	277,135	163,254	178,677	208,607
2015	40,020	158,771	219,881	168,442	210,727	187,814	221,197	144,823
2016	88,802	154,336	134,568	633,004	135,049	234,938	239,899	96,595
2017	50,604	166,033	190,933	351,529	251,394	214,742	252,189	201,699
Mean	48,500	185,380	193,423	360,245	211,713	200,187	220,436	160,022

Table 5
One-tailed t-test results (Global).

Comparison	MAPE		MAE		RMSE	
	p-Value	t-Interval	p-Value	t-Interval	p-Value	t-Interval
DLM vs SVR + F.E.	0.0044	[−5.22, ∞]	0.0053	[−144,257, ∞]	0.0045	[−195,007, ∞]
DLM vs ANN + F.E.	0.0005	[−5.58, ∞]	0.0008	[−145,111, ∞]	0.0025	[−189,283, ∞]
DLM vs ANN	0.0102	[−25.47, ∞]	0.0099	[−500,806, ∞]	0.005	[−524,011, ∞]
DLM vs SVR	0.0021	[−7.42, ∞]	0.0042	[−184,529, ∞]	0.0042	[−231,497, ∞]
DLM vs ARIMA	0.00002	[−6.00, ∞]	0.0001	[−162,190, ∞]	0.0002	[−153,647, ∞]
DLM vs ARIMAX	0.00002	[−6.61, ∞]	0.0001	[−182,772, ∞]	0.0002	[−173,870, ∞]
DLM vs Naive	0.0041	[−5.63, ∞]	0.0098	[−151,932, ∞]	0.0095	[−168,317, ∞]

Table 6
MAPE comparison for 5 years forecast (Mainland China).

Year	DLM	ANN + F.E.	SVR + F.E.	ANN	SVR	ARIMA	ARIMAX	Naive
2013	3.194	11.42	8.12	23.6	10.43	7.34	9.182	9.58
2014	1.844	9.35	6.85	12.64	14.2	8.49	11.173	12.08
2015	1.532	7.88	7.14	9.78	8.24	12.68	10.912	6.37
2016	1.657	5.17	4.58	6.38	6.43	8.24	9.972	3.13
2017	1.549	6.32	4.55	5.87	8.49	8.99	8.151	8.82
Mean	1.9552	8.02	6.248	11.654	9.538	9.14	9.878	7.996

Table 7
MAE comparison for 5 years forecast (Mainland China).

Year	DLM	ANN + F.E.	SVR + F.E.	ANN	SVR	ARIMA	ARIMAX	Naive
2013	49,927	212,808	159,343	369,842	168,702	130,664	157,221	150,283
2014	32,174	175,702	120,589	233,452	263,028	165,568	198,820	218,350
2015	27,319	136,844	118,494	160,917	136,979	203,604	170,102	103,770
2016	31,561	95,834	77,343	109,241	110,641	136,979	163,312	54,773
2017	28,958	109,842	75,581	109,874	165,568	157,298	156,982	167,558
Mean	33,987	146,206	110,270	196,665	168,983	158,822	169,287	138,946

Table 8
RMSE comparison for 5 years forecast (Mainland China).

Year	DLM	ANN + F.E.	SVR + F.E.	ANN	SVR	ARIMA	ARIMAX	Naive
2013	64,029	260,368	206,153	391,121	202,332	162,523	172,218	174,360
2014	41,161	231,644	140,911	282,370	301,997	205,286	220,010	242,180
2015	29,877	168,888	151,186	183,486	171,182	230,795	189,910	143,599
2016	35,702	117,528	127,488	129,448	132,807	171,182	190,205	69,854
2017	31,465	134,955	113,514	134,216	205,286	190,461	220,103	192,676
Mean	40,446	182,676	147,850	224,128	202,720	192,049	209,429	164,533

forecasting. We train the proposed deep network architecture by using all 92 monthly Macau observations from January 2011 to August 2018, and then identify the raw features assigned with high attention scores in the trained model.

In order to interpret the attention score, the attention scores from 92 months are utilized by sum and then mean calculation. After the calculation, 10 search keywords for Google and nine search keywords for Baidu are identified, and the attention scores for their

Table 9
One-tailed t-test results (Mainland China).

Comparison	MAPE		MAE		RMSE	
	p-Value	t-Interval	p-Value	t-Interval	p-Value	t-Interval
DLM vs SVR + F.E.	0.0007	[−6.25, ∞]	0.0019	[−119,024, ∞]	0.0003	[−150,780, ∞]
DLM vs ANN + F.E.	0.0015	[−9.09, ∞]	0.0018	[−171,428, ∞]	0.0016	[−217,787, ∞]
DLM vs ANN	0.0148	[−18.6, ∞]	0.0114	[−298,144, ∞]	0.0072	[−321,905, ∞]
DLM vs SVR	0.0019	[−11.23, ∞]	0.0031	[−206,099, ∞]	0.0021	[−239,335, ∞]
DLM vs ARIMA	0.0015	[−9.71, ∞]	0.0007	[−160,046, ∞]	0.0004	[−185,228, ∞]
DLM vs ARIMAX	0.0002	[−9.47, ∞]	0.00007	[−156,715, ∞]	0.0002	[−184,707, ∞]
DLM vs Naive	0.0068	[−10.21, ∞]	0.0095	[−182,011, ∞]	0.0057	[−202,564, ∞]

Table 10
Attention scores.

Services	Keywords	1	2	3	4	5	6	7	8	9	10	11	12
Google	Restaurant in Macau	0.91				0.58		0.71					
	Macau Tauranga	0.95			0.25	0.64		0.78					
	Macau airlines	0.75	0.42		0.31		0.29						
	travel to Macau	0.76				0.60		0.56					
	best hotel Macau	0.48		0.34	0.21								
	Macau ferry terminal	1.07			0.68			0.85					
	weather in Macau	0.84			0.51	0.64							
	casino in Macau	0.54				0.29		0.35					
	ferry to Macau	1.17			0.52	0.75							
	shopping in Macau	0.50			0.30	0.34							
Baidu	澳门住宿攻略	0.78				0.65		0.54					
	澳门美食	0.95				0.55		0.70					
	澳门旅游景点大全	0.80				0.53		0.42					
	澳门航空	1.09				0.74		0.91					
	澳门酒店	0.77				0.63		0.52					
	澳门地图	0.41			0.31		0.26						
	澳门天气	0.89			0.59	0.73							
	香港到澳门	0.68		0.61				0.51					
	澳门赌场	0.15		0.06	0.08								

12 lags are provided in Table 10. It should be noted that we do not include any seed keyword in Chinese, but the proposed approach successfully identify those influential Chinese keywords.

We obtain the following observations from the table. First, all identified influential keywords have 1–7 months of lags, and none of them is useful for the prediction of travel demand in 8–12 months. Second, all identified influential keywords with one-month lag are highly useful for the incoming month's travel demand. This result may indicate that many visitors to Macau plan their trips within a short time period, and this applies to both Google and Baidu users. Third, many keywords with 4–7 months of lag are identified as influential, and this result implies that another group of visitors to Macau would plan their trip around six months in advance.

The time lags findings of influential keywords actually match the results of a large scale visitor profile survey conducted in Macau which is an ongoing and systematical collection of comprehensive data on the characteristics of visitors since 2011 (Fong, 2017). The survey finds that 47.8% of the visitors planned to visit Macau on the same day of arrival, 28.7% of the visitors planned to visit Macau within one month and 23.5% of the visitors took more than one month to plan for their trips in 2017.

By examining the keywords with attention scores larger than 1.0, we find that the Google keywords “ferry to Macau” and “Macau ferry terminal” and the Baidu keyword “香港到澳门” (Hong Kong to Macau) are related to transportation. This observation may indicate that many visitors go to Macau from Hong Kong, and many Chinese visitors directly fly to Macau from Mainland China. In the same study, 12% of the tourists visited Hong Kong before they visited Macau and 9.6% of the tourists visited Hong Kong after they had visited Macau in 2017 (Fong, 2017).

Focusing on other keywords with attention scores larger than 0.90, we find that the identified keywords “restaurant in Macau,” and “澳门美食” (Macau food) are related to food and dining. This result indicates that Macau is a fine-dining destination, and it confirms the uniqueness of Macau's cuisine, which combines Western and Chinese cuisines (Song & Witt, 2006). Actually, Macao was designated as a new member city of UNESCO Creative Cities Network (UCCN, https://www.gov.mo/en/news/88728/) in the field of Gastronomy on 31 October 2017, making it the third city in China to join UCCN.

The next group of keywords with high attention scores pertains to weather: “weather in Macau” on Google and “澳门天气” (Macau weather) on Baidu. Notably, gaming-related keywords, such as “casino in Macau” and “澳门赌场” (Macau casinos), are influential but not as influential as the cuisine-related keywords. On one hand, this result might reflect that most of the revisit tourists are loyal to the casinos which they perceived to bring them good luck. On the other hand, some websites with the word of “casino” are blocked in mainland China so that it may not reflect the attention scores relatively.

In addition, the accommodation-related keywords have low attention scores for Google or Baidu users. This result might reflect the fact that many visitors to Macau are not overnight visitors. As reflected by the identified attention scores, tourists search for relevant travel-related information and make different travel decisions at different phases of the travel planning process. This result validates the model interpretation property of the proposed conceptual framework.

Conclusions

In the tourism industry, precise and timely demand forecasts are critical for informed decision making by most, if not all, providers of products and services. Time-series, econometrics and AI models have been extensively examined in the past decades. Traditionally, the accuracy of tourism demand forecasting models rely on the goodness of the set of features. Poor selection of determinants or indicators often leads to lower accuracy compared with that of good selection. The selection of influential factors and the determination of their lag order are domain specific and require extensive human effort. At present, only marginal improvements can be attained despite the considerable effort in advancing traditional approaches (Coshall & Charlesworth, 2011). Therefore, alternative strategies are required to further improve the capability of tourism demand forecasting models.

By addressing two practical barriers in most tourism demand forecasting models, this work transfers the task of feature engineering from researchers to the model itself. The proposed conceptual framework utilizes the deep learning technique to identify and extract discriminative features with minimal human effort. The comparison with baseline models shows that the deep network architecture performs much better in all three accuracy measures. The success of the deep learning approach may be partially attributed to two reasons. First, the deep network architecture mimics how the human brain operates. The successive layers of the network extract low-level features from the initial input layer and further abstract high-level features that represent the semantic relationships between features in subsequent layers. Second, the attention mechanism integrated into LSTM automatically identifies a set of influential features at each time step.

Compared with previous studies on tourism demand forecasting, our research makes two contributions. The first contribution is that we propose a systemic conceptual framework of tourism demand forecasting and validate the deep learning's capability in tourism demand forecasting. The proposed framework fully utilizes all available tourism demand forecasting factors and reduces the human effort required in feature engineering. Another contribution is that we utilize the attention score to interpret the trained deep network architecture. This usage provides tourism industry practitioners a novel method to update their tourism demand forecasts promptly on the basis of a set of influential indicators at different time steps. For example, a surge in the SII data of a particular keyword may imply an increase in tourist arrivals several months later.

Moreover, the results of this work confirm the capability of deep learning in selecting a set of influential factors and determining their suitable lag orders. This confirmation encourages two future extensions of this work. First, various types of indicators other than SII data, such as Tweets and Blogs, can be comfortably incorporated into forecast tourist demand. Large-scale social media data could potentially alleviate the challenges in training data availability, and the deep learning technique allows these media data to be utilized automatically. Moreover, sets of features with suitable lag orders can be used as an input to other tourism demand forecasting models. The combined power of deep learning and existing forecasting models could facilitate further theory development.

This study has its limitations. The primary one is that because of data availability, we only utilize SII data, the secondary factors, as input features, and other superior but untapped determinants are not included in the empirical study. Furthermore, the generalization of the identified search keywords with lag orders is limited, and different source markets may have different sets of features. Additional studies on the deep learning with data from other destinations and empirical studies with different search engine platforms are required to overcome these limitations.

Acknowledgement

The work described in this paper was partially supported by the National Natural Science Foundation of China under Grant No. 71471011, International Cooperation Project of Institute of Information Engineering, Chinese Academy of Sciences under Grant No. Y7Z0511101, and research grants funded by the Hong Kong Polytechnic University (G-YBXG, G-UAE8).

References

Akin, M. (2015). A novel approach to model selection in tourism demand modeling. *Tourism Management, 48*, 64–72.
Alvarez-Diaz, M., Mateu-Sbert, J., & Rossello-Nadal, J. (2009). Forecasting tourist arrivals to Balearic Islands using genetic programming. *International Journal of Computational Economics and Econometrics, 1*(1), 64–75.
Andrew, W. P., Cranage, D. A., & Lee, C. K. (1990). Forecasting hotel occupancy rates with time series models: An empirical analysis. *Hospitality Research Journal, 14*(2), 173–182.
Athanasopoulos, G., & Hyndman, R. J. (2008). Modelling and forecasting Australian domestic tourism. *Tourism Management, 29*(1), 19–31.
Baldi, P., & Pollastri, G. (2003). The Principled design of large-scale recursive neural network architectures-DAG-RNNs and the protein structure prediction problem. *Journal of Machine Learning Research, 4*, 575–602.
Bangwayo-Skeete, P. F., & Skeete, R. W. (2015). Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management, 46*, 454–464.
Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLoS One, 12*(7), e0180944.
Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control.* Hoboken, NJ: John Wiley & Sons.
Chan, C. K., Witt, S. F., Lee, Y. C. E., & Song, H. (2010). Tourism forecast combination using the CUSUM technique. *Tourism Management, 31*(6), 891–897.
Chan, Y.-M., Hui, T.-K., & Yuen, E. (1999). Modeling the impact of sudden environmental changes on visitor arrival forecasts: The case of the Gulf War. *Journal of*

Travel Research, 37(4), 391–394.
Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1724–1734). Doha, Qatar: Association for Computational Linguistics.
Choi, H., & Varian, H. (2012). Predicting the present with Google trends. *The Economic Record, 88*(s1), 2–9.
Coshall, J. T., & Charlesworth, R. (2011). A management orientated approach to combination forecasting of tourism demand. *Tourism Management, 32*(4), 759–769.
Dergiades, T., Mavragani, E., & Pan, B. (2018). Google Trends and tourists' arrivals: Emerging biases and proposed corrections. *Tourism Management, 66*, 108–120.
Fesenmaier, D. R., Xiang, Z., Pan, B., & Law, R. (2011). A framework of search engine use for travel planning. *Journal of Travel Research, 50*(6), 587–601.
Fong, K. C. (2017). *Macau visitor profile study 2017.* Institute for the Study of Commercial Gaming, University of Macau.
Frechtling, D. C., & Frechtling, D. C. (2001). *Forecasting tourism demand: methods and strategies.* Oxford, Boston: Butterworth-Heinemann (OCLC: ocm48401821).
Goh, C., & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management, 23*(5), 499–510.
Goh, C., & Law, R. (2003). Incorporating the rough sets theory into travel demand analysis. *Tourism Management, 24*(5), 511–517.
Goh, C., Law, R., & Mok, H. M. K. (2008). Analyzing and forecasting tourism demand: A rough sets approach. *Journal of Travel Research, 46*(3), 327–338.
Guizzardi, A., & Stacchini, A. (2015). Real-time forecasting regional tourism with business sentiment surveys. *Tourism Management, 47*, 213–223.
Gunter, U., & Onder, I. (2015). Forecasting international city tourism demand for Paris: Accuracy of uni- and multivariate models employing monthly data. *Tourism Management, 46*, 123–135.
Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research, 3*, 1157–1182.
Hinton, G., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural Computation, 18*(7), 1527–1554.
Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.
Kim, Y., Denton, C., Hoang, L., & Rush, A. M. (2017). Structured attention networks. arXiv preprint:1702.00887 [cs] https://arxiv.org/abs/1702.00887.
Law, R. (2000). Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tourism Management, 21*(4), 331–340.
Law, R., & Au, N. (1999). A neural network model to forecast Japanese demand for travel to Hong Kong. *Tourism Management, 20*(1), 89–97.
Le Cun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.
Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. *Tourism Management, 59*, 57–66.
Lim, C. (1997). Review of international tourism demand models. *Annals of Tourism Research, 24*(4), 835–849.
Martins, L. F., Gan, Y., & Ferreira-Lopes, A. (2017). An empirical analysis of the influence of macroeconomic determinants on World tourism demand. *Tourism Management, 61*, 248–260.
Meleddu, M., & Pulina, M. (2016). Evaluation of individuals' intention to pay a premium price for ecotourism: An exploratory study. *Journal of Behavioral and Experimental Economics, 65*, 67–78.
Moutinho, L., & Witt, S. F. (1995). Forecasting the tourism environment using a consensus approach. *Journal of Travel Research, 33*(4), 46–50.
Önder, I., & Gunter, U. (2016a). Forecasting city arrivals with Google Analytics. *Annals of Tourism Research, 61*, 199–212.
Önder, I., & Gunter, U. (2016b). Forecasting tourism demand with Google Trends for a major European city destination. *Tourism Analysis: An Interdisciplinary Journal, 21*, 203–220.
Pai, P.-F., Hong, W.-C., Chang, P.-T., & Chen, C.-T. (2006). The application of support vector machines to forecast tourist arrivals in Barbados: An empirical study. *International Journal of Management, 23*(2), 375.
Pan, B., & Yang, Y. (2017). Forecasting destination weekly hotel occupancy with big data. *Journal of Travel Research, 56*(7), 957–970.
Poprawe, M. (2015). A panel data analysis of the effect of corruption on tourism. *Applied Economics, 47*(23), 2399–2412.
Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., ... Iyengar, S. S. (2018). A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys, 51*(5), 92 (1-92:36).
Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., & Cottrell, G. (2017). A dual-stage attention-based recurrent neural network for time series prediction. arXiv:1704.02971 [cs, stat]. https://arxiv.org/abs/1704.02971.
Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., ... Sabeti, P. C. (2011). Detecting novel associations in large data sets. *Science, 334*(6062), 1518–1524.
Rivera, R. (2016). A dynamic linear model to forecast hotel registrations in Puerto Rico using Google Trends data. *Tourism Management, 57*, 12–20.
Saha, S., & Yap, G. (2015). Corruption and tourism: An empirical investigation in a non-linear framework. *International Journal of Tourism Research, 17*(3), 272–281.
Shen, S., Li, G., & Song, H. (2011). Combination forecasts of International tourism demand. *Annals of Tourism Research, 38*(1), 72–89.
Socher, R., Bengio, Y., & Manning, C. D. (2012). Deep learning for NLP (without magic). *Tutorial abstracts of ACL 2012, ACL '12* (pp. 5). Stroudsburg, PA: Association for Computational Linguistics.
Song, H., & Li, G. (2008). Tourism demand modelling and forecasting. A review of recent research. *Tourism Management, 29*(2), 203–220.
Song, H., & Witt, S. F. (2006). Forecasting international tourist flows to Macau. *Tourism Management, 27*(2), 214–224.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Vol. Eds.), *Advances in neural information processing systems. Vol. 30. Advances in neural information processing systems* (pp. 5998–6008). Red Hook, NY: Curran Associates, Inc.
Volchek, K., Liu, A., Song, H., & Buhalis, D. (2018). Forecasting tourist arrivals at attractions: Search engine empowered methodologies. *Tourism Economics*. https://doi.org/10.1177/1354816618811558.
Witt, S. F., & Witt, C. A. (1995). Forecasting tourism demand: A review of empirical research. *International Journal of Forecasting, 11*(3), 447–475.
Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. *Neural Computation, 8*(7), 1341–1390.
Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation, 1*(1), 67–82.
Wu, D. C., Song, H., & Shen, S. (2017). New developments in tourism and hotel demand modeling and forecasting. *International Journal of Contemporary Hospitality Management, 29*(1), 507–529.
Xiang, Z., & Pan, B. (2011). Travel queries on cities in the United States: Implications for search engine marketing for tourist destinations. *Tourism Management, 32*(1), 88–97.
Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management, 46*, 386–397.
Yang, Y., Pan, B., & Song, H. (2014). Predicting hotel demand using destination marketing organization's web traffic data. *Journal of Travel Research, 53*(4), 433–447.
Zhang, B., Huang, X., Li, N., & Law, R. (2017). A novel hybrid model for tourist volume forecasting incorporating search engine data. *Asia Pacific Journal of Tourism Research, 22*(3), 245–254.
Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied Artificial Intelligence, 17*(5–6), 375–381.
Zhao, R., Yan, R., Wang, J., & Mao, K. (2017). Learning to monitor machine health with convolutional bi-directional LSTM networks. *Sensors, 17*(273), 1–18.
Zhou, Z.-H. (2012). *Ensemble methods: Foundations and algorithms. Chapman & Hall/CRC machine learning & pattern recognition series*Boca Raton, FL: Taylor & Francis.

Rob Law, rob.law@polyu.edu.hk, Ph.D. His research interests are technology applications to tourism and information management.

Gang Li, gang.li@deakin.edu.au, Ph.D. His research interests are technology applications to tourism and hospitality, data science, and business intelligence.

Davis Ka Chio Fong, davisfong@umac.mo, Ph.D. His research interests are social and economic impacts of gambling, responsible gambling, and tourist behaviour.

Xin Han, xin.han@tulip.org.au, Master by research student, his research interests are tourism and hospitality management, and data science.