Tourism demand forecasting: a deep learning model based on spatial-temporal transformer

Jiaying Chen, Cheng Li, Liyao Huang and Weimin Zheng

Abstract

Purpose – Incorporating dynamic spatial effects exhibits considerable potential in improving the accuracy of forecasting tourism demands. This study aims to propose an innovative deep learning model for capturing dynamic spatial effects.

Design/methodology/approach – A novel deep learning model founded on the transformer architecture, called the spatiotemporal transformer network, is presented. This model has three components: the temporal transformer, spatial transformer and spatiotemporal fusion modules. The dynamic temporal dependencies of each attraction are extracted efficiently by the temporal transformer module. The dynamic spatial correlations between attractions are extracted efficiently by the spatial transformer module. The extracted dynamic temporal and spatial features are fused in a learnable manner in the spatiotemporal fusion module. Convolutional operations are implemented to generate the final forecasts.

Findings – The results indicate that the proposed model performs better in forecasting accuracy than some popular benchmark models, demonstrating its significant forecasting performance. Incorporating dynamic spatiotemporal features is an effective strategy for improving forecasting. It can provide an important reference to related studies.

Practical implications – The proposed model leverages high-frequency data to achieve accurate predictions at the micro level by incorporating dynamic spatial effects. Destination managers should fully consider the dynamic spatial effects of attractions when planning and marketing to promote tourism resources.

Originality/value – This study incorporates dynamic spatial effects into tourism demand forecasting models by using a transformer neural network. It advances the development of methodologies in related fields.

Keywords Tourism demand prediction, Dynamic spatial effects, Deep learning model, Transformer

Jiaying Chen is based at the Business and Tourism School, Sichuan Agricultural University, Chengdu, China. Cheng Li, Liyao Huang and Weimin Zheng are all based at the School of Management, Xiamen University, Xiamen, China.

Received 1 May 2023
Revised 10 July 2023
23 October 2023
Accepted 23 October 2023

Received 1 May 2023
Revised 10 July 2023
23 October 2023
Accepted 23 October 2023

This research received grants from the National Natural Science Foundation of China (No. 71971179), Natural Science Foundation of Fujian Province, China (No. 2020J01033), Sichuan Center for rural development research (No. CR209), Sichuan Tourism Development Research Center (No. LY22-29).

Author contribution: Jiaying Chen: Data collection, critical revision of the article, drafting the article. Cheng Li: Conception or design of the article, data analysis and interpretation, drafting the article. Liyao Huang: Conception or design of the work, data analysis and interpretation, drafting the article. Weimin Zheng: Final approval of the version to be published, critical revision of the article, drafting the article.
Previsión de la demanda turística: Un modelo de aprendizaje profundo basado en el transformador espaciotemporal

Resumen

Objetivo: La incorporación de efectos espaciales dinámicos ofrece un considerable potencial para mejorar la precisión de la predicción de la demanda turística. Este estudio propone un modelo innovador de aprendizaje profundo para capturar los efectos espaciales dinámicos.

Diseño/metodología/enfoque: Se presenta un novedoso modelo de aprendizaje profundo basado en la arquitectura transformadora, denominada red de transformador espaciotemporal.Este modelo tiene tres componentes: el transformador temporal, el transformador espacial y los módulos de fusión espaciotemporal. El módulo transformador temporal extrae de manera eficiente las dependencias temporales dinámicas de cada atracción. El módulo transformador espacial extrae eficientemente las correlaciones espaciales dinámicas entre las atracciones. Las características dinámicas temporales y espaciales extraídas se fusionan de manera que se puede aprender en el módulo de fusión espaciotemporal. Se aplican operaciones convolucionales para generar las previsiones finales.

Conclusiones: Los resultados indican que el modelo propuesto obtiene mejores resultados en la precisión de las previsiones que algunos modelos de referencia conocidos, lo que demuestra su importante capacidad de previsión. La incorporación de características espaciotemporales dinámicas supone una estrategia eficaz para mejorar las previsiones. Esto puede proporcionar una referencia importante para estudios afines.

Implicaciones prácticas: El modelo propuesto aprovecha los datos de alta frecuencia para lograr predicciones precisas a nivel micro incorporando efectos espaciales dinámicos. Los gestores de destinos deberían tener plenamente en cuenta los efectos espaciales dinámicos de las atracciones en la planificación y marketing para la promoción de los recursos turísticos.

Originalidad/valor: Este estudio incorpora efectos espaciales dinámicos a los modelos de previsión de la demanda turística mediante el empleo de una red neuronal transformadora. Supone un avance en el desarrollo de metodologías en campos afines.

Palabras clave: Predicción de la demanda turística, Efectos espaciales dinámicos, Modelo de aprendizaje profundo, Transformador

Tipo de papel: Trabajo de investigación

1. Introduction

Significant developments have occurred in the tourism industry, while vulnerabilities have emerged (Li and Jiao, 2020). Economic crises, a global pandemic and unforeseen events have forced the industry to face high levels of uncertainty, leading to significant fluctuations in tourism demand (Hu, 2022; Song et al., 2019). These increased uncertainties have incurred substantial revenue losses and misallocations of resources, and even reduced tourism experience (Law et al., 2019). Precise demand forecasting is widely valued by industry practitioners and academics, because it can help managers make adequate foresights for the future and take advanced countermeasures (Wu et al., 2017). However, complex and nonlinear fluctuation patterns hidden in tourism demand time series may impose high requirements for the development of forecasting models (Xie et al., 2020). Introducing further advanced models and strategies to improve the accuracy of tourism demand forecasting is a common and sustained interest of researchers.

Studies have driven this field toward model innovation and predictor variable incorporation to improve forecasting accuracy continuously (Huang et al., 2022). From traditional time series and econometric models to currently popular artificial intelligence (AI)-based models, the feature extraction capacities of models, particularly for nonlinear and complex features, are constantly evolving (Huang et al., 2022). Variable incorporation has also received increasing attention. Factors, such as price (Huang et al., 2022), search index (Yi-Chung and Geng, 2022) and online review (Wu et al., 2021), have been integrated into model construction to explain demand fluctuation and improve forecasting performance. Spatial effects, which refer to the direct or indirect effects of tourism activities in one destination on those of neighboring destinations, serve as a notable predictor variable (Yang and Wong, 2012). Several studies have demonstrated the significant potential of integrating spatial effects into forecasting improvement (Yang and Zhang, 2019; Zheng et al., 2021).
Previous studies have adopted two strategies for incorporating spatial effects into modeling tourism demand forecasting. One strategy aims to improve traditional econometric models (Jiao et al., 2020; Wen et al., 2018) by adding spatial lag terms. The other strategy aims to introduce deep learning models based on methods with spatial feature extraction capacity, such as convolutional neural networks and graph convolutional networks (GCNs) (Bi et al., 2023; Li et al., 2022). These spatiotemporal models can improve forecasting accuracy compared with models that do not consider spatial effects, but they still have limitations. For temporal feature extraction, the lag order selection and fixation of traditional spatial econometric models lead to their shortcoming in dealing with high-frequency, nonstationary and nonlinear data, limiting their applications (Law et al., 2019). Moreover, long short-term memory (LSTM) is one of the most popular models (Law et al., 2019). However, when forecasting is required for extremely high-frequency data (e.g. hourly forecasts), LSTM may not cover a sufficiently long period of fluctuations, and thus, cannot accurately capture all fluctuations in the time series (Vaswani et al., 2017).

When capturing spatial features, spatial relationships among attractions are assumed static in existing studies because they adopt predetermined spatial weighting matrices to capture spatial information, leading to misestimations and negatively affecting their prediction performance (Zhao et al., 2022). Spatial relationships among tourist attractions are dynamic. They vary by factors, such as transportation, weather conditions and cooperation and competition (Almeida et al., 2021). For example, the spatial correlation among tourist attractions varies during different periods (e.g. morning and evening peaks, weekdays and weekends and holidays). Therefore, spatial relationships among attractions should be dynamically described during feature extraction to better extract spatial information. If dynamic changes in tourism demands are disregarded, then the spatial effects of tourism may only be partially captured and explained (Zhou et al., 2020). Therefore, dealing with dynamics in terms of spatial effects and time dependence remains an urgent issue when constructing forecasting models.

This study aims to introduce an innovative deep learning model, called the spatiotemporal transformer network (STTN), to fill the aforementioned research gaps. This model has three components: the temporal transformer, spatial transformer and spatiotemporal fusion modules. The transformer is the core of the first two modules. It can effectively extract the temporal dependencies of each attraction and the spatial correlations between attractions on the basis of a multiheaded self-attention mechanism. The transformer can treat all parts of the input data equally regardless of the temporal distance. It can effectively handle long-term dependencies in a demand time series. The transformer can also calculate attention scores between input variables. Such scores can be regarded as relationships between attractions in this study. Spatial relationships between attractions at different periods can be automatically learned as input data changes and used for the further dynamic extraction of spatial features (Yan et al., 2021). The extracted temporal and spatial features are fused in a learnable manner in the spatiotemporal fusion module. Convolutional operations are then applied to produce the final forecasts. The validation of this model is verified by forecasting the hourly demands of several attractions in Beijing. The forecasting results indicate that proposed model is superior to some popular benchmark models, demonstrating significant forecasting performance. Thus, incorporating dynamic spatiotemporal features is an effective method for improving forecasting. It can provide an important reference to related studies.

2. Literature review

2.1 Tourism demand forecasting

The prediction of tourism demands is important in planning and allocating resources, and thus, it has elicited extensive attention from academics and practitioners (Law et al., 2019; Wu et al., 2023). Scholars have devoted considerable effort to updating forecasting models.
and incorporating appropriate explanatory variables to improve the accuracy of tourism demand forecasting continuously (Huang et al., 2022).

Researchers have initially developed time series and econometric models to predict tourism demand. Time series models achieve forecasting by fitting trends from historical data and extrapolating such trends (Bufalo and Orlando, 2023; Jiao et al., 2020). Econometric models achieve forecasting by quantifying the relationship between influencing factors and demands (Zheng et al., 2021). Both models share the disadvantage of not capturing the nonlinearity and volatility of time series data, such as daily demand data (Bi et al., 2023). The application of AI-based models effectively addresses this challenge. AI-based models exhibit the advantages of handling nonlinear relationships and extracting features automatically (Bi et al., 2021), with LSTM as the dominant method. However, no single model can perfectly fit all forecasting scenarios. Demand forecasts with finer granularity (e.g. hourly forecasts) are being increasingly emphasized, but even advanced AI-based models find these tasks challenging. Fine-grained forecasts imply that temporal dependencies are extracted over a longer time horizon, imposing more stringent requirements on the long-term feature processing capability of a model. LSTM functions well in long-term modeling, but it still fails to deal with such time series perfectly and cannot completely avoid gradient vanishing and exploding problems (Li et al., 2019). As a popular AI model, the transformer exhibits potential for improving forecasting performance. The core of the transformer is its multi-head attention, through which all parts of the input data are treated equally regardless of temporal distance. The transformer can easily handle long-term dependencies. It has been adopted to process forecasting tasks in many fields, such as transportation, energy and medicine; it has demonstrated unparalleled forecasting performance (Yan et al., 2021). However, such models have rarely been discussed in tourism demand forecasting. Future studies should consider this subject.

In addition to developing new models, incorporating more appropriate predictor variables into model construction is a valid approach for enhancing the accuracy of tourism demand forecasting. These variables involve valuable contents about future trends of tourism demand (Pan and Yang, 2017). Therefore, incorporating them into a model can increase the precision and explanatory power of the model, provided that they are associated with tourism demand (Peng et al., 2014). Explanatory variables in tourism demand forecasting models cover two factors: economic factors, such as price (Li et al., 2005), gross disposable income (Onafowora and Owoye, 2012) and income levels (Song et al., 2011); and noneconomic factors, such as search engine (Li and Law, 2020), Web traffic (Yang et al., 2014) and online review (Wu et al., 2021).

In summary, deep learning models are currently favored as updating forecasting models. Their major advantage is their ability to extract nonlinear relationships hidden in data automatically, demonstrating high accuracy in predicting tourism demand (Song et al., 2019). For the variables involved in the model, studies have attempted to incorporate spatial effects into tourism demand forecasting, achieving good results (Jiao et al., 2020; Wen et al., 2018; Zheng et al., 2021). Good results were achieved because spatial dependence may exist among attractions within a destination due to supply interactions between attractions and tourists’ multi-destination travel (Yang and Zhang, 2019). Therefore, incorporating spatial effects presents the potential to improve forecasting accuracy. Related studies remain few, and the extraction of spatial effects in existing studies exhibits some limitations. Accordingly, attempts should be made to promote relevant literature.

### 2.2 Spatial effects in tourism demand forecasting

Spatial effects describe the effects of a geographic location and the interrelationships among different regions on economic and social development (Jiao et al., 2020). They consider interactions among different regions, including the direct and indirect effects of destination tourism growth on tourism development in neighboring regions...
Tourists prefer to visit multiple destinations during a single trip; thus, situations where similar demand change patterns may exist among attractions within a destination, demonstrating evident spatial effects (Bi et al., 2023). The integration of spatial information into tourism demand forecasting models is considered an effective strategy for improving forecasting accuracy by tourism researchers (Yang and Zhang, 2019). This strategy has become a new hotspot in tourism demand forecasting research. Therefore, tourism researchers must explore and exploit spatial effects in depth to improve forecasting performance.

A review of the extant literature reveals that several studies have attempted to incorporate spatial effects into tourism demand forecasting models to improve their forecasting performance. Yang and Zhang (2019) and Wen et al. (2018) conducted tourism demand forecasting on the basis of a global spatial-temporal model in which they assumed that spatial interactions were constant for all destinations. Jiao et al. (2020) applied a local spatial-temporal model for forecasting on the basis of previous work, enabling unique estimates of the spillover effect for individual destinations. Jiao et al. (2021) considered the spatial correlation between tourism demands and explanatory variables. The aforementioned studies incorporate spatial effects into forecasting models by constructing econometric models. Traditional econometric models are extended by imposing spatial structures (e.g. spatial weighting matrices) on the time series associated with different locations; the spatial weighting matrices specify the intercorrelation patterns of each spatial unit in advance (Yang and Zhang, 2019).

In addition to econometric models, Zheng et al. (2021) and Li et al. (2022) captured spatial effects by using deep learning models, such as GCN and LSTM. These models must provide spatial weights expressed by distance or other relations in advance when capturing spatial effects. In summary, researchers deal with spatial effects by assuming in advance and fixing the estimated coefficients associated with spatial weight matrices, leading to misestimations and negatively affecting forecasting performance (Zhao et al., 2022). A single spatial weight may not represent the complex time-varying pattern of tourism demands, producing inaccurate hourly forecasts. Therefore, dealing with dynamic spatial effects in modeling tourism demand forecasting remains unresolved.

This study intends to address this problem by proposing a novel deep learning model based on a transformer. This model captures the dynamic temporal and spatial dependencies, which are incorporated into the forecasting model to improve the accuracy of demand forecasting. This model can predict the demands for multiple attractions within a tourist destination with hourly forecasting accuracy. Finally, a prediction experiment is conducted to evaluate the performance of the model by using the hourly demand data of 30 attractions in Beijing.

3. Methodology

A novel deep learning model, STTN, is proposed in this study for capturing the spatial correlations between the tourist attractions in an area and the temporal dependencies of each tourist attraction. STTN is constructed on the basis of the transformer, which was originally proposed for the machine translation problem (Vaswani et al., 2017). The transformer, which is an attention-based model, exhibits powerful feature extraction ability, and it has been introduced in many fields (Liu et al., 2022).

3.1 Overview architecture

As depicted in Figure 1, the proposed model has three parts. STTN requires the following steps to generate accurate forecasts. Given the historical tourism demand time series \( X \), two transformer modules receive and address \( X \) simultaneously to capture temporal dependency \( TD(x) \) and spatial correlation \( SC(x) \). The extracted features, i.e. \( TD(x) \) and \( SC(x) \),...
SC(x), are transformed through a two-layer linear function to increase the representation ability of the model. The processed TD(x) and SC(x) are integrated into the fused spatiotemporal features FSTF(x) in a learnable manner. Two convolution operations are applied to FSTF(x) after layer normalization to generate the final forecasting results \( \hat{X} \).

Backpropagation is performed through the Adam optimizer to update the model parameters and reduce the loss value of the prediction results. Each module is introduced in detail in the subsequent sections.

### 3.2 Temporal transformer module

The transformer with a strong learning capacity for complex and time-varying information is applied in this module (Zhang et al., 2019) to extract the temporal dependency of tourism demand time series effectively. This transformer treats all parts of the input data equally, regardless of temporal distance (Shen and Wang, 2022). Consequently, it can process longer input variables than the LSTM model and has been widely adopted in time series forecasting and classification (Liu et al., 2022). The core component of this transformer is the self-attention mechanism that performs feature extraction in accordance with the following steps. First, three different vectors with dimension \( d_k \), namely, query (Q), key (K) and value (V), are transformed from the input data. Second, scaled dot-product attention is applied to Q and K to calculate attention scores. The scores are normalized using the softmax function. The scores are weighted with V to generate the self-attention results. The detailed calculation processes are defined in equations (1) and (2), where \( W_Q, W_K \) and \( W_V \) are parameter matrices:

\[
Q = XW_Q, \; K = XW_K, \; V = XW_V
\]

\[
\text{Attention}(Q,K,V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V
\]

The framework of the proposed transformer is depicted in Figure 2. In contrast with the original transformer, which is an encoder–decoder framework (Vaswani et al., 2017), the proposed transformer adopts only encoder layers. A similar transformer framework was used in Castangia et al. (2022) and Hu and Xiao (2022). In the proposed transformer, the extraction of temporal features comprises the following steps. First, given that the self-attention mechanism is order-agnostic to the input data, positional encoding is adopted on the input data. In accordance with Vaswani et al. (2017), wave functions are adopted in this study to mark the relative position of the input data. The detailed definition of positional encoding is described in equations (3) and (4), where \( \text{pos} \) denotes the position within the input data, and \( i \) indicates the index of the embedding value. Second, multiple self-attention modules, i.e. multi-head self-attention, are implemented on the encoded
input data to extract temporal features. In accordance with the principle of self-attention mechanism, Q, K and V are first divided into h parts (heads) with dimension $d_k/h$. Then, equation (2) is applied to each head to calculate the corresponding self-attention results. The results of h heads are concatenated into the final results. The processes can be summarized by equation (5), where $W^T$ is the parameter matrix. Third, linear transformation is applied to the results of the multi-head self-attention mechanism to generate the output vectors, i.e. temporal dependency hidden in input data. Layer normalization is implemented for the residual connection between the multi-head self-attention layer and the feed-forward layer:

$$PE_{pos,2i} = \sin\left(\frac{pos}{10000^{2i/d_k}}\right)$$ \hspace{1cm} (3)

$$PE_{pos,2i+1} = \cos\left(\frac{pos}{10000^{2i/d_k}}\right)$$ \hspace{1cm} (4)

$$\text{Multihead}(Q, K, V) = \text{Concat}(\text{head}_1, \text{head}_2, \ldots, \text{head}_i) \cdot W^T$$ \hspace{1cm} (5)

### 3.3 Spatial transformer module

Previous studies have demonstrated that spatial correlations exist among the tourism demands of tourist destinations (Jiao et al., 2021; Yang and Zhang, 2019). Considering spatial effects is beneficial for improving forecasting accuracy (Wen et al., 2018). In addition to capturing temporal features, the strong spatial feature capture capacity of the...
transformer has been validated in previous studies (Yan et al., 2021). A spatial transformer module is constructed to extract spatial correlations among tourist attractions in the same region. A transformer framework similar to the one used in the temporal module is adopted to extract spatial features. By contrast, attention scores are calculated with different vector dimensions. The proposed transformer in temporal module mostly calculates attention scores among various time points in each input sequence. In the spatial module, scores among input sequences at the same time are calculated. In tourism forecasting, such scores can be regarded as spatial relationships among attractions. Dynamic spatial relationships can be obtained as data move along the time axis. Accordingly, dynamic spatial correlations among attractions can be effectively captured.

### 3.4 Spatiotemporal fusion module

This module fuses the extracted temporal dependencies $TD(x)$ and spatial correlations $SC(x)$ and generates the final forecasts. Temporal and spatial features exert varying effects on forecasting performance. Thus, they are fused in a learnable manner defined in equation (6), where $W_1$ and $W_2$ are learnable parameters. After feature fusion, layer normalization is applied to $FSTF(x)$ to increase aggregation speed and decrease the risk of gradient disappearance and explosion. Finally, convolution operations are applied to $FSTF(x)$ to generate forecasts:

$$FSTF(x) = W_1TD(x) + W_2SC(x)$$

(6)

### 4. Experimental study

#### 4.1 Data collocation and preprocessing

A forecasting experiment is conducted in Beijing, China, to demonstrate the effectiveness of the proposed model. With its well-developed urban transportation and dense distribution of attractions, Beijing is an ideal area for examining spatial effects in tourism demands. Furthermore, data on the demands of Beijing attractions are available from the Beijing Tourism Network (www.visitbeijing.com.cn), which offers valuable information support for this study. Considering data availability and attraction popularity, only the hourly demand data of 30 attractions from July 1, 2021, to December 31, 2021, are selected for the experiment (unit: 10,000). In accordance with the operating hours of most attractions, the demand time series is constructed by keeping only data from 09:00 to 18:00 each day (Zheng et al., 2021). Figure 3 shows the locations of attractions and the demand data for several attractions. Hourly demand data have many complex and nonlinear patterns. The tourism demands of these attractions exhibit similar characteristics in terms of trends and periods, indicating that demand interactions may exist among attractions. The sample data are divided into training (1,008 observations), validation (324 observations) and testing (324 observations) data sets with a ratio of 6:2:2 for model training and validation. The three sets are used to train the model, select model parameters and evaluate model performance, respectively.

#### 4.2 Performance measures

The performance of the forecasting models is evaluated on the basis of three common criteria: mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE). The calculation formulas of the three criteria are presented by equations (7) to (9), where $N$ refers to the number of prediction samples (i.e. testing data set), and $y_n$ and $\hat{y}_n$ represent the true and forecasted tourism demand,
respectively. When the values of these criteria are smaller, the forecasting performance of the model is better:

\[
MAE = \frac{1}{N} \sum_{n=1}^{N} |y_n - \hat{y}_n| \tag{7}
\]

\[
RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (y_n - \hat{y}_n)^2} \tag{8}
\]

\[
MAPE = \frac{1}{N} \sum_{n=1}^{N} \frac{|y_n - \hat{y}_n|}{y_n} \tag{9}
\]

4.3 Benchmarks and parameter selection

Popular forecasting models are selected as benchmark models to demonstrate the superiority of the proposed model. Five non-spatial models and two spatial-temporal models are selected. The non-spatial models are as follows: exponential smoothing (ETS); seasonal autoregressive integrated moving average model (SARIMA) constructed by the auto.arima function in R; trigonometric exponential smoothing state space model with Box-Cox transformation, ARMA errors and trend and seasonal components (TBATS); vector autoregression (VAR); and LSTM. The spatial-temporal models are as follows: correlation time series-oriented LSTM with attention mechanism (CTS-LSTM-AM) incorporated in Zheng et al. (2021) and spatial-temporal fused GCN (ST-FGCN) introduced in Li et al. (2022).

The selection of appropriate parameters is important for the performance of deep learning models (Bi et al., 2020). For STTN, the parameters are learning rate, batch size, epoch, dropout rate, time step and amounts of cell units and heads. A detailed search for each parameter may result in considerable computational costs. Accordingly, the first four parameters are set appropriately in accordance with Zheng et al. (2021). The last three parameters are searched using the exhaustive grid search (GS) method. On the basis of MAPE values, GS chooses the optimal parameter combination from time step \( \in \{1,2,\ldots,15\} \), unit number \( \in \{2,4,8,\ldots,128\} \) and head number \( \in \{2,4,6,8\} \). The training and validation data

Source: Created by authors
sets are used. The former is first adopted to train STTN with different parameter combinations. Then, the trained model generates forecasts for the validation data set. Forecasting under each parameter combination is repeated five times. The average MAPE value is used for evaluation. Finally, the best parameter combination is selected and used to conduct further forecasts for the testing set. The parameters of the other benchmarks are similarly determined. The results are presented in Table 1.

### 4.4 Forecasting results and comparisons

Table 2 provides the average one-step-ahead forecasting results of the proposed model and the benchmarks with regard to the three criteria. The prediction performance of STTN is superior to all the benchmarks, demonstrating its effectiveness and superiority. The Diebold–Mariano (DM) test (Diebold and Mariano, 2002) is used to determine whether the differences between STTN and the benchmarks are statistically significant. The DM values are negative, indicating that STTN considerably outperforms the benchmarks in most scenarios. To present the superiority of STTN clearly, improvement rate (IR) is computed using equation (10). Table 2 provides the results:

\[
IR = \frac{\text{MAPE(model 2)} - \text{MAPE(model 1)}}{\text{MAPE(model 1)}}
\]

The proposed model generally achieves impressive predictive performance in terms of MAE, RMSE and MAPE, which are estimated as 0.317, 0.435 and 8.11%, respectively. The positive effect of incorporating spatial information into forecasting is demonstrated, because all the spatial-temporal models provide better forecasts than the non-spatial models. For example, CTS-LSTM-AM provides 12.60%, 11.10% and 16.77% improvements on the three criteria, respectively, compared with the single LSTM model. LSTM is the leading non-spatial model. As presented in the first two comparisons in Table 3, LSTM achieves a performance improvement of more than 20% in all three criteria compared with the two best traditional models (i.e. TBATS and VAR). This result indicates that deep learning models can extract temporal features from complex, nonlinear time series in a

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Optimal parameter setting of each model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model</strong></td>
<td><strong>Parameter setting</strong></td>
</tr>
<tr>
<td>ETS</td>
<td>(\alpha = 0.998, \gamma = 0.002)</td>
</tr>
<tr>
<td>SARIMA</td>
<td>(p = 5, d = 1, q = 0, P = 2, D = 0, Q = 0, s = 9)</td>
</tr>
<tr>
<td>TBATS</td>
<td>(\alpha = 0.018, \gamma_1 = -4.346, \gamma_2 = -1.77e - 4)</td>
</tr>
<tr>
<td>VAR</td>
<td>lag lengths = 9</td>
</tr>
<tr>
<td>LSTM</td>
<td>lr = 0.002, bs = 32, epoch = 100, dr = 0.5, ts = 9, un = 128</td>
</tr>
<tr>
<td>CTS-LSTM-AM</td>
<td>lr = 0.005, bs = 64, epoch = 100, dr = 0.3, ts = 9, un = 64</td>
</tr>
<tr>
<td>ST-FGCN</td>
<td>lr = 0.001, bs = 32, epoch = 200, dr = 0.1, ts = 9, un = 64</td>
</tr>
<tr>
<td>STTN</td>
<td>lr = 0.003, bs = 32, epoch = 100, dr = 0.1, ts = 9, un = 32, hn = 2</td>
</tr>
</tbody>
</table>

**Note:** \(lr, bs, dr, ts, un\) and \(hn\) denote learning rate, batch size, dropout rate, time step, unit number and head number, respectively

**Source:** Created by authors

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Average one-step-ahead forecasting performance of each model (unit: 10,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ETS</strong></td>
<td><strong>SARIMA</strong></td>
</tr>
<tr>
<td>MAE</td>
<td>0.094</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.146</td>
</tr>
<tr>
<td>MAPE</td>
<td>0.309</td>
</tr>
</tbody>
</table>

**Source:** Created by authors
more effective and accurate manner. The forecasting performance of spatial-temporal models also differs evidently. ST-FGCN presents improvements of nearly 20% over CTS-LSTM-AM in all three criteria with the same temporal feature extraction method. This finding suggests that GCN may be a better approach for spatial feature extraction than spatial weighted LSTM. The application of the transformer to temporal and spatial feature extraction results in remarkable forecasting improvement of the proposed model. For hourly demand data, temporal dependency extraction requires a long data span to cover its multiple fluctuation periods. LSTM performs excellently in capturing long-term dependencies, but it is still inferior to the transformer in processing longer input data (Vaswani et al., 2017). The transformer can adaptively learn spatial relationships among input sequences, and thus, it contributes to dynamic spatial feature extraction. More adequate and accurate spatial feature extraction can improve forecasting accuracy. The proposed model exhibits significant superiority in this study.

Future forecasts are performed for h = 3, 6 and 9 steps and then compared with those of the benchmark models to examine the performance of the proposed model in multi-step-ahead forecasting. Figure 4 presents the values of the mean MAPE of each model. STTN performs better in every horizon than the benchmark models, indicating that it is robust and perfectly suitable for one-step-ahead and multi-step-ahead forecasting tasks. Interestingly, VAR demonstrates better forecasting performance in multi-step-ahead

![Figure 4](source: Created by authors)

<table>
<thead>
<tr>
<th>Table 3</th>
<th>IR of several model pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comparison</strong></td>
<td><strong>IR(MAE) (%)</strong></td>
</tr>
<tr>
<td>LSTM vs TBATS</td>
<td>↑25.61</td>
</tr>
<tr>
<td>LSTM vs VAR</td>
<td>↑21.77</td>
</tr>
<tr>
<td>CTS-LSTM-AM vs LSTM</td>
<td>↑12.60</td>
</tr>
<tr>
<td>ST-FGCN vs CTS-LSTM-AM</td>
<td>↑22.66</td>
</tr>
<tr>
<td>STTN vs CTS-LSTM-AM</td>
<td>↑40.64</td>
</tr>
<tr>
<td>STTN vs ST-FGCN</td>
<td>↑23.24</td>
</tr>
</tbody>
</table>

**Source:** Created by authors
4.5 Model interpretation

The forecasting results signify that the proposed model can better capture spatial features for demand forecasting. In contrast with using static and pre-defined spatial relationship descriptions, the proposed model can adaptively learn spatial relationships among tourist attractions on the basis of the given data, dynamically capturing spatial features for forecasting. For example, the spatial relationships between the Forbidden City and other sample attractions are visualized at different periods (October 1, 3 and 6, 2021), as shown in Figure 5. The attractions that exert spatial influences on the Forbidden City are marked with red circles. A large circle corresponds to considerable spatial influence. Overall, the spatial relationships are dynamic because the influential attractions are different at the three periods. Geographical distance should not be the only criteria used to measure spatial relationships among attractions, because some neighboring attractions exert no significant spatial influence on the target on October 3 and 6. The influential attractions on October 1 are relatively few and concentrated, while those on October 3 and 6 exhibit different spatial relationships.

Figure 5  Spatial relationships learned from the demand data at different periods: October 1 (a), 3 (b) and 6 (c), 2021

Source: Created by authors
characteristics because the number of tourists always reaches its peaks in the middle of a holiday (October 1 to 7 is the National Day holiday in China). These tourists are distributed throughout the city. Therefore, spatial effects can be observed on a larger scale.

5. Conclusion

Accurate tourism demand forecasting has been increasingly emphasized by tourism practitioners because of its important role in formulating relevant strategies and decisions (Zheng et al., 2021). Improving the accuracy of demand forecasting may present several challenges. The first challenge is attributed to the intangible and experiential nature of tourism, resulting in travel decisions becoming a common scenario for uncertainty decisions (Yoo and Chon, 2008). In accordance with bounded rationality theory (Simon, 2000), people are influenced by various factors in uncertain situations, particularly irrational factors, such as emotion and cognition (Isen and Patrick, 1983). Each tourist has varying cognitive levels and emotional states. Consequently, travel demands exhibit a variety of complex nonlinear patterns, especially in high-frequency data.

In the second challenge, time constraint is considered one of the most important factors (Huang et al., 2020) that affects tourism demands in accordance with time geography theory (Hgerstraand, 1970). Spatial relationships among destinations change continuously over time. They are influenced by a variety of factors, such as transportation, weather, competition and cooperation. Therefore, temporal dynamics is crucial for demand forecasting. This study introduces a novel spatial-temporal model based on the transformer architecture to address the aforementioned challenges. A forecasting experiment is conducted as performance evaluation by using the hourly demand data of Beijing attractions. The model can effectively extract nonlinear features in a demand time series, demonstrating excellent computational ability and good forecasting performance. This model also captures spatial effects from a dynamic perspective. Therefore, it helps toward extracting features more comprehensively and accurately, explaining why it produces better forecasting results.

This study contributes to tourism demand forecasting literature in theory and practice. It theoretically proposes an innovative deep learning model for making tourism demand forecasts. This model exhibits excellent applicability and forecasting performance, contributing to model development in tourism forecasting. This work is among the pioneering attempts to apply the transformer architecture to tourism demand forecasting. Tourism demands may fluctuate periodically in weeks, months and years. Therefore, temporal dependencies should be captured over long time spans to provide more accurate tourism demand forecasts. This issue can be effectively addressed by the proposed model. This study significantly contributes to the exploration of dynamic spatial effect extraction to improve forecasting performance. Influenced by traffic, weather conditions and other factors, spatial relationships among tourism destinations may change over time. Therefore, extracting spatial effects from a dynamic perspective may benefit adequate and exact feature extraction, enabling the proposed model to provide better forecasting results.

On the basis of accurate forecasting results, destination managers can flexibly adjust the allocation of tourism resources in response to demand changes over time and space. Human resources, equipment and services can be deployed to different areas or attractions in a targeted manner to meet demand variability during peak and low seasons. The analysis of dynamic spatial effects can also provide guidance for the emergency management and regulation of tourism destinations. Tourist flows can be controlled during peak seasons or special events in accordance with the spatial distribution and variation of demands. In this manner, the integrity of the tourism environment is maintained and the overall experience of tourists is enhanced.
This study also has limitations. First, the addition of explanatory variables can significantly enhance the predictive power of a model. Tourism demands are affected by various factors, including weather conditions, holidays, economic factors, and pandemics (Song et al., 2019). Therefore, investigating how the proposed model will perform when these factors are incorporated may be a notable future direction. Second, only 30 major tourist attractions in Beijing were selected for the forecasting experiments after considering data availability, attraction popularity and computation burden. The accurate extraction of spatial features may be obtained further if the effects of more attractions are effectively explored and exploited. Third, deep learning models exhibit excellent forecasting performance, but they are criticized for their lack of interpretability. Improving the interpretability of deep learning models without compromising their forecasting accuracy may be a promising research direction.

References


**Corresponding author**

Liyao Huang can be contacted at: huangly@stu.xmu.edu.cn

---

For instructions on how to order reprints of this article, please visit our website: [www.emeraldgrouppublishing.com/licensing/reprints.htm](http://www.emeraldgrouppublishing.com/licensing/reprints.htm)

Or contact us for further details: permissions@emeraldinsight.com