Deep learning approach and topic modelling for forecasting tourist arrivals

Original Scientific Paper

Houria Laaroussi
Mohammed V University in Rabat,
Mohammadia School of Engineering, EST SALE, Department of Engineering Science
RABAT, Morocco
hourialaaroussi@research.emi.ac.ma

Fatima Guerouate
Mohammed V University in Rabat,
Mohammadia School of Engineering, EST SALE, Department of Engineering Science
RABAT, Morocco
fatima.guerouate@est.um5.ac.ma

Mohamed Sbihi
Mohammed V University in Rabat,
Mohammadia School of Engineering, EST SALE, Department of Engineering Science
RABAT, Morocco
mohamed.sbihi@yahoo.fr

Abstract *– Online review data attracts the attention of researchers and practitioners in various fields, but its application in tourism is still limited. The social media data can finely reflect tourist arrivals forecasting. Accurate prediction of tourist arrivals is essential for tourism decision-makers. Although current studies have exploited deep learning and internet data (especially search engine data) to anticipate tourism demand more precisely, few have examined the viability of using social media data and deep learning algorithms to predict tourism demand. This study aims to find the key topics extracted from online reviews and integrate them into the deep learning model to forecast tourism demand. We present a novel forecasting model based on TripAdvisor reviews. Latent topics and their associated keywords are captured from reviews through Latent Dirichlet Allocation (LDA), These generated features are then employed as an additional feature into the deep learning (DL) algorithm to forecast the monthly tourist arrivals to Hong Kong from USA. We used machine learning models, artificial neural networks (ANNs), support vector regression (SVR), and random forest (RF) as benchmark models. The empirical results show that the proposed forecasting model is more accurate than other models, which rely only on historical data. Furthermore, our findings indicate that integration of the topics extracted from social media reviews can enhance the prediction.*

Keywords: *Online review data, Tourist arrivals forecasting , Latent Dirichlet Allocation, Deep Learning*

1. INTRODUCTION

The tourism sector has expanded greatly over the last several years. Tourism demand forecasting has become a striking issue in the area of prediction research because of the important economic impact of the rapid-growing tourism industry [1]. Forecasting is vital to the tourism planning process. Because of the perishable nature of tourism [2]. Tourist arrivals forecasting provides valuable information to decision makers in order to make crucial decisions and planning [3]. Accurate forecasts can serve to allocate resources, and contribute to reducing the risks of decision failures and the costs of attracting [4].

Overall, tourism demand forecasting techniques can be categorized into three classes: time series models, econometric models, and artificial learning techniques [5,6]. Econometric and time series models fail to learn the nonlinear dependencies in the data. Machine learning techniques usually need to have manual features [7]. Deep learning (DL) is a neural network with many layers. DL methodologies have gained the interest of scholars due to their successful applications. Many hidden layers can often capture the non-linear characteristics of data, it can handle the nonlinearity that exists in tourism demand data and it aims to build important features automatically [8].

In earlier studies, the most crucial data were historical data and search engine data [8,9]. On the other hand, social media data can make a better contribution to predicting tourism demand. Therefore, Social media data can supplement traditional data. Furthermore, search engine data can enhance the prediction performance of tourism demand [9], but it is not easy for users to choose the most relevant information in search engines with a such large number of information available over search engines and it cannot completely describe tourists' preferences. Then, to increase forecasting accuracy, incomplete data must be reinforced with relevant information. Social media can be used as a solution to address this problem.

The latest approaches for forecasting tourism demand are seen to have two significant limitations. First, the majority of studies employed deep learning methods that perform well in tourism demand prediction, but they didn't take into account the influence of social media data [5,8,10,11]. In the deep learning techniques, they only utilized the motor engine data or traditional data. For instance, [8] employed a deep learning approach and search engine data to predict the monthly Macau tourist arrival volumes. [10] used a deep learning model called the Bayesian Bidirectional Long Short-Term Memory (BBiLSTM) network in conjunction with past tourist volumes, tourism prices, and income to forecast quarterly tourist arrivals to Singapore. [11] used LSTM networks with multisource time series data to predict daily tourism demand. [5] Incorporate multivariate time series data into a deep learning model to forecast the daily tourism volume. This study tries to incorporate social media data into the deep learning model.

Second, the studies that take social media into account focus on quantitative data. For example, [12] utilized likes on DMO Facebook pages as a novel explanatory variable of tourism demand. [13] Employed the average review rating and the volume of the review as indicators from online review data to forecast tourist arrivals. [14] Used review volume and ratings collected from internet review sources to forecast tourism demand. .There are few studies that use qualitative data in the prediction of tourism demand. For example, [15] adopted news topics to forecasting tourism demand. In this paper, we seek to fill this gap. We discuss the impact of textual data as qualitative data that can be obtained from social media to enhance the prediction of tourism demand. We use Topic modeling to discover coherent and interesting topics of reviews extracted from social media.

The purpose of this paper is to develop a new model using social media data and strong features for tourism demand prediction. In particular, we predict tourist arrivals to Hong Kong from USA in the period from January 2013 to January 2020. First, we collected the reviews shared on Trip Advisor. Then, we extract the important topics and the key keywords from posts using a topic modeling technique, latent Dirichlet allocation (LDA), which focuses on identifying topics within a collection of reviews. It can present the distribution of keywords by topic and the distribution of topics by review. All the keywords are used as new features in the prediction model. Furthermore, we predict tourist arrivals in Hon Kong from USA based on the long short-term memory (LSTM) algorithm. To do this, text mining frequency inverse document frequency (TF-IDF) is applied for vector representation. TF-IDF converts each keyword into numeric vectors to construct an LSTM model to predict tourism demand. In addition, an empirical analysis was performed to measure the accuracy of our model. Random forest, support vector regression, and artificial neural network are used for experiments.

This article will present the related work in Section 2. We will describe the proposed methods in section 3. Section 4 will provide the results and discussion. In Section 5, we will present the conclusions of the proposed study and future work.

2. RELATED WORK

2.1. TOURISM DEMAND FORECASTING MODELS

Several researchers have used Time series, econometric, and AI methods for predicting tourism demand [16]. Time series methods include naïve, moving average, exponential smoothing, and BoxeJenkins models. The autoregressive moving average (ARIMA) and Seasonal ARIMA are widely employed models and give better performance [17]. Time series models are limited by their incapacity to forecast fluctuations that are not used the past observations [18].

Causal econometric models seek to analyze the causal relationships between tourist arrivals and their determinants (such as Exchange rates, income; relative prices, and expenditure [19]. The insufficient information on the causal structure is the main limitation of econometric models [20]. Popular econometric forecasting models comprise the autoregressive distributed lag model [19], the error correction model [21], the vector autoregressive model [22], and the time-varying parameter model [23]. The most important limitations of econometric models are the existence of multicollinearity among the independent variables, difficulties in the data collection [24] and the model may depend on some predictor variables which are not available at the moment of prediction. They are limited also by the large amount of time and substantial skills required to establish correct relationships. Time series forecasting models and econometric models are often incapable of simulating complicated nonlinear properties of the destination demands [25] when non-linearity and noise exist in tourism demand data.

Artificial Intelligence (AI) forecasting methods, including Artificial Neural Networks, rough sets theory, fuzzy time series method, support vector regression, grey theory, and modern deep learning. AI models have higher adaptability and can explore non-linear relationships. Artificial NNs (ANNs) are one of the most popular non-linear modeling methods used in tourism demand studies [26]. For example, Kon and Turner [26] showed that ANNs often had best accurate than time-series models. Cho [27] examined three time-series methods and the ANN model for predicting visitor arrivals. It was found that ANN is more accurate than other models.

Overfitting, slow convergence, and unpredictable solutions during training, NN redundancy, and getting stuck in local minima are the most drawbacks of ANNs. Deep learning is a neural network with many layers. Deep learning can explore more complex non-linear patterns in the data, can handle complex data with various structures, prevents over fitting problems for a large number of inputs, and can Improves model performance. In the last few years, a few publications used the deep learning to model tourism demand [28]. For example, Law et al. [8] compared the forecasting performances of LSTM and other methods (naive method, SVR, ANN), to predict Monthly Macau tourist arrival volumes. The experiments demonstrated that the DL technique (LSTM) performs better than other models. Li and Cao [29] forecast tourism flow based on the LSTM technique. The results showed that the LSTM technique is more accurate than the benchmark models.

2.2. TOURISM DEMAND FORECASTING WITH ONLINE DATA

A different kind of data has been used for tourism demand predicting. Based on the data type, previous literature related to tourism demand can be decomposed into two categories: traditional data and online data. Traditional data is statistical data. The majority of statistical data are extracted daily [30], monthly [31], quarterly [32], and annually [33] from tourism industry organizations; considering the major lack of data about tourist arrivals, it is difficult to apply these data to forecasting tourism demand. Moreover, the data size is limited [34]. Due to these disadvantages, the application of traditional data in tourist demand studies is limited.

Several studies have considered internet data as a complement to traditional data and as a new feature; hence, they employed big data to understand tourist satisfaction [35]. One of the major advantages of online data provided by search engines and social media is that it is real-time [36]. Two kinds of online data have emerged in the tourist arrivals prediction literature: search engine data and social media data [13]. Several studies have focused on the impact of search engine data on tourism demand forecasting performance [37]. For example, Law et al. [8] used search engine data as features in the LSTM method to forecast the monthly Macau tourist arrival volumes; likewise, Wen et al. [9] used search queries for forecasting tourism demand. The results demonstrate that models that contain a composite search index show good prediction performance; Bangwayo-Skeete and Skeete [31] try to define that Google data can be a factor for tourism prediction. And found that Google search data can improve forecasting performance.

Social media plays an important role in information search. It has been used by tourists to share their experiences. Limited researchers have attempted to predict tourist arrivals using online reviews data. For instance, Önder et al. [12] indicate that Facebook likes data incorporated into econometric models can be exploited as a new variable to explain tourism demand at different destinations. Park et al. [15] apply the news as data for predicting tourist arrivals in Hong Kong. Peng et al. [38] implemented social network data, sentiment analysis, and Gradient Boosting Regression Trees to forecast Huang Shan tourism demand, which has always resulted in good forecasting performance. Fronzetti et al. [39] employed Factor Augmented Autoregressive and Bridge models with social network and semantic variables which have the highest performance than other algorithms based on GoogleTrend data. In most of these studies, the researchers attempted to capture only the likes or the sentiments to build a predictive model. However, topics in the reviews can be used as factors in the predictive model.

3. METHODS

As schematically illustrated in Fig. 1, this framework includes six stages. In the first step, we collect online reviews about tourism demand from the Trip Advisor travel forum and the number of tourist arrivals. Then, the unstructured data were pre-processed. In Step 3, the LDA topic model is used to extract topics and their keywords. The keywords that have impacted tourism demand can be defined as features using TF-IDF representation and Pearson correlation. These features are used as input in the LSTM method to generate a predictive model. Finally, we evaluated the forecasting performances of the proposed model. .

3.1. DATA EXTRACTION

Historical data: Tourist arrivals to Hong Kong from the USA are used to measure tourism demand. We collected the monthly number of tourists From January 2013 to January 2020 from the partnernet website (https://partnernet.hktb.com/).

Online post data: The online review data used in this study were collected from the TripAdvisor travel forum. TripAdvisor is one of the most popular travel platforms and contains several topics or posts, submitted by users. We collected posts using WebHarvy software From January 2013 to January 2020.

3.2. DATA PREPROCESSING

All the data collected were preprocessed using three essential steps:

Tokenization: The first thing in text was breaking the social media posts into words, phrases, or other significant pieces named tokens. We used NLTK word tokenizer to tokenize the posts.

Stop words removal: are words which do not contribute to meaning and are dropped from the text to provide a simpler analysis of the text. For our study, the stop words were removed using the nltk corpus of stopwords.

Lemmatization: tried to get the base form of a word, known a lemma. In our study, lemmatisation is implemented using the WordNet Lemmatizer.

3.3. FEATURE CONSTRUCTION

Our paper uses topic modeling to determine important topics impacting the tourism demand from the posts. Topic modeling is adopted for discovering a set of topics. The most frequently employed topic modeling is the Latent Dirichlet Allocation (LDA) model introduced by Blei [17] in 2003. LDA is very useful and effective to find topics; it is a probabilistic generative model that was used to solve the latent semantic analysis; LDA supposes that each post can be represented as a probabilistic distribution over topics, and each topic in the LDA model is also represented as a probability distribution over words [18], as shown in Fig. 2.

The graphical representation of LDA is illustrated in Fig. 3. Each topic has a corresponding probability distribution for different words. Each post $m \in \{1,...,M\}$ is considered as a multinomial distribution $\theta(d)$ over K topics, and each topic Z is supposed to have a multinomial distribution ϕ over the collection of words W. In concept, LDA detect the topics and the distribution of these topics in each post from a corpus of posts M.

As illustrated in Fig. 4., the LDA process takes all the posts as input. This data must be preprocessed. The result of LDA is K topics, which contain N words.

Selecting the right number of topics K is a crucial component of topic modeling. The topic coherence is utilized in this study to evaluate the topic model findings, which detects semantic similarity between high-ranking words in the topic. In particular, the coherence measures the frequencies of occurrence in documents with which the high-ranking and lower-ranking terms that are related to the same topic.

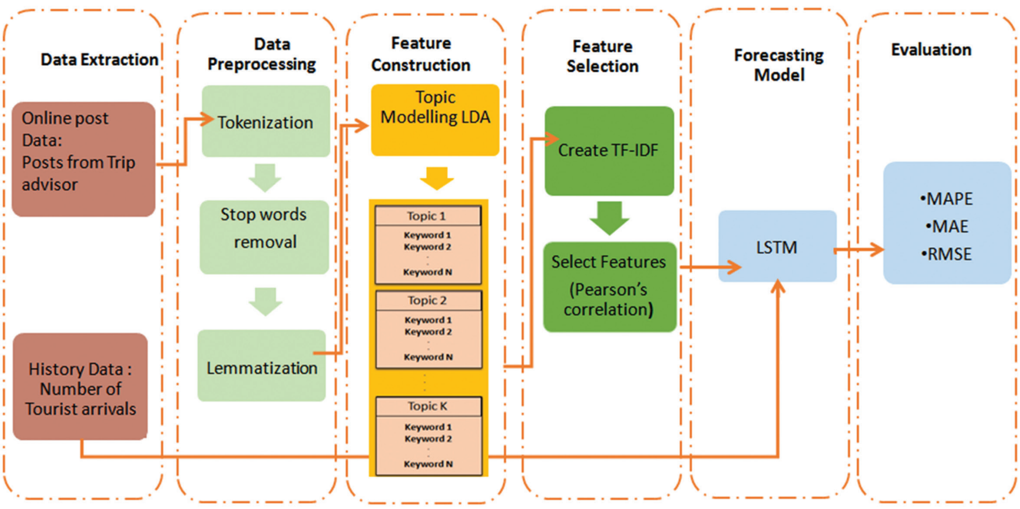

Fig. 1. Framework of proposed model

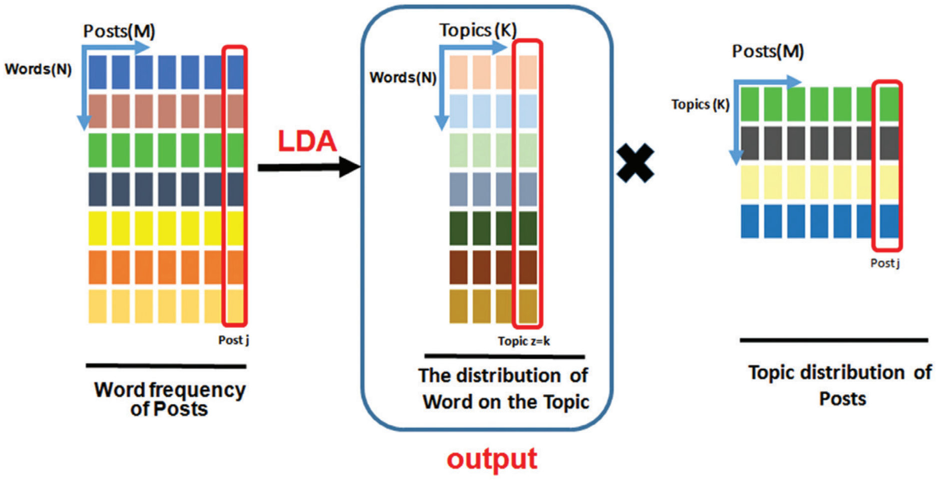

Fig. 2. The implementation of the LDA model

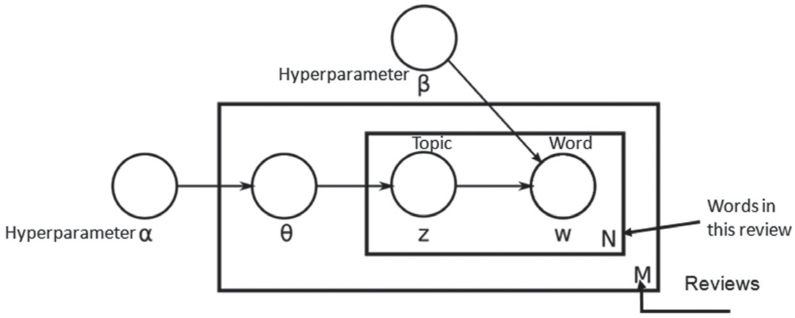

Fig. 3. The graphical model of LDA.

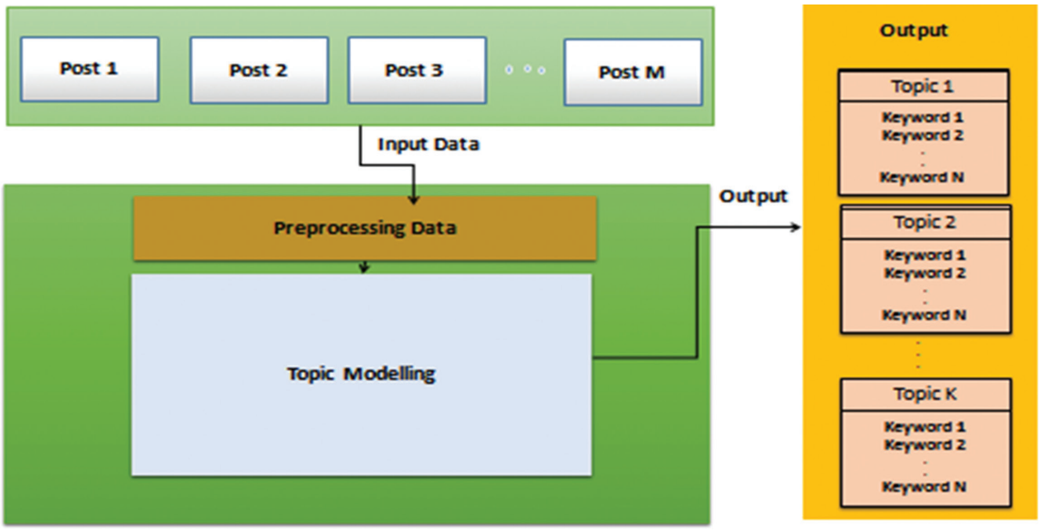

Fig. 4. Output of LDA model

3. 4. FEATURE SELECTION

- Numerical representation of the word: The TF-IDF for Term frequency- inverse Document Frequency, is employed to convert a review into structured format. It measures the importance of a given word in a given review, and the meaning of a word is defined by its frequency.
- Pearson's correlation: Pearson's correlation is the procedure of choosing an influential factor from a large collection of processed data.

3.5. TOURISM DEMAND PREDICTION BASED ON DEEP LEARNING

After identifying the appropriate features, the prediction models may be employed to predict the tourism demand in Hong Kong. In this paper, one deep learning model is chosen as the principal method, and three machine learning models are chosen as benchmark methods, i.e. LSTM, SVR, RF, and ANN. They are recurrent prediction technique, kernel prediction technique, ensemble prediction technique, and nonlinear prediction technique, respectively.

LSTM is a kind of Recurrent Neural Network (RNN) with additional variables to remember the sequence of data. LSTM is developed by Hochreiter to resolve the vanishing gradient problem of RNN. Each LSTM unit is a cell that saves information which is updated by three gates: the input gate, the forget gate, and the output gate. The input gate determines what significant information can be added from the current step and the output gate decides the part of the cell state being outputted. An illustration of LSTM unit structure is shown in Fig. 5.

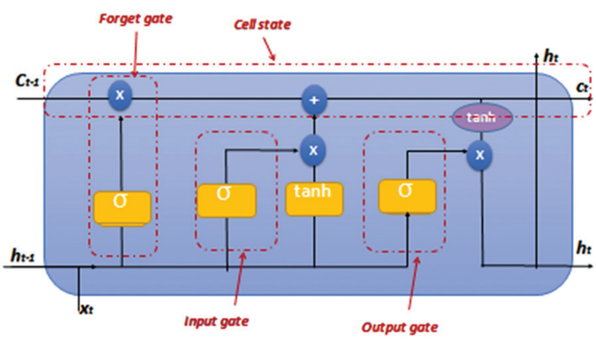

Fig. 5. The LSTM unit structure

Fig. 6. displays the model implemented for tourism demand forecasting with LSTM. The dataset used in this model is the association between the data that arrived from the TF-IDF model and the number of tourists monthly. This model consists of one hidden layer of LSTM units followed by an output layer for forecasting tourism demand.

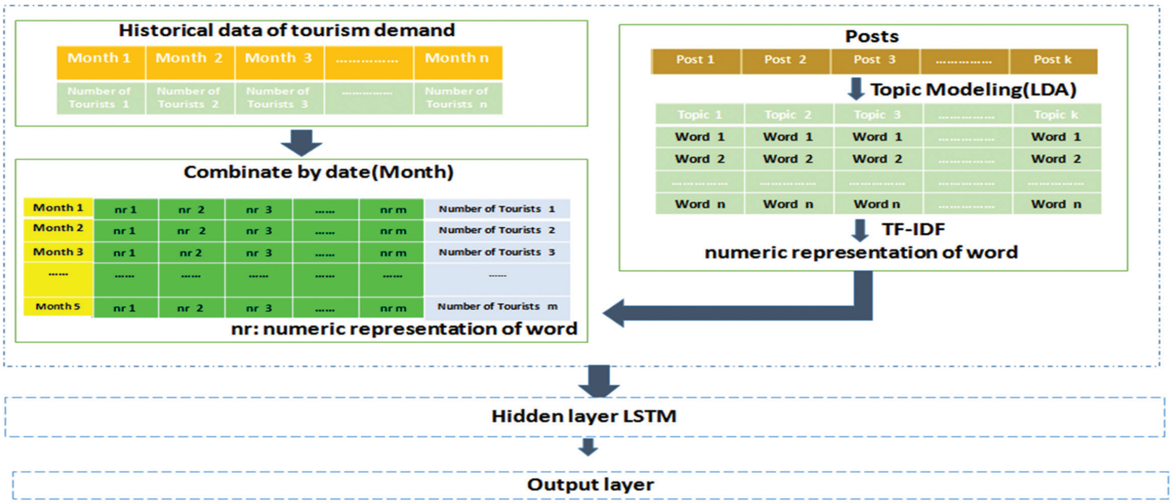

Fig. 6. Lstm in our framework

3.6. EVALUATION METRICS

For the evaluation of all models, we use Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root-Mean-Square Error (RMSE) measures. The smaller values of RMSE, MAE, and MAPE conduct to higher accuracy of the best model.

4. RESULTS AND DISCUSSION

The monthly number of tourists is illustrated in Fig. 7. In this study, first, WebHarvy software is used to collect reviews from TripAdvisor, which are registered in the excel document. Various kinds of data were crawled, comprising the origins of the posters (just from USA), the year and the month of the post, the title of the post, and the contents of the post. In total, 4987 posts and 17532 words were obtained. Right away after the collection of the posts. We undertake data cleaning or data preprocessing. Approximately 5484 words were left.

4.1. FEATURE CONSTRUCTION

To construct the LDA model, it is essential to define the optimal number of topics K. K ranges from 1 to 20. Figure 8 displays the coherence score values. The model that provided the highest coherence score is chosen. In our case, the top coherence score was reached at K = 10. Ten topics are extracted for analysis and discussion because they have the best coherence score among the other topics.

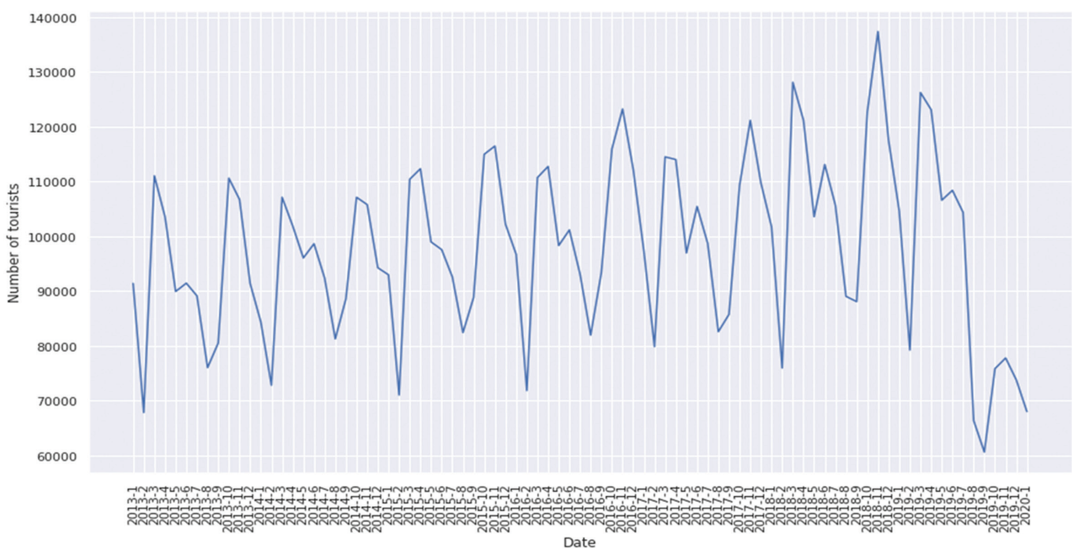

Fig. 7. Actual tourist arrivals to HongKong from USA

After trying several values for the number of topic K, we set it to 10. LDA does not generate a label for each topic; we manually assign a label to a topic word. Table 1 shows the top five words from the top ten topics identified from the posts using the Latent Dirichlet Allocation (LDA) topic modeling method. LDA can distill essential information regarding tourism demand from social media data. The ten topics are: Transport, lodging, Dining, Weather, Visit, Experience, Currency, Shopping, Busy place, Guide.

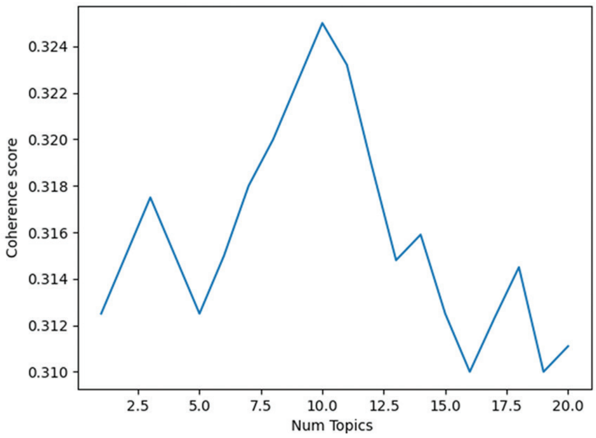

Fig. 8. The Coherence score

Table 1. Important topics identified using Latent DirechletAllocation.

Topics	Words
transport	Airport, metro, tickets, reservations, train.
lodging	Hotel, accommodation, street, Apartment, views
Dining	Bar, restaurant, food, eat, casino
Weather	Weather, degrees, down, heat, rain
Visit	Visit, trip, Disneyland, Oceanpark, museum.
Experience	Good, Bad, very interesting, Amazing, fun
Currency	Exchange rate, credit card, People, want, Money
Shopping	Shop, Price, Gift, Fashion, Clothing
Busy place	Busy, Place, People, crowd, Holiday
Guide	Guide, Sightseeing,Location, Hong Kong , Ticket

4.2. PEARSON CORRELATION

The data is randomly divided in certain percentages, 80% of our data is selected as the training data and the rest is considered as testing data.

We calculated the correlation between the numbers of tourists and all other features and selected the words that have a correlation of more than 0.01. It reduced the number of variables by more than 40%.

4.3. HYPERPARAMETER OF LSTM MODEL

For training our proposed model, it should discover the optimal combination of hyper parameters to get the best performance. After many experiments, we obtain the following combination: The model is trained for 50 epochs with a batch size of 64 and learning rate was 0.002. The optimizer used is the Adam optimizer.

4.3. MODEL EVALUATION

The error indices of the proposed model are calculated and illustrated in Table 3. The LSTM method outperforms the SVR, RF and the ANN in terms of RMSE, MAE and MAPE. The RMSE value of the LSTM method is 840.87 The MAE value of the LSTM method is 743.32, while the MAPE value is 5.145.

To test the effectiveness of integrating reviews data and topics in the predictive model, we calculate the error indices for the LSTM, RF, SVR and ANN methods with only historical data (the number of tourists only). The results are shown in Table 3. In terms of RMSE, MAE and MAPE, we observe that models based only on the number of tourists as a feature show unsatisfactory results and achieve higher values compared to models that use topics as features (Table 2). Furthermore, our approach considering LSTM model outperforms all other models with topics and with only traditional data.

Fig. 9. demonstrate the comparison between actual tourist arrivals and predictions of all models that use topics as features. The actual tourist arrival and prediction of the proposed model are shown as a blue line and an orange line, respectively. In contrast, the predictions of other benchmark models are denoted by a green dashed line for RF, a red dashed line for SVR and a purple dashed line for ANN. The forecasting accuracy of the LSTM forecasting model is higher than that of the machine learning forecasting models.

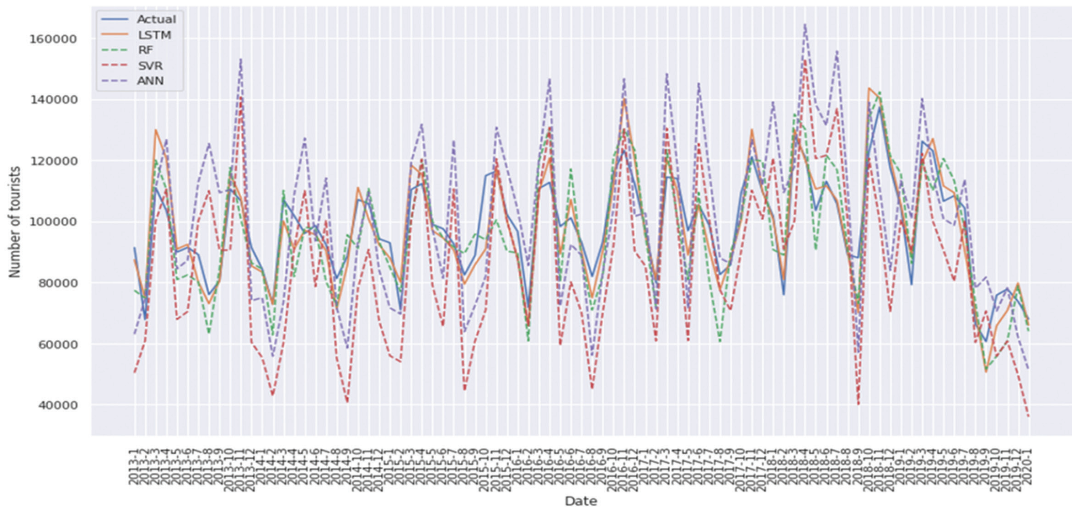

Fig. 9. Actual tourist arrivals and predictions of all models

Table 2. Prediction performance for each model with social media data

	Training data			Test data		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
LSTM	**840.87**	**743.32**	**5.145**	**960.54**	**855.46**	**5.475**
SVR	1653.98	1399.42	8.835	1721.39	1585.74	8.934
RF	1139.48	932.62	6.483	1183.73	1024.83	6.947
ANN	1463.82	1256.39	7.719	1594.35	1594.36	8.475

Table 3. Prediction performance for each model with only historical data

	Training data			Test data		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
LSTM	**1680.93**	**1655.68**	**8.879**	**1764.46**	**1710.32**	**9.121**
SVR	1826.29	1802.90	9.209	1884.35	1865.46	9.532
RF	1734.73	1729.52	8.971	1799.64	1784.63	9.249
ANN	1772.38	1768.31	9.138	1862.35	1855.63	9.438

5. CONCLUSIONS& FUTURE WORK

In this article, we attempt to construct an accurate model for tourism demand prediction, which is a very challenging task. Researchers have used search engine data and traditional data in a forecasting model. Although search engine data can complement traditional data, it produces less information than social media data. We presented a novel model that uses social media data by determining the main topics present in the online reviews and adding them as new factors in the deep learning model to predict tourism demand. We extracted data from the TripAdvisor review platform; we discovered a list of influential variables with topic modeling. We used all features in the deep learning LSTM method to predict tourism demand. Experiments indicated that our model achieved better accuracy than the other machine learning methods with social media data on the one hand and with only the number of tourists on the other hand. This study showed that forecasting performance can be greatly enhanced by using online review data to predict tourism demand; In future work, we will take sentiment analysis of social media reviews into consideration to predict tourism demand. Moreover, search query data and other types of data (such as weather data, and temperature data) can be combined to supplement the social media data to provide accurate forecasts.

6. REFERENCES:

[1] X. Xu, R. Law, W. Chen, L. Tang, "Forecasting tourism demand by extracting fuzzy Takagi–Sugeno rules from trained SVMs", CAAI Transactions on Intelligence Technology, Vol. 1, No. 1, 2016, pp. 30-42.

[2] F. L. Chu, "Forecasting tourism: A combined approach", Tourism Management, Vol. 19, No. 6, 1998, pp. 515-520.

[3] M. E. Nor, A. I. M. Nurul, M. S. Rusiman, "A Hybrid Approach on Tourism Demand Forecasting", Journal of Physics: Conference Series, Vol. 995, No. 1, 2018.

[4] A. G. Assaf, G. Li, H. Song, M. G. Tsionas, "Modeling and Forecasting Regional Tourism Demand Using the Bayesian Global Vector Autoregressive (BGVAR) Model", Journal of Travel Research, Vol. 58, No. 3, 2019, pp. 383-397.

[5] J. W. Bi, Y. Liu, H. Li, "Daily tourism volume forecasting for tourist attractions", Annals of Tourism Research, Vol. 83, 2020, p. 102923.

[6] H. Song, G. Li, "Tourism demand modelling and forecasting-A review of recent research", Tourism Management, Vol. 29, No. 2, 2008, pp. 203-220.

[7] J. W. Bi, H. Li, Z. P. Fan, "Tourism demand forecasting with time series imaging: A deep learning model", Annals of Tourism Research, Vol. 90, 2021.

[8] R. Law, G. Li, D. K. C. Fong, X. Han, "Tourism demand forecasting: A deep learning approach", Annals of Tourism Research, Vol. 75, 2019, pp. 410-423.

[9] L. Wen, C. Liu, H. Song, "Forecasting tourism demand using search query data: A hybrid modelling approach", Tourism Economics, Vol. 25, No. 3, 2019, pp. 309-329.

[10] A. Kulshrestha, V. Krishnaswamy, M. Sharma, "Bayesian BILSTM approach for tourism demand forecasting", Annals of Tourism Research, Vol. 83, 2020.

[11] J. W. Bi, C. Li, H. Xu, H. Li, "Forecasting Daily Tourism Demand for Tourist Attractions with Big Data: An Ensemble Deep Learning Method", Journal of Travel Research, Vol. 61, No. 8, 2022, pp. 1719-1737.

[12] I. Önder, U. Gunter, A. Scharl, "Forecasting tourist arrivals with the help of web sentiment: A mixed-frequency modeling approach for big data", Tourism Analysis, Vol. 24, No. 4, 2019, pp. 437-452.

[13] H. Li, M. Hu, G. Li, "Forecasting tourism demand with multisource big data", Annals of Tourism Research, Vol. 83, 2020, p. 102912.

[14] M. Hu, H. Li, H. Song, X. Li, R. Law, "Tourism demand forecasting using tourist-generated online review data", Tourism Management, Vol. 90, 2022, p. 104490.

[15] E. Park, J. Park, M. Hu, "Tourism demand forecasting with online news data mining", Annals of Tourism Research, Vol. 90, 2021.

[16] H. Song, R. T. R. Qiu, J. Park, "A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting", Annals of Tourism Research, Vol. 75, 2019, pp. 338-362.

[17] Gerard S. Forecasting tourist arrivals in Barbados. Annals of Tourism Research, 1995, Vol. 22, No. 4, pp. 804-818.

[18] R. Law, "Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting", Tourism Management, Vol. 21, No. 4, 2000, pp. 331-340.

[19] V. S. Lin, A. Liu, H. Song, "Modeling and Forecasting Chinese Outbound Tourism: An Econometric Approach", Journal of Travel and Tourism Marketing, Vol. 32, No. 1-2, 2015, pp. 34-49.

[20] X. Sun, W. Sun, J. Wang, Y. Zhang, Y. Gao, "Using a Grey-Markov model optimized by Cuckoo search algorithm to forecast the annual foreign tourist arrivals to China", Tourism Management, Vol. 52, 2016, pp. 369-379.

[21] H. Choyakh, "A Model of Tourism Demand for Tunisia: Inclusion of the Tourism Investment Variable", Tourism Economics, Vol. 14, No. 4, 2008, pp. 819-838.

[22] H. Song, S. F. Witt, "Forecasting international tourist flows to Macau", Tourism Management, Vol. 27, No. 2, 2006, pp. 214-224.

[23] S. Page, H. Song, D. C. Wu, "Assessing the Impacts of the Global Economic Crisis and Swine Flu on Inbound Tourism Demand in the United Kingdom", Journal of Travel Research, Vol. 51, No. 2, 2012, pp. 142-153.

[24] P. J. Sheldon, T. Var, "Tourism forecasting: A review of empirical research", Journal of Forecasting, Vol. 4, No. 2, 1985, pp. 183-195.

[25] B. Zhang, Y. Pu, Y. Wang, J. Li, "Forecasting hotel accommodation demand based on LSTM model incorporating internet search index", Sustainability, Vol. 11, No. 17, 2019, pp. 1-14.

[26] S. C. Kon and L. W. Turner, "Neural Network Forecasting of Tourism Demand", Tourism Economics, Vol. 11, No. 3, 2016, pp. 301-328.

[27] C. Vincent, "A comparison of three different approaches to tourist arrival forecasting", Tourism management, Vol. 24, No. 3, 2003, pp. 323-330.

[28] K. He, L. Ji, C. W. D. Wu, K. F. G. Tso, "Using SARIMA-CNN-LSTM approach to forecast daily tourism demand", Journal of Hospitality and Tourism Management, Vol. 49, 2021, pp. 25-33.

[29] Y. Li , H. Cao, "Prediction for Tourism Flow based on LSTM Neural Network", Procedia Computer Science, Vol. 129, 2018, pp. 277-283.

[30] L. Shao-Jiang, C. Jia-Ying, L. Zhi-Xue, "A EMD-BP integrated model to forecast tourist number and applied to Jiuzhaigou", Journal of Intelligent and Fuzzy Systems, Vol. 34, No. 2, 2018, pp. 1045-1052.

[31] P. F. Bangwayo-Skeete, R. W. Skeete, "Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach", Tourism Management, Vol. 46, 2015, pp. 454-464.

[32] C. Goh, R. Law, H. M. K. Mok, "Analyzing and Forecasting Tourism Demand: A Rough Sets Approach", Journal of Travel Research, Vol. 46, No. 3, 2008, pp. 327-338.

[33] C. H. Wang, L. C. Hsu, "Constructing and applying an improved fuzzy time series model: Taking the tourism industry for example", Expert Systems with Applications, Vol. 34, No. 4, 2008, pp. 2732-2738.

[34] H. Song, H. Liu, "Predicting Tourist Demand Using Big Data", Analytics in Smart Tourism Design, 2017, pp. 13-29.

[35] S. Sun, Y. Wei, K. L. Tsui, S. Wang, "Forecasting tourist arrivals with machine learning and internet search index", Tourism Management, Vol. 70, 2019, pp. 1-10.

[36] Y. Yang, B. Pan, H. Song, "Predicting Hotel Demand Using Destination Marketing Organization's Web Traffic Data", Journal of Travel Research, Vol. 53, No. 4, 2014, pp. 433-447,.

[37] B. Zhang, X. Huang, N. Li, R. Law, "A novel hybrid model for tourist volume forecasting incorporating search engine data", Asia Pacific Journal of Tourism Research, Vol. 22, No. 3, 2017, pp. 245-254.

[38] T. Peng, J. Chen, C. Wang, Y. Cao, "A Forecast Model of Tourism Demand Driven by Social Network Data", IEEE Access, Vol. 9, 2021, pp. 109488-109496.

[39] A. F. Colladon, B. Guardabascio, R. Innarella, "Using social network and semantic analysis to analyze online travel forums and forecast tourism demand", Decision Support Systems, Vol. 123, 2019.