Air travel demand forecasting based on big data: A struggle against public anxiety

Xiaozhen Liang, Chenxi Hong, Wenkun Zhou and
Mingge Yang*

School of Management, Shanghai University, Shanghai, China

It is of great significance to accurately grasp the demand for air travel to promote the revival of long-distance travel and alleviate public anxiety. The main purpose of this study is to build a high-precision air travel demand forecasting framework by introducing effective Internet data. In the age of big data, passengers before traveling often look for reference groups in search engines and make travel decisions under their informational influence. The big data generated based on these behaviors can reflect the overall passenger psychology and travel demand. Therefore, based on big data mining technology, this study designed a strict dual data preprocessing method and an ensemble forecasting framework, introduced search engine data into the air travel demand forecasting process, and conducted empirical research based on the dataset composed of air travel volume of Shanghai Pudong International Airport. The results show that effective search engine data is helpful to air travel demand forecasting. This research provides a theoretical basis for the application of big data mining technology and data spatial information in air travel demand forecasting and tourism management, and provides a new idea for alleviating public anxiety.

KEYWORDS

public anxiety, big data mining, air travel demand, convolutional LSTM, artificial intelligence algorithm

Introduction

Today's world is in an unprecedented upheaval. The whole world is suffering from various psychosocial problems under the catalyst of various unstable factors. For example, from the perspective of interpersonal relationship, the occurrence of the aggregation epidemic has led to the interpersonal relationship being fragility, alienation and lack of trust (Mei et al., 2022); and from the perspective of individual pressure, the terrible economic environment and severe employment pressure have forced individuals to lower their expectations of life, whose pessimism have pervaded the whole society (Lu and Lin, 2021). Many researchers have conducted process analysis and means discussion on the adjustment of social psychological state from different perspectives, such as tourism (Chen et al., 2020), community positive psychological intervention (Montiel et al., 2021), and localized

constructive news (Ouyang et al., 2021). Among them, tourism as a "spiritual product," can not only meet complexity requirements of people (i.e., the pursuit of novelty, change and unpredictability) to strongly stimulates their dopamine, but also drive consumption and promote economic recovery (Hu et al., 2022). In addition, taking into account the freshness of scenic spots outside the province for people, long-distance travel can often generate more powerful energy (Hannah and Lisa, 2020). As an important way of long-distance travel, aviation plays a pivotal role. It is of great significance to accurately predict the future air travel demand, which contributes to not only carry out infrastructure construction, emergency plan adjustment and resource scheduling for the government, but also reasonable operation strategies formulation for tourism companies to improve the quality of tourism services, passenger satisfaction and security (Fildes et al., 2011). Therefore, improving the accuracy of air travel demand forecasting can be regarded as a new idea to promote the recovery of long-distance travel and alleviate public anxiety, which has great research value.

Most existing studies on air travel demand forecasting have always been dependent on traditional historical statistical data (Jin et al., 2020), which is released with low frequency and only reflects relatively objective information on a macro level. In the era of Internet big data, as a matter of fact, data from Internet search queries provide a new source of predictive variables. When tourists generate tourism intentions about a destination, they need to have a deep understanding of the object due to their knowledge seeking psychology. And the more frequently tourists search for information, the stronger the correlation between their tourism behavior and tourism intention (Tavitiyaman et al., 2021). Therefore, the overall tourism intention reflected by the search engine data, a kind of big data, is also highly correlated with the social travel demand. A large number of existing studies indicate that the predictive accuracy will be generally improved when search engine data of query keywords are used as explanatory variables in forecasting models (Liu et al., 2021; Aaronson et al., 2022). However, most of the previous literature have ignored the damage to the prediction accuracy caused by much noise information contained in big data. In this paper, we aim to explore how to make use of this big data in turn to achieve better air travel demand forecasting performance when the Internet is gradually dominating people's lives. Meanwhile, the novel idea is provided for the recovery of tourism and the alleviation of public anxiety in the turbulent era. The underlying logic of this study is shown in Figure 1.

In addition, the nonlinear and dynamic data characteristics of air travel demand and explanatory factors increase the problem complexity. In recent years, the common forecasting methods are mainly divided into traditional econometric models and artificial intelligence models. Traditional econometric models, such as the autoregressive integrated moving average (ARIMA) model (Jimenez et al., 2022) and the autoregressive distribution lag (ARDL) model (Madziwa et al., 2022), can capture the relationships between historical and future values to obtain better

FIGURE 1
The underlying logic of the research.
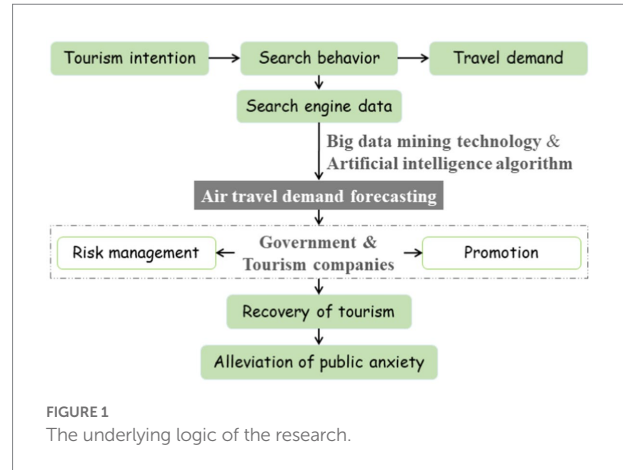
prediction results. Because of their simple principle and easy implementation, they are widely used to forecast time series to obtain high prediction accuracy for stationary time series. On the contrary, advanced artificial intelligence algorithms show strong nonlinear mapping abilities. For example, Xiong et al. (2021) designed a forecasting framework with the support vector regression (SVR), whose feasibility and effectiveness have been proved in experiments. Chen and Huang (2021) used convolutional neural network (CNN) and long short-term memory network (LSTM) respectively to implement stock price prediction based on eight input features, which fully proved the superiority of artificial intelligence models in forecasting complex time series. However, the above models are mostly used to analyze and extract the temporal relationship of data rather than the spatial relationship for air travel demand forecasting. In fact, the spatial effect of travel flows can occur and have an impact on adjacent destinations (Chhetri et al., 2013). Therefore, the spatial effect has great potential to assist the forecasting for air travel demand. The existing literature has proved the spatial correlation between tourism flows in adjacent areas (Yang and Zhang, 2019; Jiao et al., 2020; Ouchen and Montargot, 2021). However, there are few researches to make full use of the spatial correlation of series to assist air travel demand forecasting. Expanding to other fields, Zhang et al. (2021) embedded CNN model and LSTM model into traffic generative adversarial nets to deeply capture temporal and spatial correlation between road connections for road network level prediction. Fu et al. (2022) proposed a wind power prediction method considering the spatial distribution characteristics of field groups based on LSTM model and time convolution neural network (TCN). Shi et al. (2015) proposed a new method namely convolutional LSTM (ConvLSTM), which organically combines the convolutional operator with LSTM to complete the learning of spatial–temporal information in essence. Above studies have confirmed that the introduction of spatial dependence of input data is conducive to improving the performance of prediction models, and the combination of LSTM model and CNN model is a feasible choice for processing spatial–temporal data. Compared with other combination methods,

ConvLSTM model can achieve deeper learning of spatial data relevance through simple model structure. Therefore, this research used ConvLSTM model to simultaneously analyze the spatial and temporal dependence of the search engine data reflecting the spatial effect of tourism flow and the historical data of air travel demand, so as to capture more effective information for assisting prediction.

Moreover, traditional econometric models and artificial intelligence models have their own advantages and disadvantages so that cannot give full play to all the advantages when processing the input data with complex components. At present, the ensemble prediction model which combines decomposition method and prediction model is generally regarded as an effective method (Li et al., 2022). More precisely, the decomposition method contributes to exploration about information of different time scales of the series pointedly to generate better prediction. In this paper, a novel decomposition algorithm named time varying filtering-based empirical mode decomposition (TVF-EMD) was applied to disassemble the information of different scales of the original sequence, which can use time-varying filtering to alleviate the mode mixing problem, and produce the decomposition results that have strong stability (Liu et al., 2022). To sum up, based on the ensemble framework, this paper attempts to use ConvLSTM model as the main prediction model, and combine the effective search engine data that can make full use of the search psychology of passengers and the spatial effect of tourism flow to achieve high-precision air travel demand prediction. Through this design of forecasting framework, the prediction accuracy of air travel demand has the opportunity to increase significantly.

The main contributions of this paper include: (1) From the perspective of improving the prediction accuracy of air travel demand, it has developed a new direction for the recovery of tourism and the alleviation of public anxiety by providing decision support for putting forward reasonable policies and measures; (2) Based on the designed dual data preprocessing method, effective search engine data is introduced, which can help to enhance the accuracy of air travel demand forecasting model; (3) The concept of spatial effect of tourism flow is first introduced into the field of air travel demand forecasting, and the ConvLSTM model is introduced to fully extract the spatial–temporal characteristics of search engine data and historical data; (4) An ensemble forecasting framework is established by combining the econometric model with artificial intelligence model for air travel demand, which can fully avoid the impact of market environment and give play to the advantages of different models to improve the prediction performance.

The rest work is designed as follows. Section "Methodology" presents the approaches used in the ensemble forecasting framework. Section "Overall forecasting framework" provides design ideas and detailed information about the overall framework of the proposed model. Section "Experimental process" describes the process of data preprocessing and prediction as necessary. In Section "Discussion," the experimental results are introduced, and the deep insights about how to make practical contributions to social well-being interventions based on the proposed model are provided. Section "Conclusion" summarizes the conclusions of this study.

Methodology

Proposed dual data preprocessing method

First-stage data preprocessing

The herd mentality and curiosity seeking mentality of most Internet users will lead to the rapid evolution of a novel event into a popular search keyword on the Internet, even if this event has nothing to do with their travel intention. As a result, the collected search engine data contains some outliers and noise, which need to be eliminated in advance. Scrupulously, two algorithms are introduced to remove the outliers and noises of initial SED:

(1) K-means clustering algorithm. It is a classic partition clustering method proposed by Macqueen (1967). For a given dataset D containing n objects in Euclidean space, K-means algorithm will divide the data objects in D into k clusters $C_1, C_2, \ldots, C_k$. The goal of partition is to make the objects within a cluster have high similarity while the objects between clusters have low similarity. In this paper, the data excluded from large clustering by K-means algorithm are considered as outliers.

The basic flow of the algorithm is described as follows:

Step 1: Set k as the number of clustering centers in advance, and input the dataset $D = \left\{ x_i \middle| x_i \in R^m, i = 1, 2, \ldots, n \right\}$ containing n data samples.

Step 2: k sample points are randomly selected as the initial clustering centers, denoted as $y_1, y_2, \ldots, y_k$.

Step 3: Calculate the Euclidean distance between each sample point x_i and each clustering center y_j, and the calculation formula is as follows:

$$d\left(x_i, y_j\right) = \sqrt{\left(x_{i1} - y_{j1}\right)^2 + \left(x_{i2} - y_{j2}\right)^2 + \ldots + \left(x_{im} - y_{jm}\right)^2} \quad (1)$$

According to the distance, each point x_i is placed into the corresponding cluster C_j to minimize the distance between x_i and y_j.

Step 4: Calculate the criterion function E for the sum of error squares after clustering, and the formula is as follows:

$$E = \sum_{j=1}^{k} \sum_{x_i \in C_j} d\left(x_i, y_j\right) \quad (2)$$

Step 5: For each cluster C_j, recalculate its new cluster center, the formula is as follows:

$$y_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i \qquad (3)$$

where $|C_j|, (j = 1, 2, \ldots, k)$ is the number of points in the cluster C_j that completed the allocation.

Step 6: Based on the new clustering centers, recalculate the criterion function E'. If $|E - E'|$ is greater than the specified threshold, repeat steps 3–5. Otherwise, the algorithm ends to output the clustering result.

(2) TVF-EMD decomposition algorithm. By definition, a single-component signal with a specific physical interpretation is called an intrinsic mode function (IMF). The traditional EMD decomposition algorithm has an inherent defect namely mode mixing problem, that is, a certain IMF generated by decomposition has different scales, or different IMFs appear on the same scale (Zheng et al., 2017). In this context, TVF-EMD proposed by Li et al. (2017) attempts to define the IMF from another perspective and adjust the cut-off frequency to overcome the mode mixing problem. Specifically, TVF-EMD algorithm will successively extract qualified sub-sequences, namely IMF components, through a series of calculation formulas. Due to space limitations, this paper only briefly describes the operation process as follows.

Step 1: Calculate the instantaneous amplitude $A(t)$ and instantaneous phase $\varphi(t)$ of the actual signal $x(t)$:

$$A(t) = \sqrt{x^2(t) + \hat{x}^2(t)} \qquad (4)$$

$$\varphi(t) = \arctan\left[\hat{x}(t) / x(t)\right] \qquad (5)$$

where $\arctan[\ ]$ is the inverse tangent function, and $\hat{x}(t)$ is the Hilbert transform of $x(t)$.

Step 2: Calculate all local maximum values $A(\{t_{max}\})$ and minimum values $A(\{t_{min}\})$ of $A(t)$, respectively.

Step 3: The upper envelope $\beta_1(t)$ and the lower envelope $\beta_2(t)$ are estimated respectively, and then the amplitudes $a_1(t)$ and $a_2(t)$ are calculated based on the following formulas:

$$\begin{cases} a_1(t) = \left[\beta_1(t) + \beta_2(t)\right] / 2 \\ a_2(t) = \left[\beta_1(t) - \beta_2(t)\right] / 2 \end{cases} \qquad (6)$$

Step 4: Based on a series of points $\varphi'(\{t_{min}\}) A^2(\{t_{min}\})$ and $\varphi'(\{t_{max}\}) A^2(\{t_{max}\})$ for interpolation respectively, accordingly

estimate $\eta_1(t)$ and $\eta_2(t)$, so as to calculate the instantaneous frequencies $\varphi_1'(t)$ and $\varphi_2'(t)$. The mathematical expressions are as follows:

$$\begin{cases} \varphi_1'(t) = \eta_1(t) / \left(2a_1^2(t) - 2a_1(t)a_2(t)\right) + \\ \eta_2(t) / \left(2a_1^2(t) + 2a_1(t)a_2(t)\right) \\ \varphi_2'(t) = \eta_1(t) / \left(2a_2^2(t) - 2a_1(t)a_2(t)\right) + \\ \eta_2(t) / \left(2a_2^2(t) + 2a_1(t)a_2(t)\right) \end{cases} \qquad (7)$$

Step 5: Calculate the local cut-off frequency $\varphi_{bis}'(t)$:

$$\varphi_{bis}'(t) = \left[\varphi_1'(t) + \varphi_2'(t)\right] / 2 = \left(\eta_2(t) - \eta_1(t)\right) / 4a_1(t)a_2(t) \qquad (8)$$

Step 6: Rearrange $\varphi_{bis}'(t)$ based on local cut-off frequency realignment.

Step 7: Calculate signal $h(t)$ accordingly:

$$h(t) = \cos\left[\int \varphi_{bis}'(t) dt\right] \qquad (9)$$

$x(t)$ is filtered by B-spline approximation, and the approximate result is $m(t)$.

Step 8: Update $x(t)$ according to the specified rules. That is, if $\theta(t) \leq \xi$, $x(t)$ is considered as an IMF component and the calculation is terminated. Otherwise, set $x(t) = x(t) - m(t)$, and repeat Step 1–Step 7. The calculation formula of the stopping criterion $\theta(t)$ is as follows:

$$\theta(t) = B_{Loughlin}(t) / \varphi_{avg}(t) \qquad (10)$$

where $\varphi_{avg}(t)$ is the weighted average of instantaneous frequencies, and $B_{Loughlin}(t)$ is *Loughlin* instantaneous bandwidth. Until the IMF is no longer generated, the remainder is denoted as a residue (R). Then, the IMF with the highest frequency in TVF-EMD decomposition results is considered as noise information (Li et al., 2017).

Second-stage data preprocessing

In order to develop search engine data that have real predictive ability for air travel demand as explanatory variables, it is necessary to adopt appropriate filtering methods to process the search engine data. Specifically, the augmented Dickey-Fuller (ADF) test, Engle-Granger (E-G) cointegration test, Granger causality test are used to select the variables with high forecasting ability for air travel demand (Huang et al., 2017). After the above statistical tests, the selected explanatory variables will have the same stability as the predicted variables, and show a strong long-term cointegration relationship and causality with the predicted

variables. Moreover, Pearson correlation coefficient analysis (Lai et al., 2022) is introduced to select the objects with relatively high correlation with the predicted variables from the variables that pass the statistical test. It is worth mentioning that Pearson correlation coefficient is the most commonly used statistical indicator to reflect the degree of correlation between variables.

Convolutional long short-term memory network

Long short-term memory network

The structure of the LSTM (shown in Figure 2) determines that it is capable of solving the problem namely long-term dependence to achieve satisfactory prediction results.

Figure 2 shows that the input gate i_t, forget gate f_t, and output gate o_t constitute the main structure of LSTM. After being activated, i_t can selectively store new input information, f_t can omit unnecessary information, and o_t can control the part of output. Finally, the current output h_t is updated. The above process can be expressed as follows:

$$f_t = \sigma\left(W_f \circ \left[h_{t-1}, x_t\right] + b_f\right) \tag{11}$$

$$i_t = \sigma\left(W_i \circ \left[h_{t-1}, x_t\right] + b_i\right) \tag{12}$$

$$o_t = \sigma\left(W_o \circ \left[h_{t-1}, x_t\right] + b_o\right) \tag{13}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tanh\left(W_c \circ \left[h_{t-1}, x_t\right] + b_c\right) \tag{14}$$

$$h_t = o_t \circ \tanh\left(c_t\right) \tag{15}$$

where W_f, W_i, W_o and W_c are weight vectors, b_f, b_i, b_o, and b_c are bias vectors, x_t is the current input, h_{t-1} represents the output of the previous moment, c_t and c_{t-1} are memory units of the current moment and the previous moment, respectively. In addition, $\sigma(\bullet)$ and $\tanh(\bullet)$ represent sigmoid function and hyperbolic tangent activation function respectively, and $\circ$ stands for the Hadamard operator.

Convolutional long short-term memory network

The ConvLSTM model is an improved version of the LSTM model. The key point of its improvement is to replace the Hadamard operators with the convolutional operators for both the input-to-state transformations and the state-to-state transformations, which makes the network enable to fully learn the hidden correlations in spatial–temporal data through limited historical data. The functions used in LSTM are modified for ConvLSTM, as shown below:

$$f_t = \sigma\left(W_f * \left[h_{t-1}, x_t\right] + b_f\right) \tag{16}$$

$$i_t = \sigma\left(W_i * \left[h_{t-1}, x_t\right] + b_i\right) \tag{17}$$

$$o_t = \sigma\left(W_o * \left[h_{t-1}, x_t\right] + b_o\right) \tag{18}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tanh\left(W_c * \left[h_{t-1}, x_t\right] + b_c\right) \tag{19}$$

$$h_t = o_t \circ \tanh\left(c_t\right) \tag{20}$$

where $*$ represents the convolutional operator.

Overall forecasting framework

Tourism can not only directly stimulate people's positive emotions, but also indirectly relieve public pressure by stimulating local economic growth (Liu et al., 2021). It is necessary to master the future air travel demand for formulating reasonable policies and measures to promote long-distance tourism. Therefore, the main purpose of this paper is to build a high-precision air travel demand forecasting framework.

In the Internet era, online search engine data can represent people's behavior and generate early indicators of future demand to a certain extent (Huang and Hao, 2021). With the rapid popularization of the Internet, search engine has become one of the primary means of modern information searching, such as
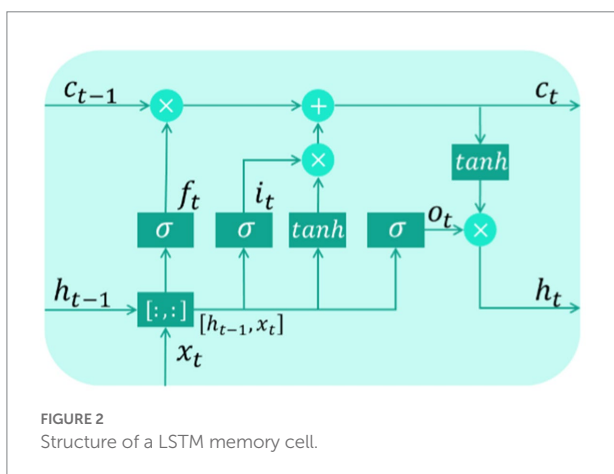
Baidu and Google. Baidu and Google generate search engine data of each keyword (i.e., Baidu Index and Google Trend) based on the search volume of hundreds of millions of users according to specific formulas, this is, the weighted sum of the search times of each keyword in their search engines (Yao et al., 2021). Both the Baidu Index (Liang et al., 2022) and Google Trend (Sun et al., 2022) can reflect the potential demand and the current focus. Considering that this paper studies the forecast of air travel demand in China, Baidu Index is a better choice.

Tourism preference is considered as an important psychological tendency (Yu and Lu, 2008). To be specific, people generally need to investigate and understand the information of tourist products, such as attractions, delicacies and overall situation, so as to determine the emotional tendency of tourist products and generate the intention of yearning or repelling. In the big data era, passengers are used to querying information in search engines before making travel plans. The relevant big data records the attention of passengers to the destination and their desire to travel to a certain extent. Therefore, appropriate search engine data can reflect the air travel demand of destinations in advance. Furthermore, considering that the travel flows of adjacent areas will have a spatial effect on it of the destination, this paper tries to obtain the required search engine data by using "name," "name + attraction" and "name + delicacies" of the destination where the airport is located and its adjacent destinations as seed keywords.

With the prevalence of new media, social self-media applications gradually carve up the market share of search engines in information query, which may lead to the change of air travel demand being reflected incorrectly by the change of the later trend of search engine data. In order to introduce available fluctuation information and eliminate unstable trend information of search engine data, decomposition-ensemble strategy can be adopted to distinguish linear trend term and nonlinear fluctuation term of variables, and appropriate prediction models can be used for analysis separately.

Based on above analysis, the proposed ensemble forecasting framework for air travel demand combining search engine data (SED), dual data preprocessing (DDP), TVF-EMD and ConvLSTM, namely SED-DDP-TVFEMD-ConvLSTM, is shown in Figure 3.

The major procedures are demonstrated as follows:

Step 1: All seed keywords are selected and searched on Baidu Index. The search volumes of these keywords and their associated keywords are found out through the Demand Map function of Baidu Index. After removing repetitive keywords and the keywords with less search, the final keywords are determined.

Step 2: The proposed DDP method is used to clean initial SEDs, including removing outliers and noises by K-means and TVF-EMD algorithms, and selecting effective SEDs by statistical tests and Pearson analysis.

Step 3: First, the TVF-EMD algorithm is applied to decompose each variable into a set of sub-sequences (i.e., IMFs and R). Then, for each variable, the R component with the lowest frequency is taken as the trend term (TT), and all IMFs are reconstructed into the fluctuation term (FT). In addition, considering that collinearity among multiple SEDs may lead to over-fitting problems in the forecasting framework, the TTs of several SEDs are further integrated into a TT component, as well as the FTs of several SEDs into a FT component, to replace the scattered SEDs to express the travel flows in the surrounding areas antecedently. Based on this, two components of air travel volume and SED can be obtained, respectively. Next, ARIMA model is employed to model and forecast the TT component of air travel volume, and ConvLSTM model is utilized to forecast the FT component of air travel volume based on the FTs of air travel volume and SED. The final forecasting result of air travel volume is obtained by simply adding the forecasting results of the FT and TT components.

Experimental process

Data preparation

Air travel volume data

In this study, the air travel volume data of Shanghai Pudong International Airport (ZSPD) was used as the experimental object to verify the positive effect of spatial–temporal search engine data and the validity of the proposed ensemble framework for air travel demand forecasting. Due to the epidemic situation, domestic airports were closed or semi closed for a long time from 2020 to the later stage, and the government had not released effective statistical data. Therefore, the data used in this paper only covers the period from January 2011 to December 2019. The required air travel volume data came from the Wind database[1], and was low-frequency monthly data containing 108 observations. In order to further enhance the performance of ConvLSTM model, the monthly data was converted into 10-day data containing 324 observations by interpolation method (Xie et al., 2021). Additionally, this dataset was divided into train set and test set based on 8:1 (Case 1) and 7:2 (Case 2) ratios, respectively, to obtain two available datasets. The corresponding visualization representation is illustrated in Figure 4.

Spatial–temporal search engine data

Keywords confirmation

In order to collect SED that can represent the travel flows of cities around Shanghai, "names of 26 urban agglomerations in the Yangtze River Delta region," "names + attraction" and "names + delicacies" were taken as seed keywords and entered into the search box of Baidu Index[2] in turn. Then, the highly relevant

---

FIGURE 3
The structure of SED-DDP-TVFEMD-ConvLSTM ensemble forecasting framework.
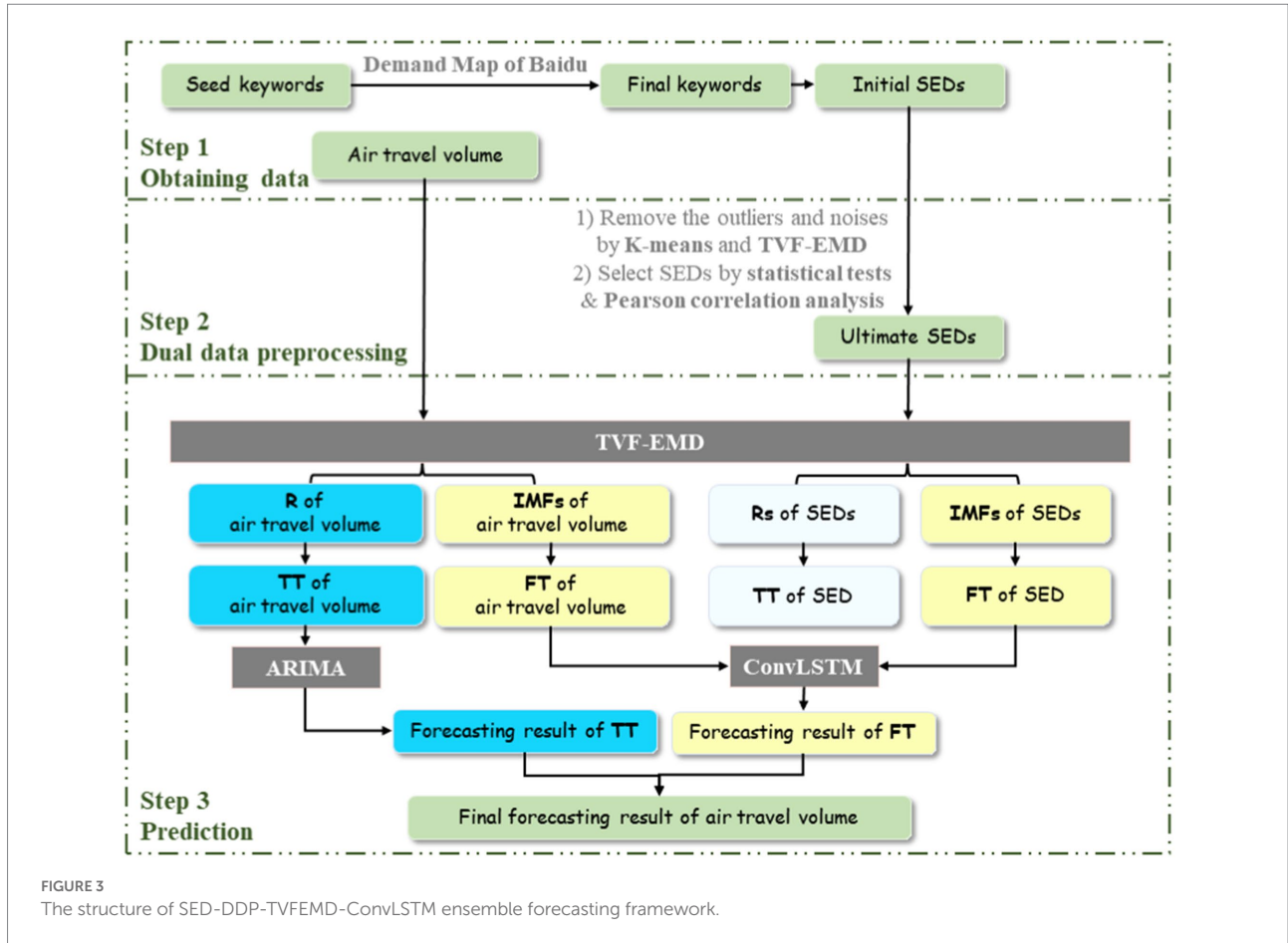
sub-keywords were screened out by the Demand Map function continuously to expand the key thesaurus. Meanwhile, the irrelevant keywords were removed by virtue of experience. Finally, total 90 keywords were determined, and the corresponding SED data from January 2011 to December 2019 were obtained according to these keywords, which were further converted into 10-day data.

First-stage data preprocessing

To optimize SED quality, K-means clustering algorithm was first adopted to identify and delete outliers of SED groups in different cities. According to the Elbow curve, the optimal number of clustering centers for each group was determined, and the outlier ratio was set to 0.05. Then, the multiple imputation method and reliability analysis were combined to generate a complete alternative dataset.

In addition, the TVF-EMD algorithm was used to decompose each SED sequence. Taking "Ningbo Attraction" SED as an example, the decomposition results are shown in Figure 5. It can be found that the high-frequency component IMF1 of the SED fluctuates violently and has a short period, which can be regarded as a random factor affecting the sequence. In other words, IMF1 contains more information unrelated to the air travel demand and no longer has the

ability to predict the air travel demand. Hence, this component was directly rounded off. Furthermore, the remaining IMFs and residue were reconstructed into a new sequence, which contains the original available information.

Second-stage data preprocessing

In this section, standard statistical tests and correlation analysis were employed to select SED variables with predictive ability for air travel demand. Specifically, the ADF test was performed on the sequences firstly, and the SEDs whose stationarity was consistent with the predicted variable were retained. Subsequently, the E-G cointegration test was used to further test whether there was a long-term equilibrium relationship between SEDs and the predicted variable, and Granger causality test was used to observe whether SEDs could explain the future variation trend of the predicted sequence. Finally, the Pearson correlation coefficient was used to obtain the SEDs that contribute greatly to the forecast of airport travel demand. We set the keyword correlation coefficient threshold to 0.4, and then reserved the keywords whose correlation coefficients are higher than the threshold.

After dual data preprocessing, five SEDs were retained, that is, "Shanghai Tour Guide," "Hangzhou," "Ningbo Attraction,"
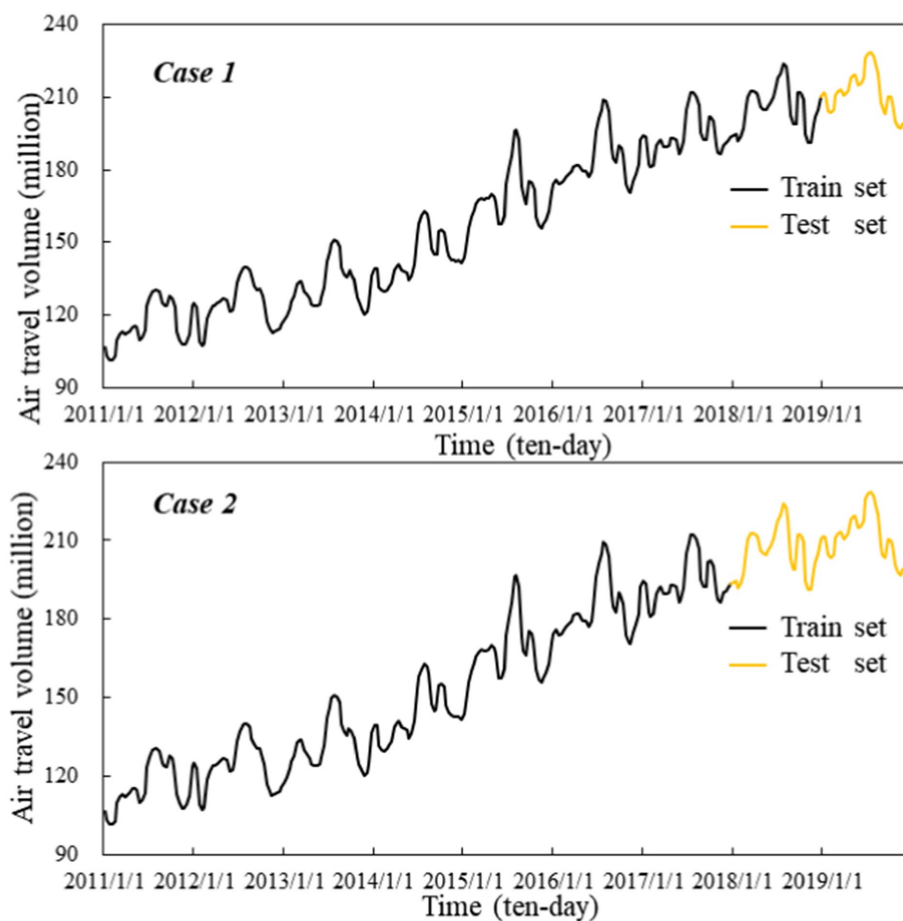
FIGURE 4
Air travel volume of Shanghai Pudong International Airport in two cases from January 2011 to December 2019.

"Jinhua" and "Zhoushan Tour Guide." These SEDs all passed the tests at a significant level of 0.01, and the Pearson correlation coefficients between them and the predicted variable are shown in Figure 6.

Prediction

Considering that different variables contain different time scales, the corresponding residue component for each variable decomposed by TVF-EMD algorithm was taken as the trend term (TT), and the remaining components were integrated as the fluctuation term (FT). To avoid the multicollinearity problem, the TT and FT components of the five SEDs were reconstructed, respectively, to obtain the final TT component and FT component of SED. Based on the introduction of SED as the explanatory variable, the FT component of air travel volume was predicted by ConvLSTM model, and the TT component of air travel volume was directly predicted by ARIMA. Finally, the predicted values of the FT and TT components were added to generate the final forecasting result of air travel volume.
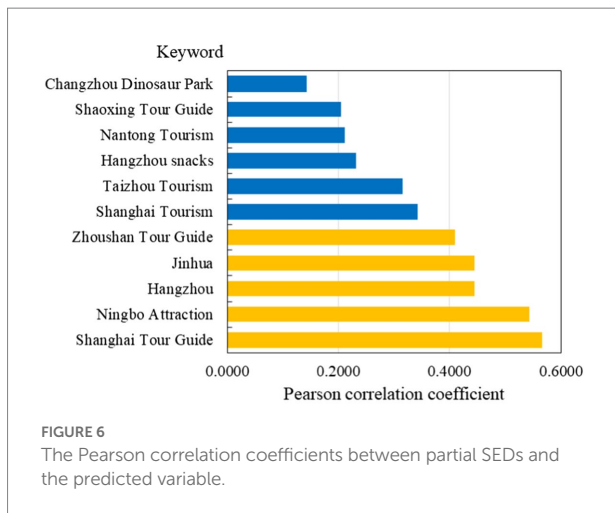
Parameter setting

As to ConvLSTM model, Adam optimization algorithm was adopted to optimize its inherent parameters, and the number of hidden layers and hidden layer nodes were determined by trial and error method, as shown in Table 1. Additionally, the B-spline order n and bandwidth threshold ξ of TVF-EMD algorithm were set to 26 and 0.25, respectively, considering its decomposition efficiency.

Evaluation criteria

To evaluate the prediction performance of the proposed ensemble forecasting framework and the comparison models, multiple evaluation metrics are applied in this study. In particular, the mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE) can effectively reflect the prediction accuracy and fitting effect. The Willmott's index of agreement (WIA) can reflect the generalization performance of the forecasting model. Their mathematical expressions are presented as follows:

FIGURE 5
Decomposition results of "Ningbo Attraction" SED after TVF-EMD algorithm.

FIGURE 6
The Pearson correlation coefficients between partial SEDs and the predicted variable.
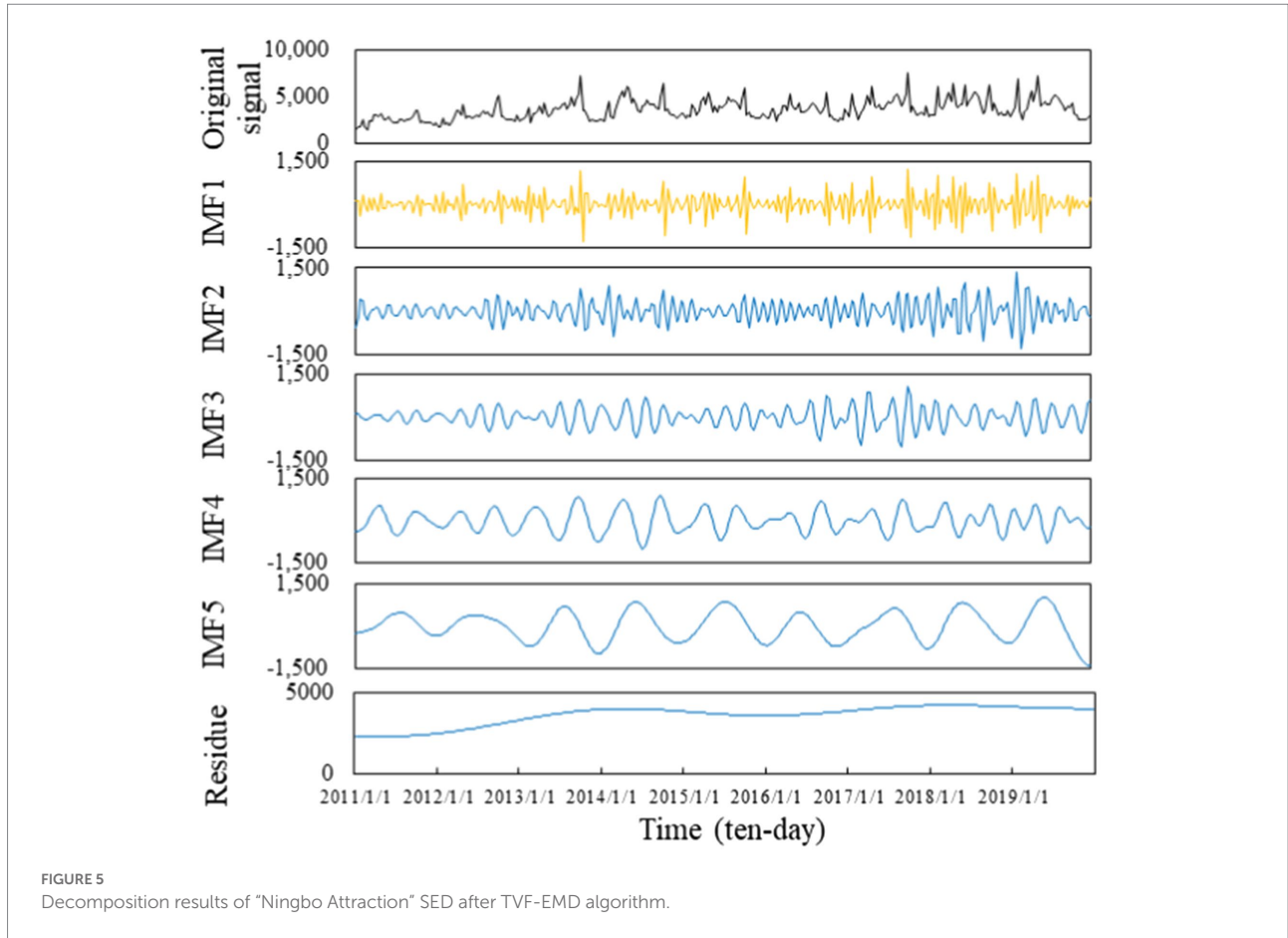
$$MAE = \sum_{i=1}^{n} \left| Y_i - \hat{Y}_i \right| / n \qquad (21)$$

$$RMSE = \sqrt{\sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 / n} \qquad (22)$$

$$MAPE = \sum_{i=1}^{n} \left| \left(Y_i - \hat{Y}_i \right) / Y_i \right| / n \times 100\% \qquad (23)$$

$$WIA = 1 - \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 / \sum_{i=1}^{n} \left(\left| Y_i - \overline{Y} \right| + \left| \hat{Y}_i - \overline{Y} \right| \right)^2 \qquad (24)$$

where Y_i and $\hat{Y}_i$ are the actual value and forecasting value at time i respectively, n is the number of testing samples, and $\overline{Y}$ represents the average value of the real data.

Furthermore, the Diebold-Mariano (DM) test (Diebold and Mariano, 1995) is introduced to evaluate the significant difference in prediction performance between the proposed framework and the comparison models from a statistical perspective, and the loss function adopts the mean square error (MSE). The null hypothesis is that the MSE value of the test model is greater than or equal to that of the comparison model, that is, the prediction performance of the test model is inferior to that of the comparison model. Assuming that model 1 and model 2 are the two models whose prediction effects need to be compared, the DM statistics can be defined by the following formula:

TABLE 1 Parameters and hyperparameters of the proposed model.

Model	Parameter	Value
ConvLSTM	Input shape	(2, 1, 3, 2)
	Filters	{12, 12}
	Kernel size	{(2, 2), (2, 2)}
	Activation function	{'relu', 'relu'}
	Loss function	Mean square error (MSE)
	Optimization algorithm	Adam
	Epoch	1,000
ARIMA	(p, d, q)	(9, 1, 2)

$$DM = \bar{d}\sqrt{\left(\sum_{i=1}^{n}\left(d_i - \bar{d}\right)\right)^2 / (n-1)} \qquad (25)$$

where $\bar{d} = \frac{1}{n}\sum_{i=1}^{n}d_i$, $d_i = \left(y_i - \hat{y}_{1,i}\right)^2 - \left(y_i - \hat{y}_{2,i}\right)^2$; $\hat{y}_{1,i}$ and $\hat{y}_{2,i}$ represent the forecasting values of model 1 and model 2 at time i respectively; y_i is the real value of the original series at time i; n is the number of testing samples. Under the null hypothesis, the DM test value obeys the standard normal distribution.

Discussion

Experimental result analysis

To comprehensively and fully evaluate the prediction accuracy and effectiveness of the proposed model (i.e., model ① SED-DDP-TVFEMD-ConvLSTM), some comparative models were designed for experiments from the following aspects:

1. In order to verify the effectiveness of the strategy of introducing search engine data, under the framework of non-decomposition, we compared the model with the preprocessed SED as explanatory variable (i.e., model ③) and the single ConvLSTM model without explanatory variables (i.e., model ⑥).
2. To verify the effectiveness of the dual data preprocessing strategy for search engine data, we established a new comparison model without decomposition (i.e., model ⑤). Specifically, five SEDs without DDP were randomly selected from the original SEDs and synthesized into an integrated SED, which was used as the explanatory variable of ConvLSTM model. Then, the feasibility of DDP strategy was further analyzed by comparing model ③ with model ⑤.
3. To demonstrate the role of the spatial–temporal artificial intelligence model in the proposed forecasting framework, two groups of comparative experiments were designed in this paper. First, in the case of

introducing decomposition-ensemble strategy, the model with ConvLSTM as the main prediction method (i.e., model ①) was compared with the model with LSTM as the main prediction method (i.e., model ②). Secondly, in the case of without introducing decomposition-ensemble strategy, the model with ConvLSTM as the main prediction method (i.e., model ③) was compared with the model with LSTM as the main prediction method (i.e., model ④).
4. To verify the feasibility of the decomposition-ensemble strategy, two groups of comparative experiments were designed. First, in the case of taking ConvLSTM as the prediction method, the ensemble forecasting framework based on TVF-EMD algorithm (i.e., model ①) was compared with the non-ensemble forecasting framework (i.e., model ③). Secondly, in the case of taking LSTM as the prediction method, the ensemble forecasting framework based on TVF-EMD algorithm (i.e., model ②) was compared with the non-ensemble forecasting framework (i.e., model ④).
5. In order to prove the comprehensive advantages of the forecasting framework considering search engine data, spatial effect and decomposition-ensemble strategy, the proposed framework (i.e., model ①) was compared with the classical single model (i.e., model ⑦ and model ⑧).

The error evaluation results of the proposed model and the comparison models in two cases are shown in Table 2, and based on the analysis of Table 2, the following conclusions can be drawn:

1. Comparing the prediction results of model ③ and model ⑥, it can be observed that the MAE value, RMSE value and MAPE value of model ③ in Case 1 decrease by 15.79%, 13.03%, and 14.98% respectively, and the WIA value increases by 1.06%; in case 2, the MAE value, RMSE value and MAPE value of model ③ decrease by 15.74%, 9.81%, and 14.73% respectively, and the WIA value increases by 2.19%. The comparison results illustrate that the search engine data can be introduced as a more flexible and fuller data source to make up for the singleness of historical statistical data, and effectively improve the prediction accuracy of the model.
2. Without decomposition, the model with DDP strategy for SED (i.e., model ③) is obviously superior to the model without DDP strategy for SED (i.e., model ⑤). For example, compared with model ⑤, the decrease rates of MAE value, RMSE value and MAPE value of model ③ are 23.62%, 15.54%, and 24.54%, respectively, in Case 1, and 23.30%, 23.99%, and 23.23%, respectively, in Case 2. It is proved that the proposed dual data preprocessing strategy has a great cleaning effect, which can correctly identify

TABLE 2 Prediction results of all the models.

Case	Model	MAE	RMSE	MAPE (%)	WIA
Case 1	① SED-DDP-TVFEMD-ConvLSTM	**1.5068**	**2.0332**	**0.7209**	**0.9866**
	② SED-DDP-TVFEMD-LSTM	2.1629	3.2654	1.0444	0.9657
	③ SED-DDP-ConvLSTM	1.9021	2.5782	0.8985	0.9808
	④ SED-DDP-LSTM	2.6234	3.3312	1.2542	0.9633
	⑤ SED-ConvLSTM	2.4902	3.0527	1.1908	0.9694
	⑥ ConvLSTM	2.2587	2.9644	1.0568	0.9705
	⑦ SVR	5.9349	6.9635	2.7454	0.8268
	⑧ ARIMA	9.3035	11.6494	4.4684	0.2080
Case 2	① SED-DDP-TVFEMD-ConvLSTM	**1.8607**	**2.4502**	**0.8872**	**0.9796**
	② SED-DDP-TVFEMD-LSTM	2.3514	3.0941	1.1289	0.9667
	③ SED-DDP-ConvLSTM	3.1509	4.0562	1.5147	0.9495
	④ SED-DDP-LSTM	3.4039	4.2216	1.6074	0.9327
	⑤ SED-ConvLSTM	4.1079	5.3361	1.9730	0.9097
	⑥ ConvLSTM	3.7394	4.4973	1.7763	0.9292
	⑦ SVR	9.1966	10.3618	4.3166	0.6704
	⑧ ARIMA	9.1321	10.9896	4.3730	0.3980

The best results are bold.

the outliers and noise information in the search engine big data, and screen the SEDs that have the ability to explain and predict the demand of air travel, so as to promote the prediction model to obtain better results

3. Based on the introduction of SED and decomposition, the performance of spatial–temporal ConvLSTM model (i.e., model ①) is more outstanding than that of non-spatial LSTM model (i.e., model ②). For example, the RMSE value of model ① is 37.74% lower than that of model ② in Case 1, and decreases by 20.81% in Case 2. Similarly, based on the introduction of SED without decomposition, the performance of spatial–temporal ConvLSTM model (i.e., model ③) is still better than that of non-spatial LSTM model (i.e., model ④). For example, the MAPE value of model ③ is reduced by 28.36% compared with model ④ in case 1, and decreases by 5.77% in Case 2. The comparison results show that ConvLSTM model can better grasp the spatial information in the input data to improve the performance of the model.

4. Based on the ConvLSTM model, the performance of the model with decomposition-ensemble strategy (i.e., model ①) is significantly better than that without this strategy (i.e., model ③). For example, compared with model ③, the MAE value of model ① is reduced by 20.78% in Case 1 and 40.95% in Case 2. Similarly, based on the LSTM model, the performance of the model with decomposition-ensemble strategy (i.e., model ②) is still better than that without this strategy (i.e., model ④). For example, compared with model ④, the WIA value of model ② increases by 0.25% in Case 1 and 3.65% in Case 2. The

comparison results show that the decomposition-ensemble strategy can avoid the adverse effects of the market environment to effectively improve the prediction precision.

5. Comparing the proposed framework (i.e., model ①) with the single SVR (i.e., model ⑦) and ARIMA (i.e., model ⑧), it can be found that the performance of the proposed framework is significantly better than that of the single models in both Cases. For example, in Case 1, the MAE value, RMSE value and MAPE value of model ① decrease by 74.61%, 70.80%, and 73.74%, respectively, compared with model ⑦, and decrease by 83.80%, 82.55%, and 83.87% compared with model ⑧. The comparison results fully prove the comprehensive advantages of the proposed framework, which not only plays the role of search engine data and spatial effect, but also combines the data analysis ability of different models.

In this paper, DM test was employed to further evaluate the superiority of the proposed forecasting framework from a statistical perspective by revealing the performance gaps between proposed model and the comparison models, respectively. The calculation results under Case 1 and Case 2, respectively, are presented in Table 3.

As illustrated in Table 3, the DM statistics of all comparison models are <0. Moreover, except for models ② and model ③, the proposed forecasting framework is significantly better than other comparative models at a significant level of 0.01 in Case 1; and have the same performance in Case 2 except for model ②. Conclusively, the proposed forecasting framework has excellently statistical validity.

TABLE 3 DM test results of the proposed model and the comparison models.

Comparison model	Case 1		Case 2	
	DM-statistic	Prob.	DM-statistic	Prob.
② SED-DDP-TVFEMD-LSTM	−1.9109*	0.0660	−2.4662**	0.0163
③ SED-DDP-ConvLSTM	−1.5064	0.1428	−4.6252***	0.0000
④ SED-DDP-LSTM	−3.5643***	0.0013	−5.2239***	0.0000
⑤ SED-ConvLSTM	−3.9033***	0.0005	−5.0077***	0.0000
⑥ ConvLSTM	−2.8186***	0.0086	−6.1822***	0.0000
⑦ SVR	−5.5977***	0.0000	−11.6820***	0.0000
⑧ ARIMA	−6.2229***	0.0000	−9.2150***	0.0000

*** is the 0.01 significance level, ** is the 0.05 significance level, * is the 0.1 significance level, and the MSE is taken as the loss function.

Deep insights

The previous analysis of the experimental results proved that the application of Internet big data and spatial information has a prominent effect on improving the accuracy of demand forecasting. Based on this research achievements, some deep insights on how the relevant bodies can benefit from it have produced.

For the government, they can organize the establishment of a tourism big data information sharing platform to achieve the linkage of public health, crisis early warning, crisis decision-making and other public management systems. Based on the specific predicted demand, a comprehensive and specific emergency response plan need be prepared to improve the public risk prevention and control capability. In addition, government is advised to strengthen the positive publicity of the effect of the safe travel policy to reduce the public's concern about the risk of long-distance travel. Meanwhile, they can strengthen the scientific literacy education related to the crisis by Internet, further enhance the citizens' scientific cognition of the effectiveness of crisis prevention and control, so as to alleviate their negative emotions. Additionally, based on the capacity of each scenic spot and the predicted results, government should not only arrange corresponding health facilities, but also complete the advance scheduling of resources (e.g., public transport, consulting services and medical services, etc.), to provide tourists with a sense of security for traveling. And preferential policies also need to be introduced to stimulate the tourism industry and drive economic development periodically.

For tourism companies, they better to fully realize information sharing among tourism enterprises, which is conducive to reducing the cost of enterprise information collection, so as to make accurate response to sudden crises. Moreover, according to the characteristics of varying demands, tourism service quality and passenger satisfaction can be improved by optimize resource allocation, strengthen human resource training and reshape service processes. And they can further analyze and respond new travel needs of the public, such as safety, health and convenience, through timely adjustment and optimization of the enterprise's operational objectives and work ideas based on the scientific prediction of travel demand. Meanwhile, it is critical to pay attention to the destinations with high travel intention of passengers, deepen the exploration of new tourism resources, or effectively integrate existing tourism resources combined with big data mining technology. This method caters to people's inherent pursuit of novelty, which is beneficial to the transformation of public positive emotions.

Conclusion

Nowadays, the social instability factors emerge in endlessly, has led to the global social psychological problems. In order to give full play to the role of long-distance travel in struggle against public anxiety and negative, starting from the search behavior related to travel behavior, this paper put forward an ensemble forecasting framework for air travel demand based on big data mining technology and artificial intelligence algorithm, and applies it to predict the air travel demand of Shanghai Pudong International Airport. The experimental results show that the proposed framework, adding search engine data with effective data preprocessing, employing the superior ConvLSTM model to mine the spatial–temporal information in the input data, and combining the advantages of different models, can alleviate the inherent content complexity of big data and obtain more satisfactory prediction results than the comparison models. The high performance of the proposed forecasting framework means that individuals and relevant departments can grasp the travel demand of passengers in advance, and make reasonable short-term planning and risk control for airports, scenic spots, hotels and so on, which greatly improves the sense of safety of passengers. On the basis of the achievements of this study, several suggestions about the formulation of policies and measures for the government and tourism companies in the unstable era are put forward for reference, which thoroughly provide a fuel for promoting travel safety, tourism revitalization and public anxiety alleviation.

In addition, although the proposed model has achieved good prediction results, there are still gaps to be filled. For example, this paper only conducted empirical research on the relevant actual data of Shanghai Pudong Airport, and more airport samples should be considered in the subsequent research. Moreover, the framework included only search engine data. In future researches, more types of big data can be incorporated into the model, such as news topics and comments on social media. Finally, this paper employed the

trial and error method to determine the parameters of ConvLSTM model, which had a negative impact on the stability and convenience of the model. In the subsequent research, advanced optimization algorithms can be used to achieve automatic optimization of parameters.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://github.com/QingZ96/Air-Passenger-Demand-Forecasting.

Author contributions

XL devised the study and collected the data. CH analyzed the data and wrote the manuscript. MY and WZ helped with writing and editing the manuscript. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aaronson, D., Brave, S. A., Butters, R. A., Fogarty, M., Sacks, D. W., and Seo, B. (2022). Forecasting unemployment insurance claims in realtime with Google trends. *Int. J. Forecast.* 38, 567–581. doi: 10.1016/j.ijforecast.2021.04.001

Chen, Y. C., and Huang, W. C. (2021). Constructing a stock-price forecast CNN model with gold and crude oil indicators. *Appl. Soft Comput.* 112:107760. doi: 10.1016/j.asoc.2021.107760

Chen, C. C., Zou, S. W., and Gao, J. (2020). Towards the recovery mechanisms of leisure travel experiences: does the length of vacation matter? *J. Travel Tour. Mark.* 37, 636–648. doi: 10.1080/10548408.2018.1525469

Chhetri, A., Arrowsmith, C., Chhetri, P., and Corcoran, J. (2013). Mapping spatial tourism and hospitality employment clusters: an application of spatial autocorrelation. *Tour. Anal.* 18, 559–573. doi: 10.3727/108354213X13782245307830

Diebold, F. X., and Mariano, R. S. (1995). Comparing predictive accuracy. *J. Bus. Econ. Stat.* 13, 134–144.

Fildes, R., Wei, Y. Q., and Ismail, S. (2011). Evaluating the forecasting performance of econometric models of air passenger traffic flows using multiple error measures. *Int. J. Forecast.* 27, 902–922. doi: 10.1016/j.ijforecast.2009.06.002

Fu, Y., Ren, Z., Wei, S., Wang, Y., Huang, L., Jia, F. (2022). Ultra-short-term power prediction of offshore wind power based on improved LSTM-TCN model. *Proc. CSEE* 42, 1–12. doi: 10.13334/j.0258-8013.pcsee.210724

Hannah, U., and Lisa, A. H. (2020). Exploring motivations and barriers for long-distance trips of adult women Vermonters. *Travel Behav. Soc.* 21, 37–47. doi: 10.1016/j.tbs.2020.05.007

Hu, M. M., Li, H. Y., Song, H. Y., Li, X., and Law, R. (2022). Tourism demand forecasting using tourist-generated online review data. *Tour. Manag.* 90:104490. doi: 10.1016/j.tourman.2022.104490

Huang, B., and Hao, H. (2021). A novel two-step procedure for tourism demand forecasting. *Curr. Issue Tour.* 24, 1199–1210. doi: 10.1080/13683500.2020.1770705

Huang, X. K., Zhang, L. F., and Ding, Y. S. (2017). The Baidu index: uses in predicting tourism flows –a case study of the Forbidden City. *Tour. Manag.* 58, 301–306. doi: 10.1016/j.tourman.2016.03.015

Jiao, X. Y., Li, G., and Chen, J. L. (2020). Forecasting international tourism demand: a local spatiotemporal model. *Ann. Tour. Res.* 83:102937. doi: 10.1016/j.annals.2020.102937

Jimenez, J. A. C., Torre, M. G. M. V., and Millán, M. G. D. (2022). Enotourism in southern Spain: the Montilla-Moriles PDO. *Int. J. Environ. Res. Public Health* 19, 1–21. doi: 10.3390/ijerph19063393

Jin, F., Li, Y. W., Sun, S. L., and Li, H. (2020). Forecasting air passenger demand with a new hybrid ensemble approach. *J. Air Transp. Manag.* 83:101744. doi: 10.1016/j.jairtraman.2019.101744

Lai, L., Ding, S. Q., Zhong, Z. Q., Mao, P., Sun, N., and Zheng, F. (2022). Association between positive mental character and humanistic care ability in Chinese nursing students in Changsha, China. *Front. Psychol.* 13:896415. doi: 10.3389/fpsyg.2022.896415

Li, H. L., Li, H. T., and Li, Z. (2022). Air passenger demand forecasting based on strategy of dual decomposition and reconstruction. *J. Comput. Appl.* Available at: https://kns.cnki.net/kcms/detail/51.1307.TP.20220310.1114.004.html

Li, H., Li, Z., and Mo, W. (2017). A time varying filter approach for empirical mode decomposition. *Signal Process.* 138, 146–158. doi: 10.1016/j.sigpro.2017.03.019

Li, X., Lv, B., Zeng, P., Liu, J. (2017). Tourism prediction using web search data based on CLSI-EMD-BP. *Syst. Eng. Theory Pract.* 37, 106–118.

Liang, X. Z., Zhang, Q., Hong, C. X., Niu, W., and Yang, M. (2022). Do internet search data help forecast air passenger demand? Evidence from China's airports. *Front. Psychol.* 13:809954. doi: 10.3389/fpsyg.2022.809954

Liu, Z. K., Jiang, P., Wang, J. Z., and Zhang, L. (2022). Ensemble system for short term carbon dioxide emissions forecasting based on multi-objective tangent search algorithm. *J. Environ. Manag.* 302:113951. doi: 10.1016/j.jenvman.2021.113951

Liu, H., Liu, Y., Li, G., and Wen, L. (2021). Tourism demand nowcasting using a LASSO-MIDAS model. *Int. J. Contemp. Hosp. Manag.* 33, 1922–1949. doi: 10.1108/IJCHM-06-2020-0589

Liu, H. H., Xiao, Y., Wang, B., and Wu, D. (2021). Effects of tourism development on economic growth: an empirical study of China based on both static and dynamic spatial Durbin models. *Tour. Econ.* 28, 1888–1913. doi: 10.1177/13548166211021175

Lu, X. Q., and Lin, Z. B. (2021). COVID-19, economic impact, mental health, and coping behaviors: a conceptual framework and future research directions. *Front. Psychol.* 12:759974. doi: 10.3389/fpsyg.2021.759974

Macqueen, J. B. (1967). "Some methods for classification and analysis of multivariate observations," in *Proc 5th Berkeley Symposium Mathematics Statist and Probability*, 281–297.

Madziwa, L., Pillalamarry, M., and Chatterjee, S. (2022). Gold price forecasting using multivariate stochastic model. *Resour. Policy* 76:102544. doi: 10.1016/j.resourpol.2021.102544

Mei, S. L., Meng, C. C., Hu, Y. Y., Guo, X., Lv, J., Qin, Z., et al. (2022). Relationships between depressive symptoms, interpersonal sensitivity and social support of employees before and during the COVID-19 epidemic: a cross-lag study. *Front. Psychol.* 13:742381. doi: 10.3389/fpsyg.2022.742381

Montiel, C., Radziszewski, S., Prilleltensky, I., and Houle, J. (2021). Fostering positive communities: a scoping review of community-level positive psychology interventions. *Front. Psychol.* 12:720793. doi: 10.3389/fpsyg.2021.720793

Ouchen, A., and Montargot, N. (2021). Non-spatial and spatial econometric analysis of tourism demand in a panel of countries around the world. *Spat. Econ. Anal.* 05, 1–22. doi: 10.1080/17421772.2021.1940256

Ouyang, X., Wang, J. C., Bai, L., and Gong, W. (2021). Emotion, trust and action: experimental study on localized communication effects of constructive journalism. *Chin. J. Journalism Commun.* 8, 73–89.

Shi, X. J., Chen, Z. R., Wang, H., Yeung, D. Y., Wong, W. K., Woo, W. C., et al. (2015). "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," in *Advances in Neural Information Processing Systems. Vol. 28* MIT Press, 802–810.

Sun, Y. P., Li, H. N., and Cao, Y. N. (2022). Effects of COVID-induced public anxiety on European stock markets: evidence from a fear-based algorithmic trading system. *Front. Psychol.* 12:780992. doi: 10.3389/fpsyg.2021.780992

Tavitiyaman, P., Qu, H. L., Tsang, W.-S. L., and Lam, C. W. R. (2021). The influence of smart tourism applications on perceived destination image and behavioral intention: the moderating role of information search behavior. *J. Hosp. Tour. Manag.* 46, 476–487. doi: 10.1016/j.jhtm.2021.02.003

Xie, G., Qian, Y. T., and Wang, S. Y. (2021). Forecasting Chinese cruise tourism demand with big data: an optimized machine learning approach. *Tour. Manag.* 82:104208. doi: 10.1016/j.tourman.2020.104208

Xiong, H. L., Zhu, R. J., Ji, H., Fan, Z., and Xu, P. (2021). Air passenger index prediction method based on MI-SVR mode. *Control Decis.* 36, 1619–1626. doi: 10.13195/j.kzyjc.2019.1446

Yang, Y., and Zhang, H. L. (2019). Spatial-temporal forecasting of tourism demand. *Ann. Tour. Res.* 75, 106–119. doi: 10.1016/j.annals.2018.12.024

Yao, L. F., Ma, R. M., and Wang, H. (2021). Baidu index-based forecast of daily tourist arrivals through rescaled range analysis, support vector regression, and autoregressive integrated moving average. *Alex. Eng. J.* 60, 365–372. doi: 10.1016/j.aej.2020.08.037

Yu, C. X., and Lu, L. (2008). A comment on overseas studies of tourists' preferences for tourism in recent years. *J. Anhui Normal Univ. (Nat. Sci.)* 31, 590–595.

Zhang, Y. X., Wang, S. Z., Chen, B., Cao, J., and Huang, Z. (2021). TrafficGAN: network-scale deep traffic prediction with generative adversarial nets. *IEEE Trans. Intell. Transp. Syst.* 22, 219–230. doi: 10.1109/TITS.2019.2955794

Zheng, W. Q., Peng, X. G., Lu, D., Zhang, D., Liu, Y., Lin, Z., et al. (2017). Composite quantile regression extreme learning machine with feature selection for short-term wind speed forecasting: a new approach. *Energy Convers. Manage.* 151, 737–752. doi: 10.1016/j.enconman.2017.09.029