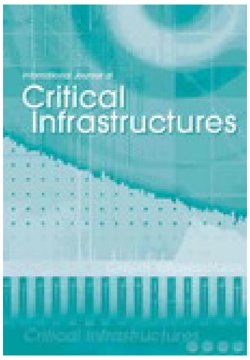


International Journal of Critical Infrastructures

ISSN online: 1741-8038 - ISSN print: 1475-3219
https://www.inderscience.com/ijcis

A smart rural tourism resources recommendation based on audience preference

Jin Lu

DOI: 10.1504/IJCIS.2026.10068089

Article History:

Received:	08 April 2024
Last revised:	27 September 2024
Accepted:	02 October 2024
Published online:	23 September 2025

A smart rural tourism resources recommendation based on audience preference

Jin Lu

School of Tourism and Culinary Arts,
Wuxi Vocational Institute of Commerce,
Wuxi, 214153, China
Email: lujin_lj@outlook.com

Abstract: How to provide users with more accurate smart rural tourism recommendation services has become a hot research topic at present. To address the short-term audience preference issue caused by data scarcity, firstly, graph convolutional networks (GCN) are applied to recommend smart rural tourism resources. For long-term tourism audiences with sufficient data, use long short-term memory (LSTM) to construct a recommendation model based on users' long-term dynamic preferences. The results showed that in the case of data scarcity, the recall and accuracy of the GCN recommendation method increased by 17.9% and 11.8%, respectively. In long-term rural tourism applications, the hits ratio (HR)@10 and HR@20 of the dynamic preference recommendation model were as high as 42% and 50%, respectively. The results indicate that the proposed method provides more reliable technical support for intelligent rural tourism recommendation and can more effectively discover audience preferences.

Keywords: audience preference; rural tourism; resource recommendation; long short-term memory; LSTM; graph convolutional network; GCN.

Reference to this paper should be made as follows: Lu, J. (2025) 'A smart rural tourism resources recommendation based on audience preference', *Int. J. Critical Infrastructures*, Vol. 21, No. 10, pp.1–18.

Biographical notes: Jin Lu received her Bachelor's in Tourism Management from Nanjing Agricultural University in China in 2004 and Master of Public Administration from Nanjing Agricultural University in 2008. Presently, she is a teacher at the School of Tourism and Culinary Arts at Wuxi Vocational institute of Commerce in China, and her areas of interest include rural tourism, smart tourism and vocational education.

1 Introduction

Rural tourism is an important part of tourism. As a sunrise industry, rural tourism has strong linkage with local planting industry and agriculture. It can promote the development of rural economy as a whole, which is an important measure in the national rural revitalisation strategy (Izquierdo-Gascón and Rubio-Gil, 2023). Smart tourism integrates new theories, technologies and methods, which provide inexhaustible impetus for the further development of tourism. In practical applications, smart tourism can provide all-round support for rural tourism. Meanwhile, smart tourism can also help rural tourism scenic spots carry out intelligent management. The organic combination of rural tourism and smart tourism can promote the development of rural tourism, enhance the attractiveness and competitiveness of rural tourism, and promote the development of rural economy. Smart rural tourism has become one of the emerging driving forces for the common prosperity of urban and rural areas (Sia et al., 2023). In smart rural tourism, the demand for tourism services has attracted much attention. The impact of tourism infrastructure on the tourism industry is comprehensive, including improving tourist experience, enhancing scenic attraction, promoting economic development, improving management level, ensuring tourist safety, promoting sustainable development, and improving tourism service quality. A well-developed tourism infrastructure can significantly enhance the tourist experience, including improving the tourist experience, ensuring tourist safety, and improving the quality of tourism services, thereby enhancing the attractiveness of scenic spots, attracting more tourists to visit, and promoting regional economic development (Ciacci et al., 2023). Various projects, including accommodation and scenic spots, greatly affect the quality of tourism. With the continuous development of tourism informatisation, the information on tourism resources in smart rural areas has grown explosively. It is difficult for the vast tourism audience to select the required information quickly and accurately. How to recommend information that meets the preferences of tourism audiences from massive rural tourism resources has become an urgent practical problem. Simultaneously, the traditional collaborative filtering and hybrid recommendation methods are often difficult to extract effective features. There are difficulties in processing heterogeneous auxiliary information data. With the development of deep learning, image and text processing technology has made rapid progress. The recommendation system has ushered in new innovation (Yu et al., 2023). Therefore, the research uses long short-term memory (LSTM) and graph convolutional networks (GCN) in deep learning to predict and analyse the preferences of tourism audiences, with a view to further improving the recommendation quality of rural tourism resources. The proposed model is analysed and the research structure is divided into four parts. The first part is to summarise the domestic and foreign scholars' research on deep learning and resource recommendation. The second part is the construction analysis of the proposed model, and constructs the model. The third part verifies the effectiveness and practical application of the model. The fourth part summarises the experimental results, points out the shortcomings in the research, and puts forward the future research direction.

2 Related works

In recent years, the development of in-depth learning has driven innovation in recommendation systems. Many researchers at home and abroad have conducted in-depth research in this area, providing method references for the recommendation of rural tourism resources. Uyar et al. (2023) explored the correlation between tourist numbers and tourism revenue by examining factors such as tourism infrastructure and favourable policy conditions to determine whether the tourism competitiveness index has promoted the development of the tourism sector. The results indicate that the infrastructure index is positively correlated with tourism revenue, and favourable policy conditions are positively correlated with the number of tourists. Al Fararni et al. (2021) developed a tourism recommendation system based on mixed recommendation method to help tourists find the required tourism information. This method realised the tourist attraction recommendation according to tourists' preferences. The results showed that the method had high recommendation accuracy. Wang et al. (2020) proposed an efficient data transmission method based on 5G technology to deal with the massive data and delayed communication problems in smart tourism. Then the internet of things was applied to tourism data processing. The results showed that the method was effective. Sardianos et al. (2021) designed an interpretable energy-saving recommendation mechanism for intelligent recommendation system. Personalise recommendations was carried out according to user preferences and habits. The result showed that the recommendation acceptance rate increased by 19% in total. To improve the generation method of personalised sorting item list in the recommendation system, Dacrema et al. (2021) compared the recommendation method based on collaborative filtering with the neural method based on the nearest neighbour heuristic. The results showed that the existing learning-based technology reduced the complexity of the recommendation system. To improve the recommendation quality in the recommendation system, Liu et al. (2021a) proposed a recommendation system based on depth matrix decomposition and comment feature learning. An alternative minimisation algorithm was introduced to optimise the loss function. The results showed that the proposed system had higher efficiency than other recommendation systems.

Yu et al. (2021) developed a privacy preserving medical wearable device recommendation algorithm for personalised recommendation of medical wearable devices. The original scoring information was factorised by the orthogonal non-negative matrix three-factor decomposition model. Finally, the user characteristics and target domain information were fused. The results showed that the method was effective. Guo et al. (2021) designed a social internet of things embedded in deep learning to cope with the sharp increase in users' demand for personalised social services. The deep learning technology was applied to process data generated in the social internet of things, achieving smarter and more efficient application services. The internet of things provided online data awareness and management functions. The results showed that it had good robustness. Huang et al. (2021) proposed a two-stage in-depth learning group recommendation method to solve the poor group recommendation effect, and constructed an undirected tripartite graph to learn group semantic features. The results showed that it was superior to the latest baseline for group recommendation. Fathma et al. (2022) proposed a heart prediction method that combined feature selection based on majority voting and deep neural networks. The results showed that the accuracy, sensitivity and F1 score had good performance. Anwar and Uma (2022) used semantic similarity of items to extend cross-domain recommendation for goods and services recommendation. The collaborative filtering method was applied to find similar projects and users. Finally, the Topseq rule mining algorithm was used to implement recommendation. The result showed that this method alleviated the new user data scarcity.

To sum up, most scholars have improved the recommendation system using learning methods such as CNN, achieving good recommendation results. However, the user preference has not been analysed, resulting in low accuracy. Therefore, the research creatively divides tourism audience preferences into long-term and short-term preferences, and improves their recommendation methods respectively. In these two cases, the recommendation models based on GCN and LSTM are proposed respectively, which effectively solve the challenges faced by the smart rural tourism recommendation system. The proposed model will provide new ideas and technical means to solve the audience preference in the smart rural tourism resource recommendation, providing a guarantee for the development and promotion of smart rural tourism.

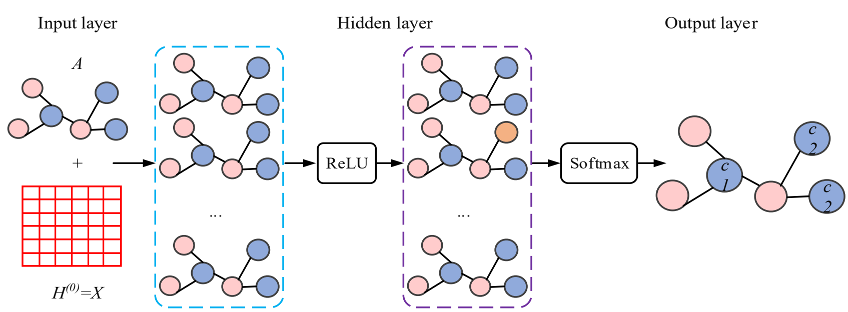
3 Rural tourism resources recommendation based on audience preference

To study the long-term preference and short-term preference of the audience in smart rural tourism, and achieve more accurate recommendation, this paper divides the audience's visit sequence into different sub-sequences by LSTM and GCN respectively, analyses the audience's characteristic preferences, visit preferences in different periods and current status, and then recommends travel places.

3.1 Tourism recommendation based on short-term audience preference

Some audiences have few interactive records. In view of the scarce data and short-term audience preference, the research proposes a rural tourism resource recommendation method based on GCN. GCN is able to extract and explore the features and patterns in the graph structure data through neural network learning, which has been well applied in social networks and recommendation systems (Ohtomo et al., 2021). In GCN, graph G is composed of edges and vertices, which is a data structure. Several data with connection can be expressed as graphs, which are represented by adjacency matrix. Under the action of specific functions, GCN maps the nodes contained in the graph into Euclidean space, which actually belongs to the neural network layer. The simple framework is shown in Figure 1.

Figure 1 A simple framework of graph convolution neural network (see online version for colours)



In GCN, the way to realise spread between layers is shown in formula (1).

$$H^{1+l} = \left(H^l W^l \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}\right) \cdot \sigma \tag{1}$$

In formula (1), H^l is the implicit vector of the first level node. $\tilde{A}$ is the adjacency matrix containing self-circulation. σ is the activation function. $\tilde{D} = \sum_{i=1}^{N} \tilde{A}_{ij}$ is the degree matrix of $\tilde{A}$. W^l is a trainable parameter of the first layer. GCN uses graph convolution to aggregate the information of nodes and neighbours, and cope with data sparsity and cold start to some extent (Liu et al., 2021b). However, when GCN is extended to higher fields, the number of learnable parameters increases, which increases the computational workload (Zhang et al., 2022). Under the average aggregation effect, nodes are excessively smooth. It is difficult to mine the potential preferences of the audience. To tap into the audience's potential preferences, the PageRank algorithm is introduced. Combined with the PageRank algorithm, factors such as correlation and weight among tourist locations can be further explored, so as to improve the accuracy and personalisation of user preferences in the recommendation system. PageRank algorithm is a 'random walk model', which uses links between pages to achieve page search ranking calculation. A tourist site is taken as a node. The probability of the node being visited by the audience is shown in formula (2).

$$PageRank(v) = \alpha \cdot \sum_{P_i} \frac{PageRank(P_i)}{Outdegree(P_i)} + \frac{1-\alpha}{N} \tag{2}$$

In formula (2), α is the damping factor. N is the total number of nodes contained in the graph convolution network. $\alpha \cdot \sum_{P_i} \frac{PageRank(P_i)}{Outdegree(P_i)}$ is the probability that the audience may access the node v. $\alpha \cdot \sum_{P_i} \frac{PageRank(P_i)}{Outdegree(P_i)}$ is the probability that the user can access the node v using the link. The total probability is obtained by adding the probabilities. PageRank combined with GCN can effectively aggregate infinite neighbourhood information in node embedding vectors, avoiding node smoothing. Therefore, the research combines the two to promote the audience preference to spread to a higher neighbourhood. The recommended method proposed in the research includes the embedding layer, the information self-coding layer, the information dissemination aggregation layer and the prediction layer. The framework is shown in Figure 2.

In Figure 2, the main function of the embedding layer is to initialise the node and embed it into a vector representation to express the characteristics of the node. In the information self-coding layer, the algorithm will self-encode the initial node features to extract the relationship and importance between the features. The information dissemination aggregation layer combines the GCN and PageRank to realise the effective aggregation of multi-neighbourhood information, avoiding the excessive smoothing of node characteristics. This can achieve audience preference spreading to higher communities. The prediction layer is the final layer that is used to predict user preferences for items or generate recommendations. By combining these layers, the proposed PTRGCN method can effectively combine GCN and PageRank algorithm. This achieves effective information aggregation and audience preference dissemination, thereby improving the performance of recommendation systems. A lookup table X is established to represent the initial values of audience u and tourist attractions i. X is shown in formula (3).

$$X = \left[x_{u_1}, x_{u_2}, \ldots, x_{u_N}, x_{i_1}, x_{i_2}, \ldots, x_{i_M} \right] \tag{3}$$

In formula (3), N represents the number of audiences. M is the number of tourist attractions. All audiences and tourist attractions are regarded as nodes in the diagram, so that all preference information can be spread in the interaction diagram between audiences and tourist attractions. Therefore, while paying attention to local neighbours, nodes can use information diffusion to obtain higher connection between audiences and tourist attractions. The cooperation information is embedded into the audience and scenic spots embedding vector to effectively alleviate the data sparsity. In the audience-attraction interaction, all attractions or audiences are interrelated. The direct interactive attraction of the audience is the local neighbour. The interactive connection and high-level connection between the audience and the tourist attraction are shown in Figure 3.

Figure 2 Diagram of rural tourism resources recommendation method based on graph convolution neural network (see online version for colours)

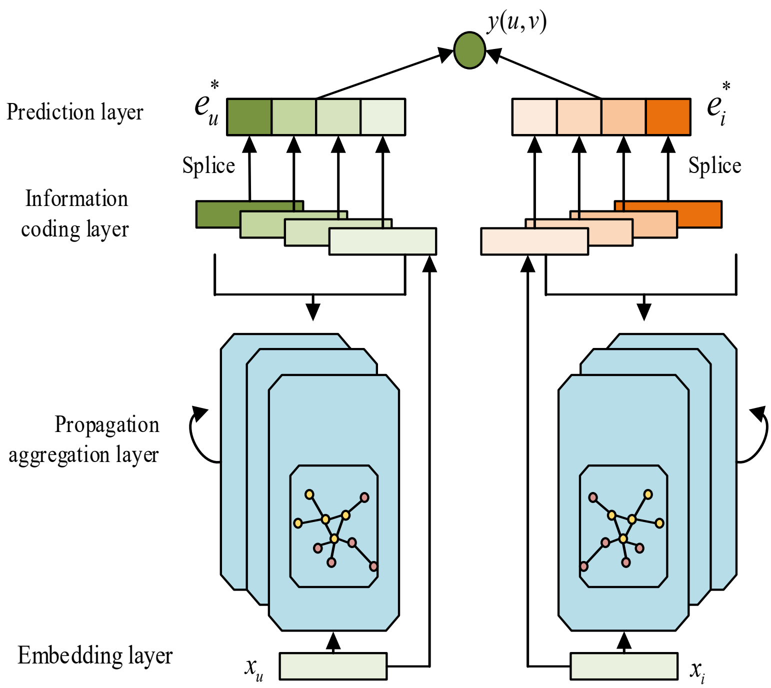
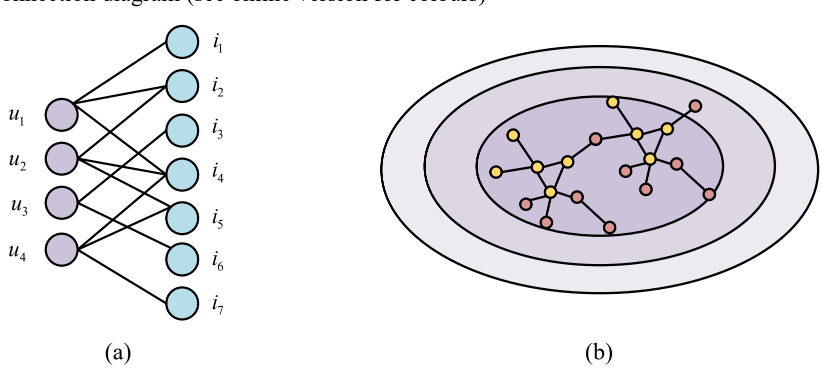


Figure 3 Interactive connection and high-level connection between audiences and tourist attractions, (a) user-attraction interaction diagram (b) user-attraction high-level connection diagram (see online version for colours)



The tourist attractions i_1, i_2 and i_4 in Figure 3 represent local neighbours of audience u_1, indicating the direct preference evidence of the audience. The potential preferences of the audience are also affected by the local neighbours. Because the audience u_2 visits the tourist attractions i_1 and establishes a connection with i_4, u_2 becomes the second-order neighbours of u_1. u_2 reflects the audience u_1's second-order potential preferences. Meanwhile, neighbour nodes that are far away from the audience u_1 have a smaller impact on the preferences of u_1. It is a process of continuous attenuation. Similar to the water wave, the audience's preference is shown as personalised communication on the interaction diagram. Therefore, even if the audience has only a few sparse interactions, it can achieve the preference diffusion in higher-order neighbourhoods through this connection and preference relationship, thus obtaining higher-order potential preferences. When the audience node x_u reaches the tourist attraction node x_i through random walk, the preference information x_u is spread to x_i, the PageRank is used to spread and adjust x_u, as shown in formula (4).

$$m(x_u) = \left[I_n - (1-a)\tilde{\tilde{A}} \right]^{-1} x_u \cdot a \tag{4}$$

In formula (4), m is message embedding. a represents the transmission probability. $I(u, i)$ represents the impact score of x_i on x_u. Different root nodes have different scores. The matrix spread of PageRank propagation rule is shown in formula (5).

$$\begin{cases} Z^0 = X = H \\ Z^{1+l} = aH + (1-a)\tilde{\tilde{A}} Z^l \end{cases} \tag{5}$$

In formula (5), Z^l represents the embedded vector of audience and tourist attractions obtained through l step. X represents the value of initial message iteration. The effective embedding vector is filtered by a self-coder, which used for the final model prediction, as shown in formula (6).

$$E^l = W_1^l Z^l \tag{6}$$

In formula (6), W_1^l represents the learnable parameters of each layer, which can extract useful information and realise message aggregation. After the multiple embedding vectors are obtained, the corresponding embedding vectors of the audience and tourist attraction nodes at different communication layers are concatenated to obtain the final embedding vector, as shown in formula (7).

$$e_u^* = e_u^0 \| \ldots \| e_u^L, \; e_i^* = e_i^0 \| \ldots \| e_i^L \tag{7}$$

In formula (7), $\| \ldots \|$ refers to the splicing operation. e_u^L is the multiple embedding vector of the audience. e_i^L is the multiple embedding vector of the tourist attractions. Finally, the preference of the target tourist attractions is estimated by the inner product, as shown in formula (8).

$$y(u, i) = e_i^* e_u^{*T} \tag{8}$$

Then the optimised Bayesian personalised ranking (BPR) loss function is used to learn the model parameters, as shown in formula (9).

$$Loss = \sum_{(u,i,j)\in O} -\ln\sigma\left(\hat{y}_{iu} - \hat{y}_{ju}\right) + \|\Theta\|_2^2 \cdot \lambda \tag{9}$$

In formula (9), O represents paired training data. $\sigma(\cdot)$represents Sigmoid function. Θ is all trainable parameters. λ is regularisation parameter that controls the over-fitting.

3.2 *Tourism resource recommendation based on long-term audience preference*

Smart rural tourism audience is not dominated by short-term behaviour. A large number of audiences have long-term tourism visit records. In view of this situation, a recommendation method based on LSTM is proposed. RNN is the predecessor of LSTM with parameter sharing and memory (Chen, 2021). RNN takes time series data as input. Recursion is implemented according to the time sequence, so as to keep the information continuously. Each unit in RNN contains two inputs, namely the current time input and the last time input (Gauch et al., 2021). The basic cycle structure is shown in Figure 4.

Figure 4 Basic cycle structure and expansion form of RNN (see online version for colours)

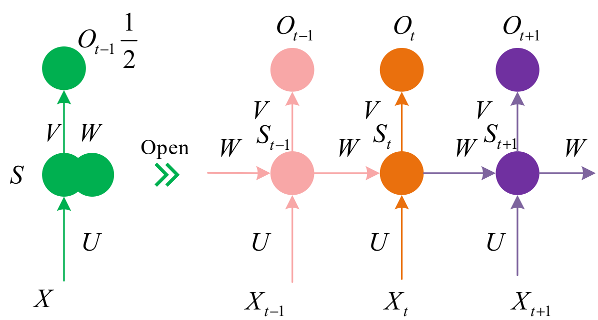
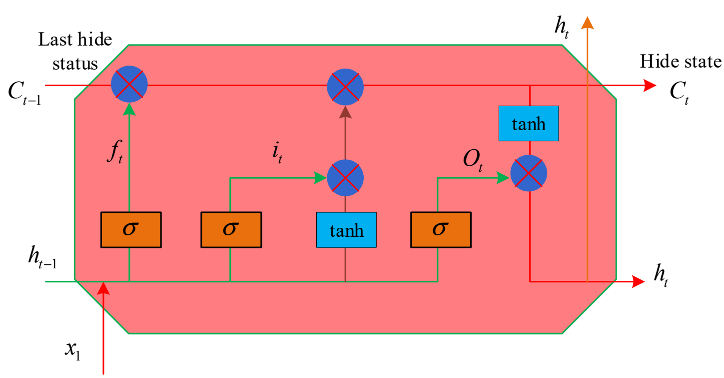


Figure 5 Basic structure of LSTM network unit (see online version for colours)



According to the structural characteristics of RNN, the infinite length sequence can be processed in theory. However, in practical application, there is a gradient explosion problem caused by too deep network depth, which leads to increased difficulty in training and long-term dependence on feature learning. LSTM adds output gate, forgetting gate and output gate to control the information of the previous moment. When it passes to the next moment, the long-term dependence problem is effectively solved (Wang et al., 2022). The basic structure of LSTM network unit is shown in Figure 5.

The audience's rural tourism behaviour is determined by short-term order preference, long-term preference and current environment. Simultaneously, interest will change over time. The research divides the audience's access sequence into different sub-sequences according to the time interval. Through LSTM, CNN and external memory matrix, the audience's long-term characteristic preferences, visit preferences in different periods and current status are analysed. The next tourist destination is recommended. The rural tourism resource prediction method proposed in the study mainly includes four parts, current access status, audience dynamic long-term preference memory, audience short-term sequence preference based on LSTM and prediction level. Based on the LSTM, the audience's long-term preference for tourism resources is accurately recorded and updated through a separate memory matrix (Wirthmüller et al., 2021). Meanwhile, considering that there are certain differences in the importance of rural tourist attractions in a single visit sequence, the long-term dynamic preferences of the audience are obtained by extracting the characteristics of the tourist attractions through the attention unit. In the rural tourism resource recommendation, the next attraction recommendation is highly related to the recent visit. Therefore, CNN is used to extract the local features of the nearest locations, so as to obtain the characteristics of the audience's current access status. According to the audience's long-term preference representation, current visit status representation, and sequence preference representation, the probability that the audience is most likely to visit the tourist destination can be obtained by inputting these three into a fully connected network. The basic framework of the proposed recommendation method is shown in Figure 6.

Figure 6 Framework of the proposed recommended method (see online version for colours)

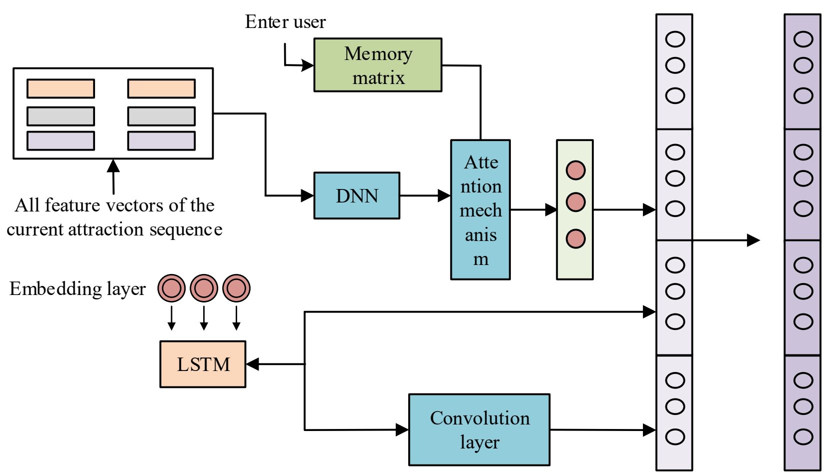


Figure 6 shows the LSTM recommendation method based on dynamic user preferences. A user memory matrix can be used in a recommendation system to predict a user's possible future interests and choices by analysing their historical behaviour and preferences. All feature vectors of the current scenic spot sequence are the feature vectors of each scenic spot in the current scenic spot sequence, which can include various features such as the geographical location of the scenic spot, the number of historical tourists, and user ratings. These feature vectors can be used to describe the attributes and characteristics of each scenic spot. Attention gates are used to control how much attention the network pays to inputs. By introducing the attention gate, the model can pay more attention to the key information in the input, thus improving the performance and generalisation ability of the model. The LSTM layer is mainly used to process time series data. The convolutional layer is mainly used to extract features from the input data. In the audience dynamic long-term memory, the research records long-term preferences through a key value. The long-term preference of the audience is highly related to the attribute characteristics of rural tourist attractions. These attribute preferences is reflected in the access record. The study establishes a corresponding feature memory matrix for all audiences to store their preferences for certain features. When the feature vector of the new rural tourist attraction sequence itself is input, the corresponding audience's feature preference vector is read by the memory matrix. Combined it with the attention gate unit and the tourist attraction feature, the fine-grained interest is obtained, as shown in formula (10).

$$m^u = \sum_{i=0}^{k} m_i^u w_i^u \tag{10}$$

In formula (10), w_k^u refers to the attention degree that the audience shows to the attribute characteristics of rural tourist attractions. m^u is fine-grained interest. The memory matrix of the audience is automatically updated with the current visit sequence. Because each visit record in the visit sequence has certain differences in the importance of the audience's long-term interest, the study screens the features of important rural tourism spots in the sequence through the attention unit. All the features in the input tourist attraction sequence is processed by the attention unit, so as to obtain the features that have the highest correlation with the long-term interest of the audience, as shown in formula (11).

$$f_k^i = f_k^i \otimes \sigma\left(f_k^i \cdot W_{g1} + m_k^u \cdot W_{g2} + b_g\right) \tag{11}$$

In formula (11), E_m represents the characteristics of tourist attractions. W_{g1}, W_{g2} and b_g represent learnable parameters. m_k^u represents the preference vector of audiences k in the memory matrix. After passing through the attention gate, the features in the tourist attractions sequence that have low correlation with the audience's long-term preferences is ignored. The implicit vector representation corresponding to the features k in the current tourist attractions sequence is obtained by using the average pooling operation. Then the implicit vector representation is used as the query value. The audience's preference degree for the feature is obtained by inner product with the preference vector. Finally, the softmax operation is performed on the preference degree of all features to obtain the preference weight, as shown in formula (12).

$$w_k^u = softmax\left(w_k^u\right) = \frac{\exp\left(w_k^u\right)}{\sum_{i=0}^{K} \exp\left(w_i^u\right)} \tag{12}$$

In formula (12), $r_{i/o}$ is the final preference weight. The long-term preference of the audience is dynamic, which is influenced by the current visit sequence. Therefore, it is necessary to update the memory matrix to the long-term preference vector according to the current scenic spot characteristics. The update degree of each long-term feature preference vector is controlled by a gate control unit to update the memory matrix, as shown in formula (13).

$$\begin{cases} z = W_z * \Phi\left(f_k^l, m_k^u\right) \\ m_k^u \leftarrow m_k^u \cdot (1-z) + f_k^l \cdot z \end{cases} \tag{13}$$

In formula (13), f_k^l is a hidden vector. Φ is inner product operation. W_z is learnable parameter. To obtain the final preferences of the audience, the long-term preference representation, short-term preference in LSTM, current access status representation and audience representation are spliced and connected to the full connection layer. Then the softmax is used to predict. The probability of all items for the audience during interaction is shown in formula (14).

$$y^{u,t} = \sigma\left(b + W\left(h_t^u, m_t^u, c_t^u, p^u\right)\right) \tag{14}$$

In formula (14), W and b are trainable parameters. m_t^u is long-term preferences. h_t^u is sequence preferences. c_t^u is current status. p^u is user representation. $y^{u,t}$ is probability. Finally, the optimisation is realised through the random gradient descent method and the cross entropy loss function. The objective function is shown in formula (15).

$$L = -\sum_{u}^{|U|}\sum_{t}^{I^u} \hat{y}^{ut}\log\left(\hat{y}^{ut}\right) + \left(1-\hat{y}^{ut}\right)\log\left(1-\hat{y}^{ut}\right) + \|\theta\|^2 \cdot \lambda \tag{15}$$

In formula (15), θ represents all parameters in the network. $\hat{y}^{ut}$ is the one-hot code of the audience u at t. λ is regularisation weight parameter.

4 Verification of recommendation effect of smart rural tourism resources based on audience preference

4.1 Recommended effect of short-term audience preference

To verify the effectiveness of the rural tourism resource recommendation method integrating PageRank and graph convolution network (PGCN), it is compared with the recommendation methods based on BPR, neural graph collaborative filtering (NGCF) and graph convolution matrix completion (GCMC). The dataset information comes from an annual data from Ctrip, Fliggy and Qunar websites. The dataset is the self-built rural tourism resource dataset (SRTR), which mainly comes from the representative rural tourism data in Jiangsu, Hainan and Chongqing, containing more than 400,000 tourist attractions records. This dataset is publicly available. The dataset contains information about tourist attractions, tourist distribution, and records of tourist behaviour at the attractions. A total of 100 scenic spots are involved in the dataset. These scenic spots are all natural scenery, historical culture, and rural customs. Tourists come from all over

China. Each attraction is recorded between 0 and 9,999. At the same time, to filter the noise records, the records of tourist attractions visited less than 10 times are deleted. The experimental environment is shown in Table 1.

Table 1 Experimental environment for research

Operating environment	*Specific configuration*
Operating system	64-bit Ubuntu 16.04 operating system
CPU	Intel (R) Core (TM) i7
GPU	Single NVIDIA GTX TITAN
Platform	Python 3 and TensorFlow deep learning platform
TensorFlow version	2.6.0
Python version	3.7

Figure 7 shows the operation results of the four methods in the SRTR dataset. From Figure 7(a), compared with the traditional BPR algorithm, the proposed PGCN method increases the recall value by about 9.7% and the precision by 5.7% for the warm start audience. From Figure 7(b), for the cold start audience, the recall and precision of PGCN method are improved by 17.9% and 11.8% respectively compared with BPR method. It shows that the method proposed in the study is better for cold start audiences, because it can effectively utilise the potential and deep connection between nodes. It is more suitable for cold start audiences.

Figure 7 Comparison of recommendation effects of four methods in cold start and warm start audiences, (a) warm start user (b) cold start user (see online version for colours)

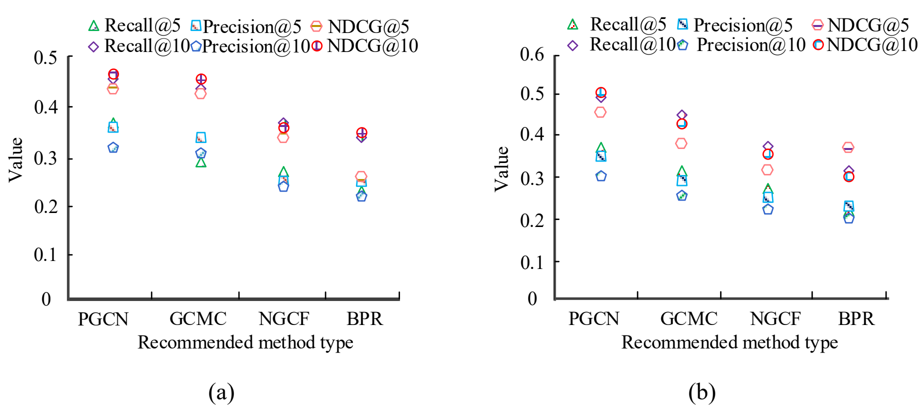


Figure 8 shows the recommended performance changes of the proposed PGCN method when the propagation depth of the graph model changes. From Figure 8(a), for the cold start audience, when the number of graph propagation layers of the model is 4, the index Recall@10 reaches the maximum value of 0.5. Compared with the fifth floor Recall@10, the index values are basically the same, indicating that this method can effectively deal with the excessive smoothing. From Figure 8(b), for the warm start audience, as the number of layers increases, the index value of Recall@10 fluctuates more frequently. Compared with the cold start audience, this method is less sensitive to the increase of the number of layers. In general, this method can improve the recommendation effect for cold start audiences by increasing the number of layers.

Figure 8 The recommended performance change of the proposed PGCN method under the change of graph model propagation depth, (a) cold start user (b) warm start user (see online version for colours)

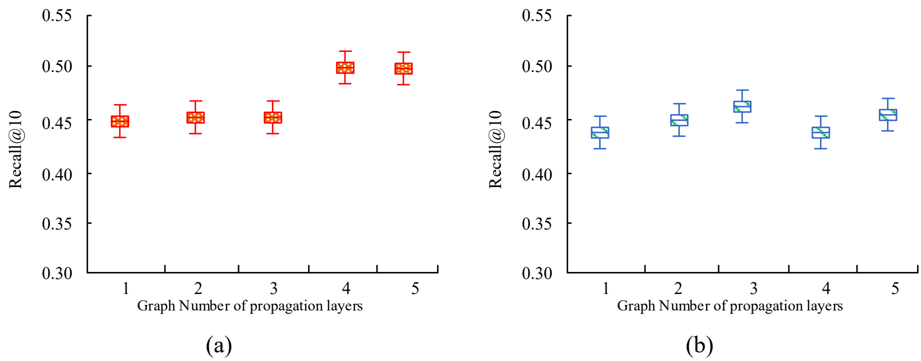
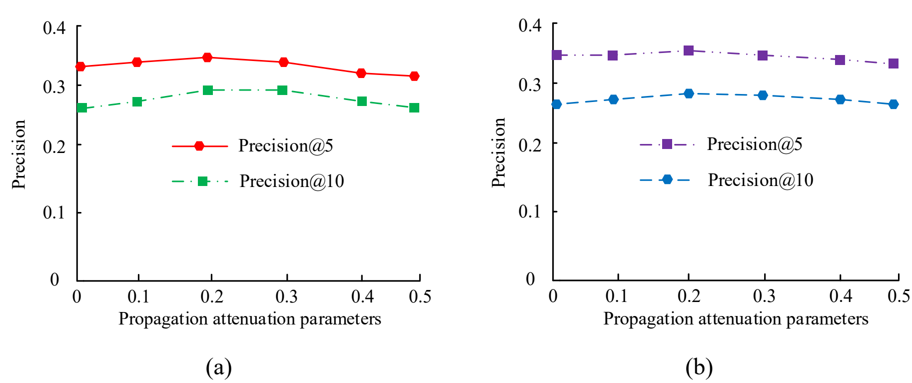


Figure 9 shows the recommended performance changes of PGCN method under the change of propagation attenuation parameters. From Figure 9, when the propagation attenuation parameter is 0, the recommended effect of the model is the worst. When the parameter is 0.2, the Precision@5 and Precision@10 are the highest, with 0.35 and 0.29, respectively. The value of warm start audience's Precision@5 and Precision@10 is the highest, with 0.36 and 0.28, respectively. After the propagation attenuation parameter is greater than 0.2, the recommended performance gradually deteriorates. When the propagation attenuation parameter is 0.2, the recommendation performance of the model is better.

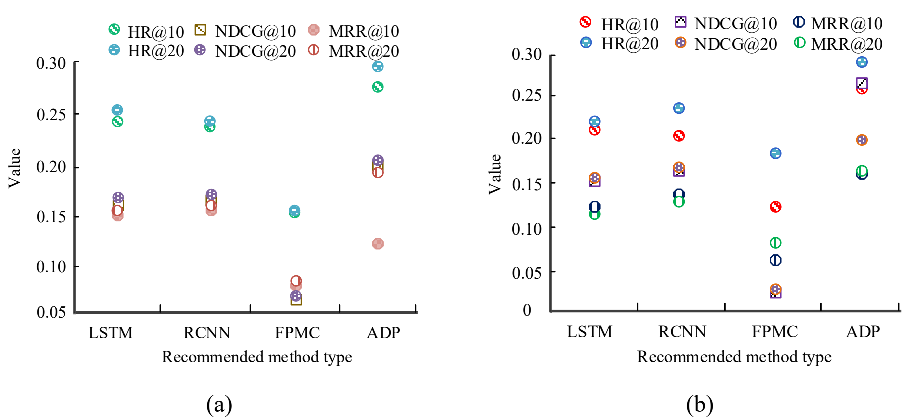
Figure 9 Recommended performance change of PGCN method under the change of propagation attenuation parameters, (a) precision of cold start user (b) precision of warm start user (see online version for colours)



4.2 Recommended effect of long-term audience preference

In view of the timing of audience tourism, the research proposes a rural tourism recommendation method based on audience dynamic preference (ADP). To verify its recommendation effect, it is first run in the New York dataset (NYC) dataset and Tokyo (TKY) dataset. Then it is compared with the recommendation methods based on recurrent revolution neural network (RCNN), gated recurrent unit (GRU) and factorised personalised Markov chains (FPMC). The number of tourist destinations is 7,222 and the audience is 1,939. The recommended performance is based on three indicators, hits ratio (HR), mean reciprocal rank (MRR) and normalised discounted cumulative gain (NDGG). The NYC dataset is from Foursquare data, which mainly contains the check-in data collected in New York, filtering out audiences with less than ten visit records and ten tourist sites with ten visits. The number of tourist sites participating in the experiment is 4,638, with a total of 1,036 audiences. The dataset for the city of Tokyo comes from the Foursquare dataset, which mainly includes check-in data collected in Tokyo, audiences with fewer than ten visits and ten tourist attractions with ten visits. The TKY dataset comes from Tokyo, with 2,584 tourist attractions and a total audience of 903 people. The two datasets contain information about tourist attractions, visitor distribution, and records of tourist behaviour at the attractions. The attractions include natural scenery, historical culture, and rural customs. Tourists come from all over the world. Each attraction is recorded between 0 and 9,999. The processing results of the four methods selected in the study in the dataset NYC and TKY are shown in Figure 10. From Figure 10(a), in the NYC dataset, the recommended performance of FPMC is the worst, with the values of the three indicators below 0.2. The NDGG and MRR indicators are below 0.1. Compared with GRU, the HR@10 of the RCNN increases by 1.6%. The HR@10 of the proposed ADP method is about 13.2% higher than FPMC, which is the best in the four methods. From Figure 10(b), in the TKY dataset, the three indicators of GRU and RCNN are in the range of 0.10~0.25. FPMC is still less than 0.2. The maximum of ADP method is 0.3, and the minimum is still above 0.15, which is about 12% higher than RCNN.

Figure 10 Study the processing results of the four methods selected in the dataset NYC and TKY, (a) NYC dataset run results (b) TKY dataset running results (see online version for colours)



To further verify the performance of the proposed ADP method, three variants of the ADP method, namely LSTM, LSTM-CNN and LSTM-USER, are designed to prove the importance of each module proposed in the ADP method. The dataset is the self-built SRTR. The effect of each module in the ADP method on the recommendation accuracy is shown in Figure 11. In terms of indicators HR@10, LSTM, LSTM-CNN and LSTM-USER modules are 30%, 32% and 36%, respectively. The ADP method integrating the three modules can reach 42%, with an increase of 12%, 10% and 8%, respectively. In terms of HR@20, the LSTM, LSTM-CNN and LSTM-USER modules are 37%, 38% and 40%, respectively. The ADP method is nearly 50%, increased by 13%, 12% and 10%, respectively. In general, after integrating the three modules, the ADP method has good accuracy in recommending rural tourism resources.

Figure 11 Results of the influence of each module in the ADP method on the HR, (a) impact results of each module on indicator HR@10 (b) impact results of each module on indicator HR@20 (see online version for colours)

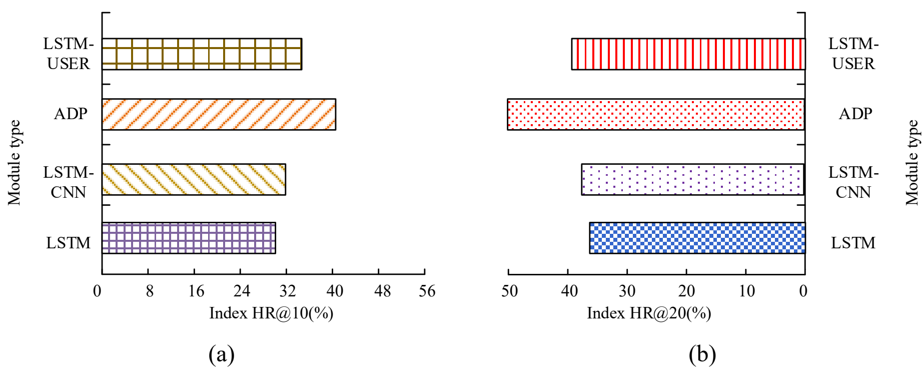
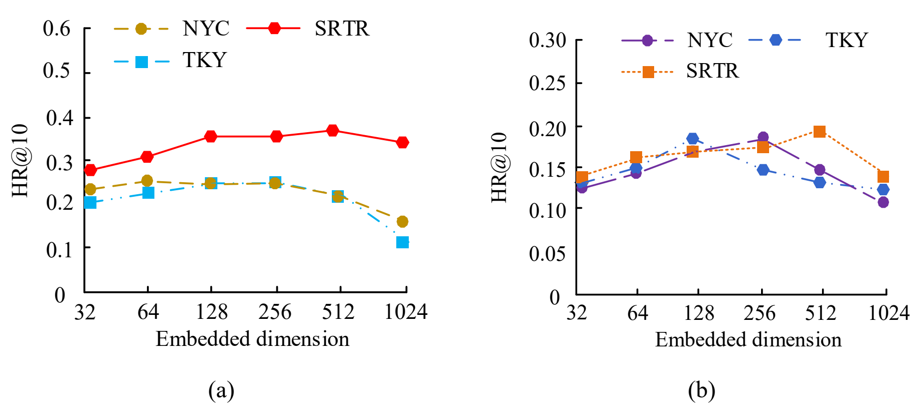


Figure 12 Influence of embedded dimension parameters on ADP method performance, (a) recommended indicator HR@10 results (b) recommended indicator NDGG@10 results (see online version for colours)



The memory unit dimension has a great impact on the performance of the LSTM. In the established recommendation model, the memory unit dimension in LSTM is consistent with the embedded dimension of audience and tourist attractions. Therefore, the research finally analyses the impact of embedded dimension parameters on ADP performance.

The results are shown in Figure 12. From Figure 12, when the embedding dimension is low, the recommendation accuracy is low. With the increase of embedded dimensions, the accuracy of recommendation gradually improves. As the dimension continues to increase, the difficulty of online training increases. The accuracy of recommendation is affected. In the NYC and TKY datasets, ADP indicators achieve the best performance in 256 and 128 dimensions respectively. In the SRTR dataset, when the ADP model is at 512 dimensions, the indicators HR and NDCG reach the highest value of 0.38 and 0.19, respectively, indicating that the method has the best performance at 512 dimensions.

5 Discussion

In recent years, the deep integration of the internet and tourism has spawned a new development hotspot of smart rural tourism, which has become an important driving force and direction for rural construction. For the recommendation service of rural tourism resources, researchers have put forward a variety of innovative recommendation models to improve the recommendation effect and meet the needs of rural tourism market. Among them, the recommendation method, which combines PageRank and GCN, shows remarkable results. This method can improve the recall and precision for the warm start audience, effectively deal with the over-smoothing problem for the cold start audience, and provide a new idea for the recommendation system optimisation. In addition, the recommendation model based on ADP and LSTM temporal modelling, has achieved satisfactory recommendation results when audience data is sufficient. In the future research, the combination with knowledge graph, augmented reality and other technologies, as well as the deep integration with other new technologies, is expected to further improve the smart rural tourism recommendation service effect, promote the development of rural tourism, and inject new vitality into the rural revitalisation strategy.

6 Conclusions

The organic combination of internet and tourism has brought new development hotspots. Smart rural tourism has become an important driving force and direction of rural construction. The research focuses on the recommendation service of rural tourism resources. Firstly, a recommendation method integrating PageRank and PGCN was proposed. Then, aiming at the timing of audience's long-term tourism, based on LSTM, a recommendation model based on ADP was designed. The effect was verified. The results showed that compared with the traditional BPR algorithm, the recall value of the PGCN method for the hot start audience increased by 9.7%. The precision increased by 5.7%. For cold start audiences, when the number of graph propagation layers of PGCN method was 4, the index Recall@10 reached the maximum value of 0.5, which was basically the same as the number of layers 5. It could effectively deal with the excessive smoothing. When the audience data was sufficient, the index HR@10 of the designed ADP recommendation method in NYC dataset was about 13.2% higher than that of FPMC. In TKY dataset, the maximum was 0.3, and the minimum was still above 0.15. Compared with RCNN, it increased by 12%, which showed the recommendation effect that the corresponding PGCN and ADP methods achieved in different audience preferences. However, the research does not use knowledge map and other technologies to mine the fine-grained preference of the audience in tourism information. In this regard, further research is needed to improve the recommendation effect. The model proposed in this study did not take into account the running speed, which is crucial to the recommendation effect of the model. Therefore, the subsequent research can consider the running speed of the model.

References

Al Fararni, K., Nafis, F., Aghoutane, B., Yahyaouy, A., Riffi, J. and Sabri, A. (2021) 'Hybrid recommender system for tourism based on big data and AI: a conceptual framework', *Big Data Mining and Analytics*, Vol. 4, No. 1, pp.47–55.

Anwar, T. and Uma, V. (2022) 'CD-SPM: cross-domain book recommendation using sequential pattern mining and rule mining', *Journal of King Saud University-Computer and Information Sciences*, Vol. 34, No. 3, pp.793–800.

Chen, S. (2021) 'Design of internet of things online oral English teaching platform based on long-term and short-term memory network', *International Journal of Continuing Engineering Education and Life Long Learning*, Vol. 31, No. 1, pp.104–118.

Ciacci, A., Ivaldi, E., Mangano, S. and Ugolini, G.M. (2023) 'Environment, logistics and infrastructure: the three dimensions of influence of Italian coastal tourism', *Journal of Sustainable Tourism*, Vol. 31, No. 7, pp.1583–1607.

Dacrema, M.F., Boglio, S., Cremonesi, P. and Jannach, D. (2021) 'A troubling analysis of reproducibility and progress in recommender systems research', *ACM Transactions on Information Systems (TOIS)*, Vol. 39, No. 2, pp.1–49.

Fathima, M.D., Samuel, S.J. and Natchadalingam, R. (2022) 'Majority voting ensembled feature selection and customized deep neural network for the enhanced clinical decision support system', *International Journal of Computers and Applications*, Vol. 44, No. 10, pp.991–1001.

Gauch, M., Kratzert, F. and Klotz, D. (2021) 'Rainfall-runoff prediction at multiple timescales with a single long short-term memory network', *Hydrology and Earth System Sciences*, Vol. 25, No. 4, pp.2045–2062.

Guo, Z., Yu, K., Li, Y., Srivastava, G. and Lin, J.C.W. (2021) 'Deep learning-embedded social internet of things for ambiguity-aware social recommendations', *IEEE Transactions on Network Science and Engineering*, Vol. 9, No. 3, pp.1067–1081.

Huang, Z., Liu, Y., Zhan, C., Lin, C., Cai, W. and Chen, Y. (2021) 'A novel group recommendation model with two-stage deep learning', *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 52, No. 9, pp.5853–5864.

Izquierdo-Gascón, M. and Rubio-Gil, Á. (2023) 'Theoretical approach to Api-tourism routes as a paradigm of sustainable and regenerative rural development', *Journal of Apicultural Research*, Vol. 4, No. 62, pp.751–766.

Liu, H., Zheng, C., Li, D., Shen, X., Lin, K., Wang, J. and Xiong, N.N. (2021a) 'EDMF: efficient deep matrix factorization with review feature learning for industrial recommender system', *IEEE Transactions on Industrial Informatics*, Vol. 18, No. 7, pp.4361–4371.

Liu, S., Wang, B., Deng, X. and Yang, L.T. (2021b) 'Self-attentive graph convolution network with latent group mining and collaborative filtering for personalized recommendation', *IEEE Transactions on Network Science and Engineering*, Vol. 9, No. 5, pp.3212–3221.

Ohtomo, K., Harakawa, R. and Ogawa, T. (2021) 'Personalized recommendation of Tumblr posts using graph convolutional networks with preference-aware multimodal features', *ITE Transactions on Media Technology and Applications*, Vol. 9, No. 1, pp.54–61.

Sardianos, C., Varlamis, I., Chronis, C., Dimitrakopoulos, G., Alsalemi, A., Himeur, Y. and Amira, A. (2021) 'The emergence of explainability of intelligent systems: delivering explainable and personalized recommendations for energy efficiency', *International Journal of Intelligent Systems*, Vol. 36, No. 2, pp.656–680.

Sia, P.Y.H., Saidin, S.S. and Iskandar, Y.H.P. (2023) 'Systematic review of mobile travel apps and their smart features and challenges', *Journal of Hospitality and Tourism Insights*, Vol. 6, No. 5, pp.2115–2138.

Uyar, A., Kuzey, C., Koseoglu, M.A. and Karaman, A.S. (2023) 'Travel and tourism competitiveness index and the tourism sector development', *Tourism Economics*, Vol. 29, No. 4, pp.1005–1031.

Wang, C., Li, Z., Outbib, R., Dou, M. and Zhao, D. (2022) 'A novel long short-term memory networks-based data-driven prognostic strategy for proton exchange membrane fuel cells', *International Journal of Hydrogen Energy*, Vol. 47, No. 18, pp.10395–10408.

Wang, W., Kumar, N., Chen, J., Gong, Z., Kong, X., Wei, W. and Gao, H. (2020) 'Realizing the potential of the internet of things for smart tourism with 5G and AI', *IEEE Network*, Vol. 34, No. 6, pp.295–301.

Wirthmüller, F., Klimke, M., Schlechtriemen, J., Hipp, J. and Reichert, M. (2021) 'Predicting the time until a vehicle changes the lane using LSTM-based recurrent neural networks', *IEEE Robotics and Automation Letters*, Vol. 6, No. 2, pp.2357–2364.

Yu, T., Yu, D., Wang, D. and Hu, X. (2023) 'Web service recommendation for mashup creation based on graph network', *Journal of Supercomputing*, Vol. 79, pp.8993–9020.

Yu, X., Zhan, D., Liu, L., Lv, H., Xu, L. and Du, J. (2021) 'A privacy-preserving cross-domain healthcare wearables recommendation algorithm based on domain-dependent and domain-independent feature fusion', *IEEE Journal of Biomedical and Health Informatics*, Vol. 26, No. 5, pp.1928–1936.

Zhang, J., Ye, G. and Tu, Z. (2022) 'A spatial attentive and temporal dilated (SATD) GCN for skeleton-based action recognition', *CAAI Transactions on Intelligence Technology*, Vol. 7, No. 1, pp.46–55.