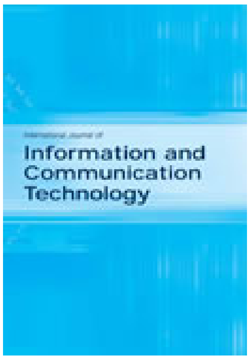


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

A spatio-temporal transformer predictive model for elderly-oriented tourism via attention mechanism

Jiya Sun

DOI: 10.1504/IJICT.2025.10074622

Article History:

Received:	12 September 2025
Last revised:	13 October 2025
Accepted:	15 October 2025
Published online:	01 December 2025

A spatio-temporal transformer predictive model for elderly-oriented tourism via attention mechanism

Jiya Sun

College of Management,
Liaoning University of International Business and Economics,
Dalian 116052, China
Email: jiya_sun@126.com

Abstract: To address the issue that current models for predicting the potential of retirement destinations overlook the spatio-temporal correlations between influencing factors, this paper first selects the influencing factors of retirement destination potential and designs an improved empirical mode decomposition algorithm to decompose these factors, obtaining the individual mode components. Then, the characteristics of each mode component are captured, and the spatio-temporal dependencies are unified through an adaptive embedding mechanism. Subsequently, a temporal self-attention module is designed to capture temporal dependencies, and a spatial self-attention mechanism is implemented to model geographical relationships. Feature fusion is achieved using a multi-head attention mechanism, and the prediction results are output through a feedforward neural network. Experimental outcome indicates that the prediction accuracy of the suggested model improves by 2.7%–11.8% compared to the baseline model, validating the superiority of the suggested model.

Keywords: potential prediction; spatiotemporal transformer; empirical mode decomposition; EMD; attention mechanism.

Reference to this paper should be made as follows: Sun, J. (2025) ‘A spatio-temporal transformer predictive model for elderly-oriented tourism via attention mechanism’, *Int. J. Information and Communication Technology*, Vol. 26, No. 42, pp.82–98.

Biographical notes: Jiya Sun received her Master’s degree from the Beijing Forestry University, China in 2013. She is currently working at the College of Management, Liaoning University of International Business and Economics. Her main research directions are senior residential tourism, innovative tourism forms and empirical mode decomposition.

1 Introduction

Recently, China’s aging population process has accelerated continuously, and the demand for elderly care services has become increasingly diversified. As a new elderly care model that integrates tourism, health, and elderly care, travel-based retirement has gradually become an important way to ease the pressure of traditional elderly care and improve the quality of life for the elderly (Barnes et al., 2016). Assessing and forecasting the potential of travel-based retirement destinations can not only provide references for the elderly to choose suitable retirement locations but also provide scientific basis for local governments in planning elderly care resources and optimising tourism facility layouts, playing a vital role in encouraging sustainable practices in the senior travel and retirement industry (Barbosa et al., 2021). However, the potential of travel-based retirement destinations is comprehensively influenced by multiple factors, which exhibit obvious spatiotemporal dynamic characteristics (Leung et al., 2025). Climate conditions in the natural environment fluctuate with seasons, tourism attractiveness varies with holidays and other time nodes, while convenience of transportation and distribution of medical resources are closely related to spatial positions. Traditional potential prediction models often find it difficult to effectively capture such complex spatiotemporal interaction relationships, leading to certain biases in prediction results (Tang et al., 2024). Enhancing the accuracy of forecasting travel-based retirement destinations by overcoming traditional prediction limitations holds significant practical importance.

The potential prediction of travel-based retirement destinations is influenced by multidimensional factors (Park and Choi, 2021). Traditional research methods predict the likelihood and potential trends of future destinations using historical data. Castán-Lascorz et al. (2022) proposed the autoregressive model, which predicts future trends through a linear combination of historical data, suitable for stationary time series. Tovmasyan (2021) proposed a destination passenger flow forecasting model in light of the autoregressive integrated moving average (ARIMA) model, but it fails to accurately capture the nonlinear relationships in the data, leading to relatively large prediction errors. Liu et al. (2018) analysed multiple factors influencing destination potential and achieved demand prediction for destinations using a spatiotemporal ARIMA model. The travel-based retirement market is in dynamic change. New elderly care policies, infrastructure construction and renovation, among others, will all influence the destinations' potential (Wang et al., 2023). However, the updating of relevant data often lags, leading models to analyse and predict based on outdated data, which cannot accurately reflect current and future reality. Machine learning models significantly improve prediction performance through nonlinear mapping and complex feature extraction, especially excelling in capturing nonlinear relationships and processing large-scale data. Ashton et al. (2019) analysed the influencing factors of travel-based retirement destination potential, removed redundant factors using principal component analysis (PCA) (Gewers et al., 2021), and used decision trees to achieve potential prediction, but the prediction errors were relatively large. Meena et al. (2024) decomposed the influencing variables of destination potential using empirical mode decomposition (EMD) (Wu and Huang, 2009), and obtained prediction results through support vector machines. Bravo et al. (2023) analysed factors affecting the development potential of travel-based retirement destinations from a climatic environment perspective, removed redundant variables using the Pearson coefficient method (Li et al., 2023a), and used the k-nearest neighbours algorithm to predict future values based on the several neighbours closest to the prediction point in historical data.

A major benefit of these machine learning models is their proficiency in managing nonlinear interactions and accommodating sophisticated data formations. However, these studies mostly model and predict the development potential of destinations through manually set rules or handcrafted feature design, resulting in poor forecasting precision. Deep learning models significantly improve prediction performance by extracting the multidimensional spatiotemporal correlations in retirement hotspot factors through artificial neural networks, especially demonstrating outstanding performance in large-scale data and complex pattern modelling. Nanjappa et al. (2024) constructed a destination potential prediction model based on long short-term memory (LSTM), designed grids that could dynamically adapt to the destination, and used a two-layer deep neural network (DNN) structure to capture interactive features among influencing factors. Wen and Chen (2025) proposed a multi-channel convolutional neural network (CNN) to extract features from different influencing variables and obtained the ultimate forecasting outcome through a fully linked level. Jeong and Shin (2020) used diffusion convolution to learn dynamic features of potential influencing variables and enhanced key features through an attention mechanism, thereby improving the prediction effect. Qiu et al. (2023) established a semantic-aware recurrent model, learned embeddings of multiple factors (user, location, time), and captured semantic-aware temporal and spatial transition patterns, achieving a prediction accuracy of 81.9%. Lu and Lin (2023) adopted a multi-module embedding approach to convert sparse characteristics into dense representations, and then used a historical attention scheme to obtain the most relevant historical potential information. However, this method failed to capture dynamic personalised features and hardly considered spatiotemporal dependencies.

The Transformer model, through self-attention mechanisms, can automatically learn the importance relationships between different data features, effectively fusing multi-source heterogeneous data. For example, when analysing the impact of climate data on the attractiveness of retirement tourism, the model can automatically focus on the interactions between temperature and humidity, as well as their associations with the health needs of the elderly population, thereby more accurately evaluating the contribution of climate factors to the destination's potential. Zhu et al. (2023) proposed a retirement destination potential prediction model based on the Transformer model, fully utilising the relationships between dynamic and static factors. However, these studies ignored the spatiotemporal nature of influencing variables. Qin et al. (2024) improved the transformer positional embedding method, allowing the model to perform better when handling long-range spatiotemporal dependencies. Giménez Manuel et al. (2024) suggested a destination potential forecasting model in light of gated recurrent units (GRU) and spatiotemporal transformers, extracting spatiotemporal interaction features of variables through a dual-layer multi-head attention (MHA) mechanism to generate prediction results.

Through an analysis of the current status of research on retirement tourism destination potential prediction, it can be seen that current studies ignore the spatiotemporal correlations among influencing factors, leading to low forecasting accuracy. To address this, this article suggests a spatiotemporal Transformer-based retirement tourism destination potential prediction model. The principal contributions of this research are outlined in the following four dimensions.

1 The singular value decomposition (SVD) method is introduced to optimise the EMD algorithm to address the mode mixing problem in EMD. EMD decomposes the influencing factors layer by layer according to the instantaneous frequency and calculates the permutation entropy of each intrinsic mode function (IMF) component. The effective number of singular values in SVD is determined through the singular value difference spectrum and the one-sided maximum principle. Then, SVD denoising is applied to components with permutation entropy greater than the threshold, and the denoised components are recomposed using the remaining IMFs

to reconstruct filtered signals, thus maximally retaining the informative content of the original driving factors.

2 To enhance the spatiotemporal feature representation capabilities, this paper proposes a novel spatiotemporal adaptive embedding module. This module integrates multiple spatiotemporal embedding features into the model, particularly through an adaptive embedding mechanism. The features of each IMF of the influencing factors are extracted, combined with daily and weekly periodic information to capture periodic changes, and the adaptive embedding mechanism unifies spatiotemporal dependencies.

3 A prediction model based on spatiotemporal transformer is proposed. This model designs a novel spatiotemporal encoder layer, optimises the spatial self-attention mechanism, and introduces a spatial attention mechanism based on key node perception to conduct spatiotemporal modelling of the potential key nodes in the IMF components of influencing factors. By mining the inherent attributes of key nodes and integrating temporal, semantic, and geographic correlations, the model significantly improves the accuracy of forecasting the potential of retirement tourism destinations.

4 Extensive simulation experiments were conducted on real datasets. The outcome showed that the prediction accuracy and F1 of the proposed model were 97.9% and 95.3%, respectively, which are superior to the baseline models, thus verifying the effectiveness and superiority in handling complex spatiotemporal data and accurately predicting the potential of retirement tourism destinations. This study provides scientific decision-making support for the selection and planning of retirement tourism destinations.

2 Relevant technologies

2.1 Attention mechanism

The attention mechanism has achieved a paradigm shift in the understanding of long texts by breaking through the limitations of traditional models. Taking the development of the sequence-to-sequence (Seq2Seq) model (Peng et al., 2022) as an example, early models compressed the entire input sequence into an individual fixed-dimension semantic vector via the encoder. This 'information bottleneck' often caused the loss of key details when translating long articles. The innovation of the attention mechanism lies in restructuring the information interaction in the encoder-decoder architecture (Li et al., 2023b). The decoder can dynamically access all hidden states of the encoder when generating each target word, and autonomously filters relevant contexts through learnable weight systems.

The attention mechanism calculation usually involves three vectors: query, key, and value. A query is the feature vector representing subjective awareness. A key is the feature vector representing prominent information. A value is the feature vector representing the object itself, and it often appears in pairs with the key. The attention mechanism achieves the attention weight distribution over the values through the attention aggregation between the query and key, generating the final output.

2.2 Transformer model

The prediction model based on transformer is an important innovation in the domain of deep learning recently. The chief feature of transformer is the use of self-attention mechanism to process sequence data. Unlike traditional sequence processing models, such as RNN or LSTM, which rely on sequential computation, transformer can process the entire sequence in parallel, enabling the model to consider information from all positions in the sequence simultaneously. In the self-attention calculation, the model computes the attention scores of each element in the sequence relative to all other elements, which helps the model, capture global context information.

A key advantage of the Transformer model in processing sequence data is its ability to extract long-range dependencies, which is crucial for many complex prediction tasks (Nassiri and Akhloufi, 2023). The transformer architecture comprises two fundamental modules: the encoder and the decoder. In the encoder, the MHA mechanism complements the shortcomings of the self-attention scheme, i.e., excessive focus on the current position of information (Jiang et al., 2024). Through this mechanism, the model is able to simultaneously pay attention to information from multiple positions, introducing multiple sub-representation spaces to the attention layer. This mechanism calculates attention scores using the key, query, and value vectors, as defined in equation (1), where the calculation of Q, K, and V is implied in equation (2).

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

$$Q = W_q X, K = W_k X, V = W_v X \tag{2}$$

2.3 EMD algorithm

EMD is an effective method for decomposing nonlinear variables, often used for analysing non-stationary variables. During the decomposition process, EMD can automatically determine the amount of decomposition levels, i.e., the number of IMFs (Yao et al., 2023). It gradually extracts IMFs with different frequency components according to the complexity and intrinsic characteristics of the signal until the remaining residue simplifies to a steady-state or unidirectionally trending output. The original variable sequence can be transformed into a series of IMFs through decomposition. In the study of the potential prediction of vacation retirement destinations, high-frequency IMFs may reflect short-term market fluctuations, such as weekend and holiday tourism peaks. Low-frequency IMFs may represent long-term development trends, such as the overall growth trend of the vacation retirement market as aging intensifies. By analysing IMFs at different scales, the change patterns can be fully understood, providing a more accurate basis for the potential prediction of destinations. The decomposition process of EMD is as follows.

Step 1 Identify all the maximum and minimum value points in the variable $x(t)$, gain the extremum envelope of $x(t)$ using the moving average method, and calculate the difference $m_1(t)$ between $h_1(t)$ and the average value $x(t)$ of the envelope line, i.e., $h_1(t) = x(t) - m_1(t)$.

Step 2 Determine whether the difference $h_1(t)$ is an IMF. If not, set $h_1(t)$ as the initial sequence in the new cycle and repeat step (1) until $h_1(t)$ becomes an IMF; if yes, set $h_1(t)$ as the first IMF component. Then $c_1(t) = h_1(t)$, record the difference between $x(t)$ and the IMF as the residual component $r_1(t)$, i.e., $r_1(t) = x(t) - c_1(t)$.

Step 3 Take $r_1(t)$ as the novel initial sequence, repeat step 1 and step 2, and continue screening to extract the remaining IMF components until the residual component $r_n(t)$ is very small or becomes a monotonic function. $x(t)$ can be expressed as below.

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{3}$$

The EMD algorithm significantly enhances the decomposition performance of nonlinear, non-stationary signals through an adaptive decomposition mechanism based on local signal features. Fourier transforms rely on predefined triangular basis functions, while wavelet transforms require prior selection of wavelet bases. When processing nonlinear, non-stationary signals, these methods may yield inaccurate decompositions due to mismatches between the basis functions and signal characteristics. EMD constructs upper and lower envelopes using the signal's own extrema, calculates mean curves, and iteratively filters them to directly generate IMFs. This decomposition method is entirely data-driven, requiring no prior assumptions, and can adaptively capture the signal's local time-varying characteristics.

3 Selection and decomposition of factors influencing the potential of travel and retirement destinations

3.1 Selection of influencing factors for the potential of vacation retirement destinations

The influencing factors for the potential of vacation retirement destinations form a multi-dimensional and comprehensive system. Based on existing research, this paper selects the current influencing factors of the potential of vacation retirement destinations and summarises them into the following aspects.

1 Natural environmental factors. These factors include climate conditions and ecological environment. A warm and humid climate with distinct seasons is generally more popular with the elderly. Moreover, clean air is critically important for the respiratory health of the elderly, so regions with good air quality are more advantageous.

2 Social service factors. These factors include medical security, elderly care services, and transportation. Complete medical facilities, including hospitals, clinics, pharmacies, etc. are important considerations for the elderly in selecting vacation retirement destinations. Diversified elderly care institutions, for example nursing homes, senior apartments, and day care centres, can meet the varied needs of different elderly individuals. A convenient transportation network, including highways, railways, and aviation, can facilitate the elderly in travelling and visiting family and friends.

3 Economic cost factors. These factors include the cost of living and the cost of elderly care. A reasonable price level, including daily expenses such as food, housing, and transportation, can alleviate the economic burden on the elderly. Reasonable elderly care expenses, including costs for care facilities and medical services, are important considerations for the elderly when selecting vacation retirement destinations.

4 Policy support factors. Tourism policies implemented by the government, such as tourism discounts and tourism safety guarantees, can enhance the attractiveness of vacation retirement destinations. Comprehensive regulations for protecting the rights and interests of the elderly can safeguard the legal rights of the elderly in their vacation retirement locations.

5 Cultural atmosphere factors. A rich historical and cultural heritage, such as ancient architecture and historical sites, can satisfy the elderly's pursuit and exploration of culture. Diverse social activities, such as senior clubs and interest groups, can meet older adults' interpersonal needs and support their psychosocial wellness.

6 Other factors include safety factors, etc. Areas with a good social security environment and low natural disaster risks are more attractive, and the elderly tend to choose safe and stable living environments.

From the above analysis, it can be concluded that the influencing factors of the potential of vacation retirement destinations include natural environment, social services, economic costs, policy support, cultural atmosphere, and other factors. They are denoted as x_1, x_2, …, x_n, and through normalisation, these influencing factors are linearly mapped to the [0, 1] interval, which facilitates the identification by subsequent models.

3.2 Influence factor decomposition based on improved EMD

After pre-processing the influencing factors of vacation retirement destination potentials, this paper designs an improved EMD algorithm to decompose these influencing factors, thus reducing the complexity of the input variables of the prediction model. The traditional EMD method suffers from the problem of mode mixing, where signal components at various time scales are incorrectly decomposed into the same IMF, or co-temporal signal characteristics appear fragmented across the mode decomposition. To address the above issues, researchers add white noise in practical applications to eliminate the mixing, yet the number of times white noise can be added is limited, leading to the noise not being completely eliminated after ensemble averaging. To this end, this paper decomposes the influencing factors layer by layer according to the EMD based on instantaneous frequency and calculates the permutation entropy of each IMF component. Then, the effective number of singular values of the SVD (Begum et al., 2022) is determined by the singular value difference spectrum and the one-sided maximum principle. The SVD denoising is applied to components with permutation entropy greater than the threshold, and the denoised signal is obtained by reconstructing with the remaining IMF components to maximally preserve the informative content of original drivers while eliminating extraneous signals.

SVD is a nonlinear filtering method that extracts information through matrix decomposition and computation. It can effectively eliminate noise interference and redundant data. By decomposing the reconstructed Hankel matrix with SVD, the components with significant singular values are preserved based on the separability of noise and signal energy to separate the noise from useful information, thereby achieving the purpose of denoising. The specific steps of SVD optimisation of EMD are as follows.

Step 1 Add two white noises $o_1^+(t)$ and $o_1^-(t)$ with opposite signs to the original influencing factors $x(t)$, yielding the following formula. Here, $\chi_1^+(t)$ and $\chi_1^-(t)$ are results after adding $o_1^+(t)$ and $o_1^-(t)$, respectively.

$$\chi_1^+(t) = x(t) + o_1^+(t) \tag{4}$$

$$\chi_1^-(t) = x(t) + o_1^-(t) \tag{5}$$

Step 2 Decompose $\chi_1^+(t)$ and $\chi_1^-(t)$ to obtain the time series of the first IMF components $\{I_{i,1}^+(t)\}$ and $\{I_{i,1}^-(t)\}$. After integration, the following formula can be obtained, where N is the length of $x(t)$ and Ne is the logarithm of the number of white noise pairs added to the original signal, $i = 1, 2, \ldots, Ne$.

$$I_1(t) = \frac{1}{2N}\sum_{i=1}^{Ne}\left[I_{i,1}^+(t) + I_{i,1}^-(t)\right] \tag{6}$$

Step 3 Calculate the permutation entropy value of the aforementioned $I_1(t)$. Determine whether it is an abnormal component based on the set threshold.

Step 4 If it is an abnormal component, re-execute step (1) until $I_q(t)$ is a normal variable and white noise is no longer added, where $1 \le q \le N$.

Step 5 Perform SVD denoising on the first q-1 abnormal IMF components separated from the original variable. The remaining signal $r(t)$ is as follows.

$$r(t) = x(t) - \sum_{j=1}^{q-1} I_j(t) \tag{7}$$

Step 6 Reconstruct the final denoised signal by integrating processed IMFs with residual components, and then perform EMD decomposition on this variable.

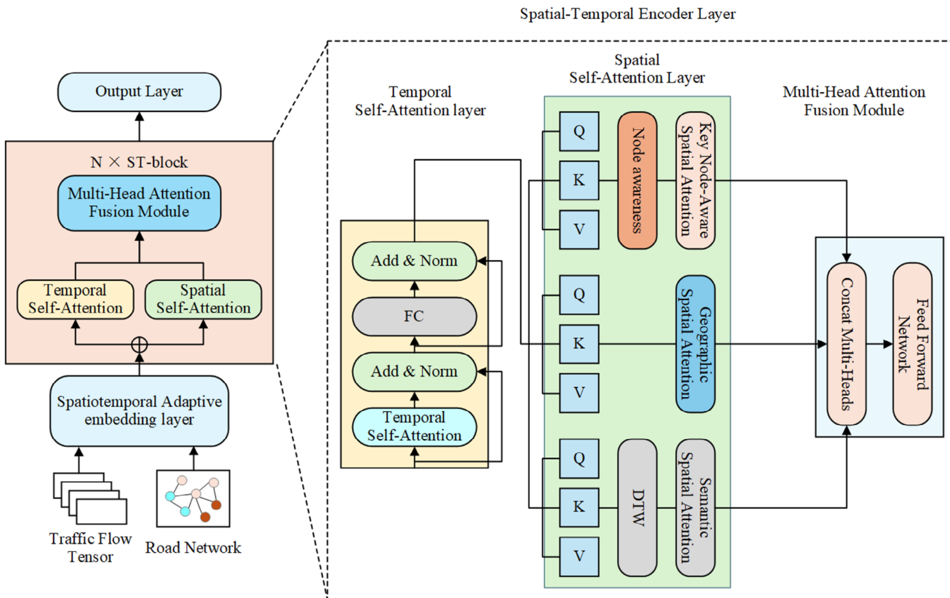
Apply SVD denoising to the abnormal IMF components screened in step (3), and recombine the processed IMFs with residual modes to reconstruct purified variables.

4 Digital media operation effect prediction based on user multimodal sentiment analysis

4.1 Prediction model structure

The influencing factors of the potential of retirement travel destinations have obvious spatiotemporal characteristics. They are not only closely related to time and location but also to spatial structure information. Current research methods usually rely on complex neural networks to capture spatiotemporal correlations among influencing factors, but they often fail to deeply mine the intrinsic properties of the data. To address the above issues, this paper suggests a retirement destination potential forecasting model in light of a spatiotemporal transformer (RDSTFormer). The model structure is implied in Figure 1.

Figure 1 RDSTFormer model structure (see online version for colours)



The RDSTFormer model mainly consists of a spatiotemporal adaptive embedding layer and multiple parallel spatiotemporal encoders. First, the spatiotemporal adaptive embedding layer extracts the features of each IMF of the influencing factors, combines daily and weekly periodic information to capture periodic changes, and unifies spatiotemporal dependencies through an adaptive embedding mechanism. Subsequently, the data is input into the spatiotemporal encoder module; the temporal self-attention module captures temporal dependencies, while the spatial self-attention module utilises geographic and semantic mask matrices to catch spatial dependencies, and the key node perception module models the spatial characteristics of the nodes. The multi-head architecture concurrently processes attention through several independent heads, each extracting distinct spatiotemporal relationships, subsequently integrated into a unified representation. Finally, each position's representation is independently transformed by a two-layer MLP, with the output stabilised through layer normalisation and combined with the original input via residual linkage, and the output layer generates the predicted results for the potential of retirement travel destinations.

4.2 Spatiotemporal adaptive embedding layer

The spatiotemporal adaptive embedding layer is used to convert the decomposed influencing factor components into high-dimensional representations while preserving the feature information of the original data. Specifically, first, a fully linked level is used to perform feature embedding on the original input data to obtain the feature embedding E_f, as shown in equation (8).

$$E_f = FC\left(concat\left(X_{[t-T+1:t]}\right)\right) \tag{8}$$

where $FC(\cdot)$ is the fully connected layer, $X_{[t-T+1:t]}$ is the data from time step $t - T + 1$ to time step t, expressed as a matrix. Building upon this, a spatiotemporal data embedding mechanism is further designed to introduce data into the RDSTFormer model, mainly including temporal cycle embedding, temporal position embedding, and adaptive embedding. Among these, temporal cycle embedding is used to model the periodic variation characteristics of the data, temporal position embedding is used to capture the relationships between different time steps in a time series, and adaptive embedding dynamically and uniformly captures complex spatiotemporal correlations. By fusing these different types of embeddings, the model can more comprehensively represent the spatiotemporal characteristics of the influencing factors of retirement travel destination potential, thereby improving the accuracy of the predictions.

To address the issue of the current model's inability to capture the sequential relationships between time steps in time series, this model employs a positional embedding method. This method generates positional embeddings for each time step using sine and cosine functions and combines these embeddings with the input data, thereby helping the model understand the relative positions of the time steps and temporal dependencies, as shown in equation (9) and equation (10), where p is the position index, j is the dimension index, and d is the embedding dimension. The final temporal position embedding can be expressed as E_t.

$$TPE(p, 2j) = \sin\left(\frac{p}{10{,}000^{2j/d}}\right) \tag{9}$$

$$TPE(p, 2j+1) = \cos\left(\frac{p}{10{,}000^{2j/d}}\right) \tag{10}$$

The spatiotemporal relationships in the prediction of retirement travel destination potential are not only influenced by periodic patterns but are also closely related to the temporal sequence. Based on this, this model designs a spatiotemporal adaptive embedding method denoted as E_a to capture complex spatiotemporal relationships in a unified manner, integrating temporal order information and spatial dependencies into a single framework, thus avoiding the complexity of separately modelling time and space in traditional methods.

Unlike the traditional static adjacency matrix method, this approach can effectively capture the temporal order information and spatial correlations in the influencing factors. Finally, by concatenating the aforementioned adaptive embeddings, the hidden spatiotemporal representation Z is obtained, as shown in equation (11), where $\oplus$ represents the concatenated embedding.

$$Z = \left[E_f \oplus E_t \oplus E_a\right] \tag{11}$$

4.3 Spatiotemporal encoder layer

This framework implements a spatiotemporal encoder leveraging self-attention to capture intricate dynamic relationships among retirement destination determinants. The encoder consists of three main modules: the temporal self-attention module captures dynamic temporal dependencies. The spatial self-attention module, composed of semantic spatial self-attention and geographic self-attention, captures dynamic spatial dependencies both at long and short distances. The key node-aware spatial self-attention module can deeply mine the edge weights and node degrees in the data. Finally, the MHA fusion module effectively fuses the outputs of the four self-attention modules. Through the collaborative effects of these modules, the encoder can more accurately characterise the dynamic features of the influencing factors, thereby improving the prediction performance.

1 Temporal self-attention layer. This model introduces the temporal self-attention mechanism to catch temporal dependencies and employs a weighted click attention mechanism within the temporal attention module. First, the query vector Q, key vector K, and value vector V for node N are achieved, as shown in equation (12), where W_Q^T, W_K^T, W_V^T are learnable parameters. Subsequently, the temporal dependencies between nodes are calculated through the self-attention mechanism on the temporal dimension, as shown in equation (13).

$$Q_n^{(T)} = X_{:n:}W_Q^T,\ K_n^{(T)} = X_{:n:}W_K^T,\ V_n^{(T)} = X_{:n:}W_V^T \tag{12}$$

$$TA\left(Q_n^{(T)}, K_n^{(T)}, V_n^{(T)}\right) = Softmax\left(\frac{\left(Q_n^{(T)}\right)\left(K_n^{(T)}\right)}{\sqrt{d_K^T}}\right)V_n^{(T)} \tag{13}$$

2 Spatial self-attention layer. This model designs a spatial self-attention ingredient that dynamically extracts the spatial characteristics of influencing factors by combining key-node-aware attention, semantic, and geographic information. First, at time t, the query matrix, key matrix, and value matrix are obtained, as shown in equation (14), where $W_Q^s, W_t^{(s)}, W_V^s$ are learnable parameters. In the spatial self-attention module, a key-node-aware mask matrix M_{key} is used to capture key nodes in the prediction. The key-node-aware spatial self-attention among all nodes at time t is calculated, as shown in equation (15), where $A_t^{(S)} = (Q_t^{(S)})(K_t^{(S)})^T \big/ \sqrt{d'}$.

$$Q_t^{(s)} = X_{:t:}W_Q^s,\ K_t^{(s)} = X_{:t:}W_t^{(s)} = X_{:t:}W_V^s \tag{14}$$

$$KeySSA\left(Q_t^{(S)}, K_t^{(S)}, V_t^{(S)}\right) = Softmax\left(A_t^{(S)} \odot M_{key}\right)V_n^{(T)} \tag{15}$$

Furthermore, since the spatial self-attention mechanism considers the entire spatial graph as a fully linked graph, but in reality only a few nodes need to interact (Salvador and Chan, 2007), this model introduces a geographic mask matrix M_{geo} and a semantic mask matrix M_{sem} to more effectively extract micro-scale and macro-scale geographical dependencies. M_{sem} masks the attention of distant nodes by limiting the distance between nodes to be less than threshold λ, thus emphasising interactions between nearby nodes. M_{sem} masks nearby nodes by adopting the dynamic time warping algorithm (DTW) to calculate the similarity of historical data, selecting the k most similar nodes for each node as semantic neighbours.

Based on these two semantic mask matrices, RDSTFormer further designs two spatial self-attention modules: geographic spatial self-attention and semantic spatial self-attention, as shown in equation (16) and equation (17), where $\odot$ is the Hadamard product. Through this operation, the spatial self-attention module effectively fuses short-range geographic neighbourhood information, long-range semantic neighbourhood information, and key-node-aware information.

$$GeoSSA\left(Q_t^{(S)}, K_t^{(S)}, V_t^{(S)}\right) = Softmax\left(A_t^{(S)} \odot M_{geo}\right)V_n^{(T)} \tag{16}$$

$$SemSSA\left(Q_t^{(S)}, K_t^{(S)}, V_t^{(S)}\right) = Softmax\left(A_t^{(S)} \odot M_{sem}\right)V_n^{(T)} \tag{17}$$

4.4 MHA fusion and output layer

The MHA mechanism integrates the four types of attention mechanisms mentioned above into a multi-head self-attention module. Specifically, each attention mechanism corresponds to an independent attention head, which individually captures different spatiotemporal characteristics in the influencing factor sequence. The MHA architecture employs parallelised attention heads that operate simultaneously, enabling the model to learn more spatiotemporal information from different perspectives. The vectors from parallel attention heads are concatenated along the feature dimension and projected to the target dimension via learned weights; the model effectively fuses spatial and temporal information. Additionally, the definition of MHA is shown in equation (18).

$$Multihead(Q, K, V) = Concat\left(Z_{[1,h_{geo}]}^{geo}, Z_{[1,h_{sem}]}^{sem}, Z_{[1,h_{bey}]}^{key}, Z_{[1,h_t]}^{t}\right)W^O \tag{18}$$

where Z^{geo}, Z^{sem}, Z^{key}, Z^t are the output concatenations, h_{geo}, h_{sem}, h_{key}, h_t are the number of attention heads for the multi-head heterogeneous attention mechanism, and W^O is the projection matrix of learnable parameters. In addition, the dimension parameters of $d' = d/h$ and $h = h_{geo} + h_{sem} + h_{key} + h_t$ are set. Finally, the MHA fusion module uses a position-wise fully linked feed-forward network on the output of the MHA mechanism to obtain output X_o, followed by layer normalisation and residual connections used in the original transformer.

The output layer of the RDSTFormer model employs a skip connection composed of 1×1 convolution. After each spatiotemporal encoder, the output X_o is transformed into a skip connection X_{SC}. Then, the outputs of all skip connection layers are summed to generate the hidden state X_h. Finally, the output layer directly converts the hidden state X_h into the target output, as shown in equation (19), where $Conv_1$ and $Conv_2$ are two 1×1 convolutional layers.

$$\hat{X} = Conv_2\left(Conv_1\left(X_h\right)\right) \tag{19}$$

5 Experimental results and analyses

5.1 Visualisation of potential prediction errors for travel and retirement destinations

This paper uses the China-aging-destination (CAD) dataset collected in literature (Liang et al., 2023) as the experimental dataset. The CAD dataset includes geographical features, climate features, and other data for 300 districts and counties across 34 provincial administrative regions in China from 2018 to 2023, as well as potential score data from experts for the regions, with a scoring scale of ten points. The experimental data is classified into training, validation, and test sets in a 6:2:2 ratio, respectively used for model training, performance validation, and final testing. All experiments in this study are run on an NVIDIA Geforce 4090 GPU with 24GB of GPU memory, and the Python version used in the virtual environment is 3.9.7. To train the RDSTFormer model, Adam adapter optimisation is used, with an initial learning rate of 0.001, a batch size of 16, a total of 200 epochs of training, and eight attention heads.

Figure 2 The predicted heat map of the RDSTFormer model (see online version for colours)

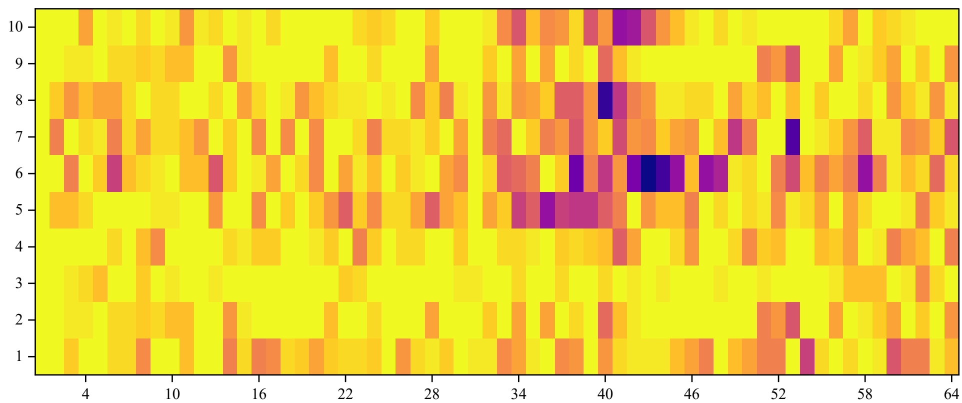
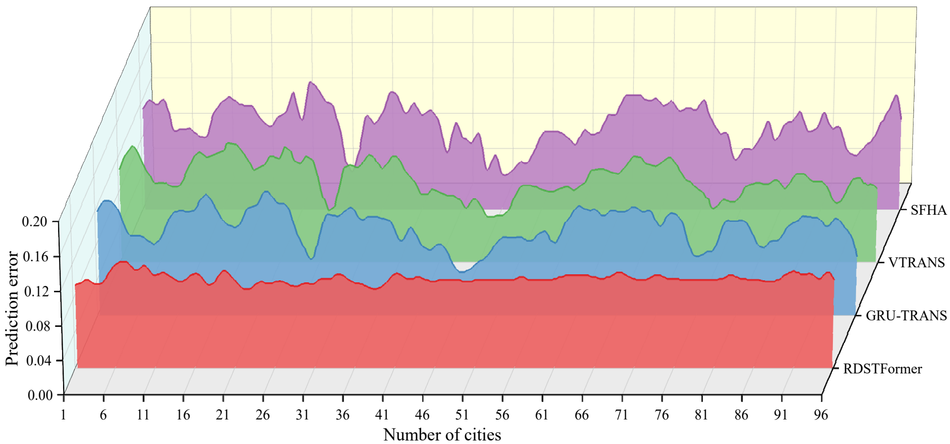


Figure 2 shows the prediction heatmap of the RDSTFormer model on the CAD dataset. The results imply that the model demonstrates good predictive performance in both short-term and long-term forecasting, effectively capturing the temporal trends of retirement destination potential. In addition, the adaptive embedding designed in RDSTFormer can adjust feature representations at different temporal scales, especially more accurately modelling temporal patterns over long time ranges, thus unifying spatiotemporal relationships and improving forecasting accuracy and robustness, and performing better in medium-and long-term forecasting tasks. Compared to traditional fixed embedding methods, adaptive embedding introduces spatiotemporal representations, enabling the model to autonomously adjust features based on time steps and spatial locations, more effectively capturing the dynamic changes of influencing factors for retirement destination potential.

5.2 Comparative experiment

To assess the prediction performance of the RDSTFormer model, this article selects SFHA (Lu and Lin, 2023), VTRANS (Zhu et al., 2023), and GRU-TRANS (Giménez Manuel et al., 2024) as baseline models. Prediction performance indicators include accuracy, F1, MAE, and RMSE. The variation of prediction accuracy across different models is shown in Figure 3. RDSTFormer rapidly converges in the early stages of training and stabilises after approximately 75 epochs, finally achieving optimal performance with a prediction accuracy of 97.9%. GRU-TRANS stabilises after approximately 80 epochs, reaching a prediction accuracy of 95.2%. VTRANS stabilises after approximately 50 epochs, achieving a prediction accuracy of 93.6%. SFHA exhibits slower accuracy growth throughout the entire training process and remains below 90% after convergence. This model fails to capture dynamic personalised preferences and barely takes into account spatiotemporal dependencies. In summary, RDSTFormer demonstrates better stability and validates the superiority of this model.

Figure 3 The changes in the prediction accuracy of different models (see online version for colours)



As shown in Figure 4, the error forecasting results of different models are presented. The SFHA model has a high deviation with an average error of 0.17. The VTRANS model has an average error as high as 0.19. The GRU-TRANS model's prediction results are the most unstable, showing a relatively significant slope decline in error. The potential curve predicted by RDSTFormer is more gradual and similar to the actual potential, proving that the prediction value of RDSTFormer has better fitting and adaptability compared to the actual value.

Figure 4 Error prediction result (see online version for colours)

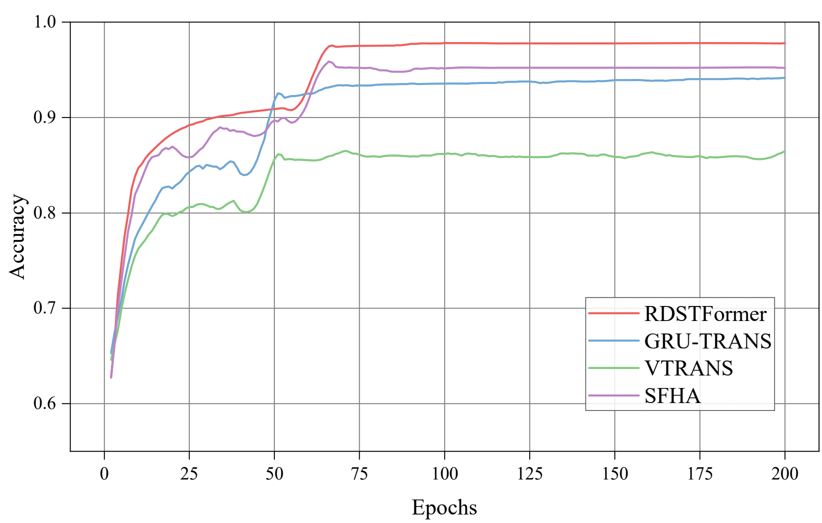


Table 1 shows the comparison of prediction performance indicators for the four models. The prediction accuracy of RDSTFormer is improved by 11.8%, 4.3%, and 2.7% compared to SFHA, VTRANS, and GRU-TRAN, respectively. The F1 values of SFHA, VTRANS, GRU-TRAN, and RDSTFormer are 83.3%, 89.8%, 92.7%, and 95.3%, respectively, indicating an improvement of 2.6%–12% for RDSTFormer compared to the baseline models. In comparison with MAE and RMSE, the MAE and RMSE of RDSTFormer are 0.102 and 0.158, respectively, which is reduced by 23.3%–48.5% compared to the other three models. RDSTFormer fully considers the spatiotemporal characteristics of factors affecting the potential of retirement travel destinations and captures complex dependencies between different time steps and spatial nodes through adaptive embedding, showing a significant advantage in the task of predicting the potential of retirement travel destinations.

Table 1 Comparison of predictive performance metrics for different models

Model	*SFHA*	*VTRANS*	*GRU-TRAN*	*RDSTFormer*
Accuracy	0.861	0.936	0.952	0.979
F1	0.833	0.898	0.927	0.953
MAE	0.198	0.176	0.147	0.102
RMSE	0.284	0.227	0.206	0.158

6 Conclusions

With the acceleration of China's aging process, travel-based retirement as a new model integrating elderly care and tourism, its destination potential prediction becomes key to meeting the diverse needs of the elderly population. Existing prediction models find it difficult to effectively capture the spatiotemporal dynamic correlations and multifactor coupling effects in the travel-based retirement scenario, leading to limited prediction accuracy and generalisation ability. In order to address the current research that ignores the spatiotemporal correlations among influencing factors, resulting in low prediction accuracy, this paper proposes a travel-based retirement destination potential prediction model based on spatiotemporal transformer. Firstly, select influencing factors of the potential for travel-based retirement destinations, and design an improved EMD algorithm to decompose the influencing factors, obtaining various modal components IMF. Then extract the features of each IMF of the influencing factors through an adaptive embedding mechanism to unify the spatiotemporal dependency relationships. Subsequently, input the data into the spatiotemporal encoder module. The temporal self-attention module captures temporal dependencies, while the spatial self-attention module uses geographical and semantic mask matrices to capture spatial dependencies. The key node perception module models the spatial characteristics of nodes. Through a MHA mechanism, calculate multiple attention heads in parallel, capturing spatiotemporal dependencies from different perspectives, and integrating the outputs of multiple attention heads. Finally, the features are processed through a position-wise feed-forward neural network with layer normalisation and residual connections, and the output layer generates the travel-based retirement destination potential prediction results. Experimental results show that the prediction accuracy of the proposed model reaches 97.9%, which is at least 2.7% higher than that of baseline models, achieving relatively accurate travel-based retirement destination potential prediction.

Acknowledgements

This work is supported by the 2024 annual school-level research project of Liaoning University of International Business and Economics titled ‘Research on the Development Prospects and Practical Paths of senior residential Tourism in Dalian under the Context of the Silver Economy’ (No. 2024XJLXYB006).

Declarations

All authors declare that they have no conflicts of interest.

References

Ashton, A.S., Scott, N. and Choibamroong, T. (2019) ‘Place development for international retirement migrants: a decision-making process model’, *Journal of Place Management and Development*, Vol. 12, No. 3, pp.449–468.

Barbosa, B., Santos, C.A. and Santos, M. (2021) ‘Tourists with migrants’ eyes: the mediating role of tourism in international retirement migration’, *Journal of Tourism and Cultural Change*, Vol. 19, No. 4, pp.530–544.

Barnes, R., Winters, M., Ste-Marie, N., McKay, H. and Ashe, M.C. (2016) ‘Age and retirement status differences in associations between the built environment and active travel behaviour’, *Journal of Transport & Health*, Vol. 3, No. 4, pp.513–522.

Begum, M., Ferdush, J. and Uddin, M.S. (2022) ‘A Hybrid robust watermarking system based on discrete cosine transform, discrete wavelet transform, and singular value decomposition’, *Journal of King Saud University-Computer and Information Sciences*, Vol. 34, No. 8, pp.5856–5867.

Bravo, J., Alarcón, R., Valdivia, C. and Serquén, O. (2023) ‘Application of machine learning techniques to predict visitors to the tourist attractions of the Moche route in Peru’, *Sustainability*, Vol. 15, No. 11, pp.89–97.

Castán-Lascorz, M., Jiménez-Herrera, P., Troncoso, A. and Asencio-Cortés, G. (2022) ‘A new hybrid method for predicting univariate and multivariate time series based on pattern forecasting’, *Information Sciences*, Vol. 586, pp.611–627.

Gewers, F.L., Ferreira, G.R., Arruda, H.F.D., Silva, F.N., Comin, C.H., Amancio, D.R. and Costa, L.D.F. (2021) ‘Principal component analysis: a natural approach to data exploration’, *ACM Computing Surveys (CSUR)*, Vol. 54, No. 4, pp.1–34.

Giménez Manuel, J.G., Giner Pérez de Lucia, J., Celdrán Bernabeu, M.A., Mazón López, J.N., Cano Escribá, J.C. and Cecilia Canales, J.M. (2024) ‘Advancing smart tourism destinations: a case study using bidirectional encoder representations from transformers-based occupancy predictions in Torrevieja (Spain)’, *IET Smart Cities*, Vol. 6, No. 4, pp.422–440.

Jeong, M. and Shin, H.H. (2020) ‘Tourists’ experiences with smart tourism technology at smart destinations and their behavior intentions’, *Journal of Travel Research*, Vol. 59, No. 8, pp.1464–1477.

Jiang, P., Obi, T. and Nakajima, Y. (2024) ‘Integrating prior knowledge to build transformer models’, *International Journal of Information Technology*, Vol. 16, No. 3, pp.1279–1292.

Leung, X.Y., Zhong, Y.S. and Sun, J. (2025) ‘The impact of social media influencer’s age cue on older adults’ travel intention: the moderating roles of travel cues and travel constraints’, *Tourism Management*, Vol. 106, pp.49–56.

Li, X., Li, M., Yan, P., Li, G., Jiang, Y., Luo, H. and Yin, S. (2023a) ‘Deep learning attention mechanism in medical image analysis: basics and beyonds’, *International Journal of Network Dynamics and Intelligence*, Vol. 42, pp.93–116.

Li, Z., Yang, Y., Li, L. and Wang, D. (2023b) 'A weighted Pearson correlation coefficient based multi-fault comprehensive diagnosis for battery circuits', *Journal of Energy Storage*, Vol. 60, pp.10–17.

Liang, Z., Luo, H. and Hui, T. (2023) 'Moving for a good life: tourism mobility and subjective well-being of Chinese retirement migrants', *Tourism Geographies*, Vol. 25, Nos. 2–3, pp.778–796.

Liu, Y-Y., Tseng, F-M. and Tseng, Y-H. (2018) 'Big data analytics for forecasting tourism destination arrivals with the applied vector autoregression model', *Technological Forecasting and Social Change*, Vol. 130, pp.123–134.

Lu, E.H-C. and Lin, Y-R. (2023) 'A self-attention model for next location prediction based on semantic mining', *ISPRS International Journal of Geo-Information*, Vol. 12, No. 10, pp.42–50.

Meena, K.K., Bairwa, D. and Agarwal, A. (2024) 'A machine learning approach for unraveling the influence of air quality awareness on travel behavior', *Decision Analytics Journal*, Vol. 11, pp.59–63.

Nanjappa, Y., Nassa, V.K., Varshney, G., Lal, B., Pandey, S. and Turukmane, A.V. (2024) 'Improving migration forecasting for transitory foreign tourists using an ensemble DNN-LSTM model', *Entertainment Computing*, Vol. 50, pp.65–72.

Nassiri, K. and Akhloufi, M. (2023) 'Transformer models used for text-based question answering systems', *Applied Intelligence*, Vol. 53, No. 9, pp.10602–10635.

Park, J-A. and Choi, B. (2021) 'Factors affecting the intention of multi-family house residents to age in place in a potential naturally occurring retirement community of Seoul in South Korea', *Sustainability*, Vol. 13, No. 16, pp.89–102.

Peng, X., Shu, W., Pan, C., Ke, Z., Zhu, H., Zhou, X. and Song, W.W. (2022) 'DSCSSA: a classification framework for spatiotemporal features extraction of arrhythmia based on the Seq2Seq model with attention mechanism', *IEEE Transactions on Instrumentation and Measurement*, Vol. 71, pp.1–12.

Qin, Z., Zhang, P. and Ma, Z. (2024) 'DeepAGS: deep learning with activity, geography and sequential information in predicting an individual's next trip destination', *IET Intelligent Transport Systems*, Vol. 18, No. 10, pp.1895–1909.

Qiu, G., Tang, G., Li, C., Luo, L., Guo, D. and Shen, Y. (2023) 'A complete and comprehensive semantic perception of mobile traveling for mobile communication services', *IEEE Internet of Things Journal*, Vol. 11, No. 3, pp.5467–5490.

Salvador, S. and Chan, P. (2007) 'Toward accurate dynamic time warping in linear time and space', *Intelligent Data Analysis*, Vol. 11, No. 5, pp.561–580.

Tang, L., Tang, C., Fu, Q. and Ma, C. (2024) 'Predicting travel mode choice with a robust neural network and Shapley additive explanations analysis', *IET Intelligent Transport Systems*, Vol. 18, No. 7, pp.1339–1354.

Tovmasyan, G. (2021) 'Forecasting the number of incoming tourists using ARIMA model: case study from Armenia', *Marketing i menedžment innovacij*, No. 3, pp.139–148.

Wang, K., Kou, L. and Xu, H. (2023) 'Enabling home and destinations for the well-being of seasonal retired tourists', *Tourism Geographies*, Vol. 25, No. 6, pp.1555–1579.

Wen, X. and Chen, X. (2025) 'A new breakthrough in travel behavior modeling using deep learning: a high-accuracy prediction method based on a CNN', *Sustainability*, Vol. 17, No. 2, pp.73–88.

Wu, Z. and Huang, N.E. (2009) 'Ensemble empirical mode decomposition: a noise-assisted data analysis method', *Advances in Adaptive Data Analysis*, Vol. 1, No. 1, pp.1–41.

Yao, Y., Zhang, Z-Y. and Zhao, Y. (2023) 'Stock index forecasting based on multivariate empirical mode decomposition and temporal convolutional networks', *Applied Soft Computing*, Vol. 142, pp.56–69.

Zhu, Y., Zhu, Q., Ma, Y., Chen, S., Meng, H. and Zubair, M. (2023) 'Identify travel and health factors influencing well-being of the older adults – a case study in China', *Frontiers in Public Health*, Vol. 11, pp.12–19.