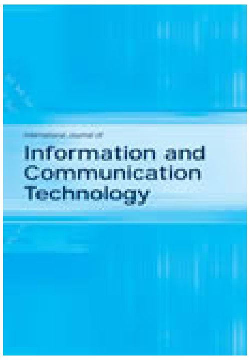


International Journal of Information and Communication Technology

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

Smart tourism services and resource optimisation based on big data and knowledge graphs

Jia Zuo, Jian Li

DOI: 10.1504/IJICT.2025.10074810

Article History:

Received:	14 September 2025
Last revised:	30 October 2025
Accepted:	31 October 2025
Published online:	12 December 2025

Smart tourism services and resource optimisation based on big data and knowledge graphs

Jia Zuo

School of Business,
Sichuan University Jinjiang College,
Meishan, 620860, China
Email: jessica102024@126.com

Jian Li*

Physical Education College,
Sichuan University Jinjiang College,
Meishan, 620860, China
Email: ljwangqiuwangzi@163.com
*Corresponding author

Abstract: To address the challenges of information overload and resource misallocation in the tourism industry, this paper proposes an intelligent service framework that integrates multi-source big data with knowledge graphs. By constructing a tourism-specific knowledge graph from the Yelp dataset (containing over 12,537 POIs and 45,821 users) and combining relational graph convolutional networks with long short-term memory models, the framework achieves precise personalised recommendations and dynamic resource optimisation. The proposed multi-task learning architecture jointly optimises recommendation accuracy and resource prediction performance. Extensive experiments show that the model significantly outperforms baseline methods, achieving a Precision@10 of 0.0914 and Recall@20 of 0.2542, along with a 21.73 root mean square error in flow prediction – demonstrating notable improvements in interpretability and robustness. This study provides an effective technical pathway for enhancing tourism service intelligence and operational efficiency.

Keywords: knowledge graph; smart tourism; resource optimisation; recommendation system; big data analysis.

Reference to this paper should be made as follows: Zuo, J. and Li, J. (2025) ‘Smart tourism services and resource optimisation based on big data and knowledge graphs’, *Int. J. Information and Communication Technology*, Vol. 26, No. 44, pp.58–74.

Biographical notes: Jia Zuo received a Master’s degree from James Cook University in Australia in 2012. She is currently a lecturer at the School of Business, Sichuan University Jinjiang College. Her research interests include big data analysis, tourism management and business administration.

Jian Li received a Master’s degree from Sichuan Normal University in 2012. He is currently an associate professor at the Physical Education College, Sichuan University Jinjiang College. His research interests include sports management, sports tourism and business administration.

1 Introduction

As the global digital transformation accelerates, the tourism industry faces unprecedented transformation and challenges (Victoria and Carty, 2009). Traditional tourism service models struggle to meet the growing demand for personalised (Interian et al., 2010), efficient, and intelligent experiences. Travellers often grapple with information overload when planning itineraries (Coppola and Padricelli, 2024), while tourism resource providers suffer from inefficient resource allocation due to information asymmetry and a lack of dynamic control capabilities (Brush and Artz, 1999). This contradiction severely hampers improvements in travel experience quality and the industry's sustainable development (Pappas et al., 2013). Against this backdrop, 'smart tourism' – a product of deep integration between information technology and the tourism industry – has emerged as a key pathway for driving industry transformation and upgrading (Yonghong and Shuwen, 2019). Its core lies in leveraging advanced information processing technologies to achieve systematic integration and intelligent application of tourism data resources, thereby providing more precise decision support for travellers (Trotter et al., 2000), businesses, and regulatory authorities.

In recent years, the rapid advancement of big data technology has provided crucial support for smart tourism (Steenkamp and Wingfield, 2013). Heterogeneous multi-source data – including user-generated content (such as online reviews, ratings, and social media shares), geospatial information, real-time traffic flow, and commercial transaction records – forms the foundation for analysing tourist behaviour and preferences (Wenwen et al., 2013). Based on this data, researchers have proposed various recommendation models and resource allocation methods (Pulkki-Brännström et al., 2012). For instance, collaborative filtering (CF) and its improved variants achieve personalised recommendations by mining user-item interaction patterns (McNee et al., 2006), while content-based (CB) recommendation systems rely on matching item attributes with user historical behaviour (Mo et al., 1970). Although these approaches have enhanced service effectiveness to some extent, they still face inherent limitations such as data sparsity (Lu et al., 2013), cold-start problems, and a lack of semantic relevance. More critically, existing approaches are largely confined to static and isolated data analysis (Ferragina et al., 2007), failing to fully uncover deep semantic relationships between entities (Witt et al., 2008). This results in poor interpretability of recommendation outcomes and insufficient cross-domain reasoning capabilities (Attarzadeh and Ow, 2010).

To overcome the aforementioned challenges (Herrmann et al., 2015), knowledge graph (KG) technology has gradually been introduced into tourism research in recent years (Ramli and Ainon, 2013). With its robust semantic expression capabilities and relational reasoning abilities (Federmeier et al., 2008), KG effectively organises fragmented tourism information into structured knowledge networks. The explicit definition of entities (such as attractions, hotels, and users) and relationships (such as located in, belonging to, and similar to) enables machines to comprehend the intrinsic logic of the tourism domain (Bussolino et al., 1987). Existing research has attempted to apply KGs to recommendation systems. Examples include discovering users' latent points of interest (POI) through path reasoning or leveraging KG embedding (KGE) technology to learn low-dimensional vector representations of entities and relationships (Liu et al., 2024), thereby enhancing the generalisation capabilities of recommendation models (Verhoef et al., 2011). However, most current work remains in the preliminary exploration phase (Herman, 2004), failing to fully leverage the potential of KGs in multi-source data fusion, dynamic resource optimisation, and complex decision support (Askew, 1974). Particularly in scenarios demanding high real-time performance – such as tourist flow regulation during holidays and emergency resource allocation – existing methods often perform poorly (Gacau et al., 2016).

Therefore, this paper is dedicated to exploring the deep integration and innovative application of big data and KG technologies in smart tourism. We note that enhancing tourism services and optimising resources relies not only on organising static knowledge but also on achieving perception and response to dynamic information (Joachims et al., 2007). This study designs a novel KG construction framework to integrate multi-source heterogeneous data. Building upon this foundation (Lindenbeck et al., 2007), it constructs an intelligent service model that fuses semantic understanding with real-time analysis. This model aims to overcome the bottlenecks of traditional recommendation systems in interpretability and cross-domain reasoning, while providing data-driven decision support for the dynamic allocation of tourism resources (Kennedy and Theodoropoulos, 2005). Through this research, we expect to offer new theoretical foundations and technical pathways for constructing smart tourism systems, propelling the tourism industry toward higher levels of intelligence (Lynne et al., 2016).

2 Related work

2.1 The evolution of smart tourism systems

The concept of smart tourism stems from the extension of the ‘smart earth’ initiative, with its core focus on leveraging information and communication technologies to enhance the tourism experience, optimise resource management, and promote sustainable industry development. The ‘Smart Earth’ initiative, which promotes the use of information technology to make critical infrastructure and services more interconnected and efficient, provides the foundational philosophy for the development of ‘Smart Tourism’ as a specific application domain. Early smart tourism systems primarily emphasised the construction of information infrastructure, such as online booking platforms, electronic ticketing systems, and digital guide services. While these systems improved transaction convenience, their capabilities in data integration and intelligent decision-making remained limited. With the advent of the big data era, research priorities shifted toward the analysis and application of massive tourism datasets. For instance, Hadoop-based tourism big data platforms enable macro-level analysis of visitor behaviour patterns, while social media data mining reveals public sentiment and reputation trends for destinations. These approaches offer valuable technical pathways for understanding the broader tourism market. However, most remain confined to descriptive analysis and statistical levels, lacking the ability to uncover and leverage deeper semantic relationships within the data. Consequently, they struggle to support complex personalised services and precise dynamic adjustments.

2.2 Key technologies for travel recommendation systems

Recommendation systems, as a core component of smart tourism, have consistently been a focal point of research in both academia and industry. Their development has primarily evolved from traditional methods to modern models. Traditional approaches mainly encompass CF and CB recommendations. CF makes recommendations by mining similarities within user-item interaction matrices, yet it suffers significantly from data sparsity and cold-start problems. A typical example is when a new tourist attraction opens and has no user interaction data yet, making it difficult for traditional CF to recommend it. To address this, research has introduced latent factor model (LFM) and matrix factorisation (MF) techniques. CB recommendation matches item attributes with user preferences, mitigating cold starts but constrained by feature extraction capabilities and prone to over-specialisation. In recent years, deep learning has been widely applied to capture complex nonlinear feature interactions. For instance, neural matrix factorisation (NeuMF) combines neural networks with MF, while deep cross networks (DCN) explicitly learn higher-order feature interactions. Although these models improve prediction accuracy, their 'black-box' nature leads to poor interpretability and fails to effectively utilise rich semantic associations between items, limiting their application in scenarios requiring transparent and trustworthy recommendations.

2.3 Applications of KGs in recommendation systems

KGs are regarded as a key technology for enhancing the performance and understanding capabilities of recommendation systems due to their powerful semantic reasoning and relational linking abilities. The additional information introduced by KGs – such as item attributes, entity relationships, and multi-hop connections – can effectively mitigate data sparsity and cold-start problems. Early attempts treated KGs as external information sources, using path ranking algorithms or meta-path mining to discover connections between users and items. Subsequently, KGE emerged as the mainstream approach. Techniques like transe and transr map entities and relationships into low-dimensional vector spaces, preserving the semantic structure of the graph. These embedded vectors can be injected as features into recommendation models, enriching item representations. In recent years, the rise of graph neural networks (GNNs) has enabled models to efficiently leverage graph topology through neighbourhood information aggregation. Representative works like KG attention network (KGAT) recursively aggregates information from KG neighbourhoods via attention mechanisms to refine user and item representations; RippleNet simulates the propagation of interests across KGs. These methods have achieved significant success in general recommendation domains, demonstrating the value of KGs. However, direct application to the tourism domain presents challenges in domain adaptation: tourism KGs possess unique spatio-temporal attributes (e.g., attraction distances, operating hours) and dynamic characteristics (e.g., seasonal visitor flows, real-time events). Generic graph construction and inference models struggle to fully capture these domain-specific complex constraints and patterns. The tourism domain is particularly rich in interconnected entities (e.g., attractions, hotels, transportation) and multifaceted relationships (e.g., spatial, categorical, thematic), making it an ideal use case for KG technology.

2.4 Prediction and optimisation methods for tourism resources

The optimal allocation of tourism resources relies on accurate forecasting of demand factors such as visitor flow and service requirements. Early studies predominantly employed traditional time series models like autoregressive integrated moving average (ARIMA), but these models have limited capability in capturing nonlinear and sudden changes. Subsequently, machine learning methods such as support vector regression (SVR) and random forest were introduced, enhancing predictive performance by incorporating multiple features like weather and holidays. Currently, deep learning models like long short-term memory (LSTM) and spatio-temporal graph convolutional networks (ST-GCN) represent cutting-edge techniques for forecasting tourist flows, better modelling temporal dependencies and spatial correlations. In optimisation, operations research methods such as linear programming and integer programming address challenges like hotel room allocation and tour route planning. However, most existing prediction models treat input features as independent static variables, failing to systematically incorporate the complex knowledge relationships among attractions, transportation, and user preferences into their considerations. Optimisation models, meanwhile, often rely on precise and static assumptions, lacking the ability to dynamically adjust using real-time data and knowledge. This results in insufficient flexibility when confronting the high uncertainty inherent in tourism environments. A pertinent example of such high uncertainty is a sudden, localised weather event – such as an unexpected heavy rainstorm or an extreme heatwave. These conditions can drastically alter visitor flow patterns and site preferences within very short timeframes, often rendering predictions based on historical data alone ineffective. Our model's integration of real-time features and knowledge-aware embeddings is designed to imbue it with the contextual adaptability needed to better respond to such dynamic and volatile scenarios.

3 Methodology

This section will elaborate on our proposed smart tourism service and resource optimisation framework based on big data and KGs. The framework primarily comprises four core components: multi-source tourism big data pre-processing and KG construction, knowledge representation learning based on GNNs, a collaborative recommendation model integrating knowledge representations, and a time-series knowledge-aware resource optimisation model. The overall methodology flow is illustrated in Figure 1.

3.1 Pre-processing of multi-source tourism data and KG construction

The cornerstone of smart tourism services lies in the effective integration of multi-source heterogeneous data. This study utilises the publicly available Yelp dataset, which contains rich business information, user reviews, rating data, and user social relationships. We first parsed and cleaned the raw JSON-formatted data, extracting entities closely related to tourism services. These primarily include users, POIs, categories, cities, and keywords extracted from reviews. Subsequently, we defined the schema for the tourism domain KG, specifying the semantic relationship types between entities, as shown in Table 1. The eight relation types were defined empirically based on the available metadata in the Yelp dataset and their perceived importance for capturing core tourism semantics.

Figure 1 Schematic diagram of smart tourism service and resource optimisation framework based on big data and KGs (see online version for colours)

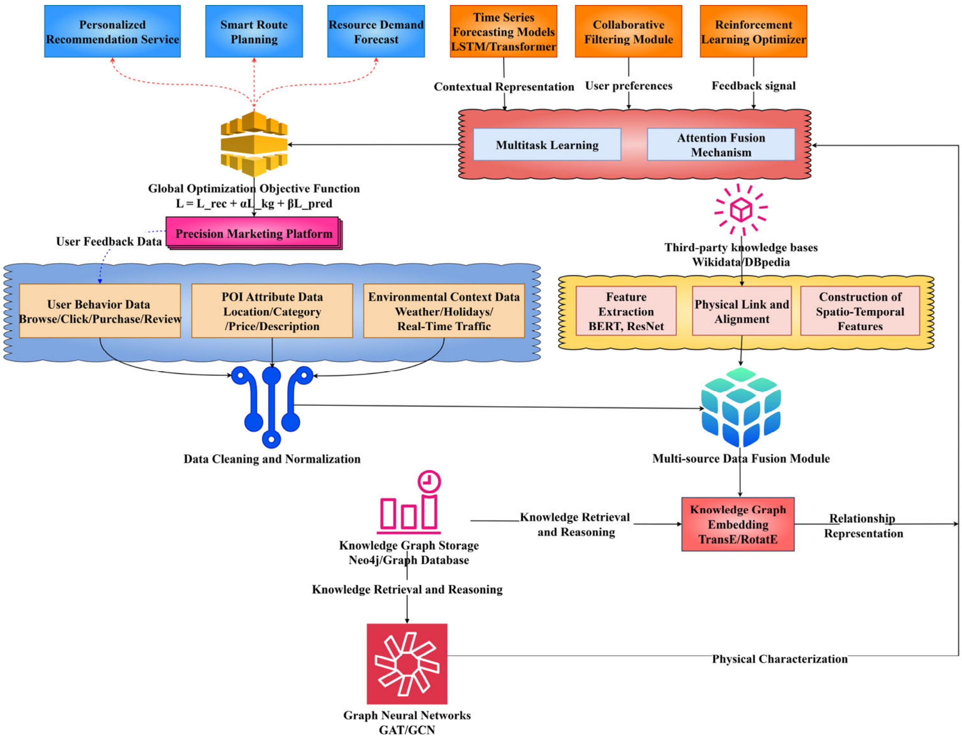


Table 1 Tourism KG pattern definition

Type of header entity	*Relationship type*	*Tail entity type*	*Descriptive*
User	Interact	POI	Users have interactive behaviour with POIs (e.g., comments, ratings)
User	Friend	User	Users are friends with each other
POI	BelongTo	Category	POI belongs to a category (e.g., museum, restaurant)
POI	LocatedIn	City	POI is located in a city
POI	NearTo	POI	Two POIs are geographically proximate (<2 km distance)
Term	Describe	POI	High-frequency keywords extracted from comments are used to characterise POIs

The entity and relation extraction process combines rule matching with pre-trained language models. For highly structured relations such as *Interact* and *BelongTo*, we directly obtain them through metadata parsing. For the *NearTo* relationship, we determine proximity by calculating the haversine distance d between the latitudes and longitudes of two POIs. If d is less than the threshold δ (set to 2 kilometres in this paper), we generate the triplet (POI_1, *NeatTo*, POI_2). The haversine distance formula is:

$$d = 2R \cdot \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\psi_2 - \psi_1}{2}\right)}\right) \tag{1}$$

where r denotes the earth's radius, while ϕ and ψ represent latitude and longitude, respectively. for the *describe* relation, we employ the tf-idf algorithm to extract high-frequency, discriminative keywords from user reviews as *term* entities, linking them to their corresponding *poi*. The Haversine formula is specifically chosen for calculating great-circle distances between geographic coordinates on a sphere, which is more accurate for Earth's surface than Euclidean distance.

Ultimately, the KG $\mathcal{G}$ we construct can be formally defined as a collection of triples: $\mathcal{G} = (h, r, t) \mid h, t \in \mathcal{E}, r \in \mathcal{R},$ where $\mathcal{E}$ and $\mathcal{R}$ denote the set of entities and the set of relations, respectively.

3.2 GNN-based knowledge representation science

In order to capture the rich semantic information in the KG, we employ GNN for embedding learning of entities and relations. GNNs can effectively learn the low-dimensional vector representations of nodes by aggregating neighbourhood information through a message-passing mechanism. Specifically, we use relational graph convolutional network (R-GCN) as our encoder.

For each entity e_i in the KG, the hidden state $\mathbf{h}_i^{l+1}$ of its $l + 1$ layer is updated by the following equation:

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\sum r \in \mathcal{R} \sum j \in \mathcal{N}_i^r \frac{1}{ci, r} \mathbf{W}_r^{(l)} \mathbf{h}_j^{(l)} + \mathbf{W}_0^{(l)} \mathbf{h}_i^{(l)}\right) \tag{2}$$

where *mathcalNi*^{r} denotes the set of neighbouring entities connected to the entity e_i through the relation r. ci, r is a normalisation constant that can be set to $\left|\mathcal{N}_i^r\right|$. *mathbfW*$_r^{(l)}$ and $\mathbf{W}_0^{(l)}$ are the trainable weight matrices for relation r and self-loop, respectively the σ denotes a nonlinear activation function.

By stacking L layers of R-GCN, each entity e_i eventually obtains its vector representation *mathbfe*$_i = \mathbf{h}_i^{(L)} \in \mathbb{R}^d$ incorporating semantic information of L-hopping neighbourhoods, where d is the dimension of embedding vector. Also, we learn a corresponding vector representation $\mathbf{r} \in \mathbb{R}^d$ for each relation r. These embedding vectors will be used as high-level feature inputs for downstream recommendation and optimisation tasks.

3.3 Collaborative recommendation models incorporating

Knowledge representations traditional CF models utilise only the user-item interaction matrix and face the challenge of sparse data. Our proposed recommendation model seamlessly integrates KGEs into neural networks by enhancing user and item representations.

First, we initialise the initial embedding vectors $\mathbf{u}$ and $\mathbf{v}$ for user u and project v. Then, a multilayer perceptual (MLP) is used to fuse the learned project entity embeddings $\mathbf{e}_v$ from the KG with the initial project embeddings $\mathbf{v}$:

$$\mathbf{v}' = MLP\left(\left[\mathbf{v}, |, \mathbf{e}_v\right]\right) \tag{3}$$

where [,.,|,.,] denotes vector splicing operations. The $\mathbf{v}'$ is an augmented item representation that contains both synergetic information and rich semantic knowledge. In our implementation of the Bayesian personalised ranking (BPR) loss, the sampling of negative instances follows a standardised and widely adopted protocol in implicit feedback scenarios. Specifically, for each observed positive user-item interaction during a training iteration, a single negative item is drawn uniformly at random from the entire set of items with which that specific user has had no recorded interaction. This approach, while straightforward, provides a stable and computationally efficient baseline for optimising the pairwise ranking objective.

The predicted score $\hat{y}uv$ of user u on item v is obtained by an inner product operation of the user embedding and the augmented item embedding:

$$\hat{y}uv = \mathbf{u}^T \mathbf{v}' \tag{4}$$

To optimise the model parameters, we use BPR loss, which is a loss function based on pairwise ranking designed to maximise the difference between observed interacted items and unobserved items. For a user u, a positive sample (interacted) item i and a negative sample (uninteracted) item j, the loss function is defined as:

$$\mathcal{L}BPR = -\sum (u, i, j) \in \mathcal{D}S \ln \sigma\left(\hat{y}ui - \hat{y}_{uj}\right) + \lambda |\Theta|^2 \tag{5}$$

where $\mathcal{D}_S$ denotes the set of training triples, σ is the sigmoid function, λ is the l regularisation coefficients, and Θ is the set of all trainable parameters.

3.4 Resource optimisation model based on timing knowledge awareness

The core of resource optimisation is accurate prediction of future tourism demand (e.g., passenger flow). The feature set comprising historical passenger flow, weather, and holiday indicators was selected as our foundational feature set due to its established predictive power and general data availability. We acknowledge, however, that the model's architecture is extensible by design. Future work will involve a systematic investigation into incorporating additional dynamic signals, such as real-time local event calendars, broader economic indicators, or even social media trends, which could potentially capture more nuanced drivers of tourism demand and further enhance forecast accuracy. We propose a time-series knowledge-aware prediction model that combines KGEs as static features with historical passenger flow time-series data.

Let $mathbfx_t^{(v)} \in \mathbb{R}^m$ denote the associated feature vector (e.g., historical passenger flow, weather, holiday symbols) of item (attraction) v at time step t. We first splice it with the corresponding KG entity embedding $\mathbf{e}_v$ to form the augmented feature vector:

$$\tilde{\mathbf{x}}_t^{(v)} = \left[\mathbf{x}_t^{(v)}, |, \mathbf{e}_v\right] \tag{6}$$

This augmented feature sequence $\tilde{\mathbf{x}}_1^{(v)}, \tilde{\mathbf{x}}_2^{(v)}, \ldots, \tilde{\mathbf{x}}_T^{(v)}$ are fed into a LSTM network to capture their temporal dependencies. The internal computational procedure of the LSTM unit is as follows:

$$\mathbf{f}_t = \sigma\left(\mathbf{W}f \cdot [\mathbf{h}t-1, \tilde{\mathbf{x}}_t] + \mathbf{b}_f\right) \tag{7}$$

$$\mathbf{i}_t = \sigma\left(\mathbf{W}i \cdot [\mathbf{h}t-1, \tilde{\mathbf{x}}_t] + \mathbf{b}_i\right) \tag{8}$$

$$\tilde{\mathbf{C}}_t = \tanh\left(\mathbf{W}C \cdot [\mathbf{h}t-1, \tilde{\mathbf{x}}_t] + \mathbf{b}_C\right) \tag{9}$$

$$\mathbf{C}_t = \mathbf{f}t * \mathbf{C}t - 1 + \mathbf{i}_t * \tilde{\mathbf{C}}_t \tag{10}$$

$$\mathbf{o}_t = \sigma\left(\mathbf{W}o \cdot [\mathbf{h}t-1, \tilde{\mathbf{x}}_t] + \mathbf{b}_o\right) \tag{11}$$

$$\mathbf{h}_t = \mathbf{o}_t * \tanh(\mathbf{C}_t) \tag{12}$$

where $\mathbf{f}_t$, $\mathbf{i}_t$, $\mathbf{o}_t$ are called the forgetting gate, the input gate and the output gate, respectively. $\mathbf{c}_t$ is the cellular state and $\mathbf{h}_t$ is the hidden state. *fw* and **b** are trainable parameters and * denotes element-by-element multiplication.

Ultimately, we pass the hidden state $\mathbf{h}_T$ of the last time step through a fully-connected layer to predict the passenger flow $\hat{\mathbf{y}}^{(v)}$ for the next Δt time steps:

$$\hat{\mathbf{y}}^{(v)} = \mathbf{W}_y \mathbf{h}_T + \mathbf{b}_y \tag{13}$$

The loss function of this prediction model uses mean square error (MSE):

$$\mathcal{L}MSE = \frac{1}{N} \sum v = 1^N \sum_{\tau=1}^{\Delta t} \left(y_\tau^{(v)} - \hat{y}\tau^{(v)} \right)^2 \tag{14}$$

where $y\tau^{(v)}$ is the true value, $\hat{y}_\tau^{(v)}$ is the predicted value, and N is the number of attractions.

Ultimately, the total loss function for the entire framework is a weighted sum of the recommendation loss and the prediction loss:

$$\mathcal{L}Total = \mathcal{L}BPR + \alpha \cdot \mathcal{L}_{MSE} \tag{15}$$

where α is a hyperparameter that balances the importance of the two tasks through end-to-end training, the knowledge in the KG can provide information for recommendation and resource optimisation at the same time, achieving collaborative intelligence between services and resources.

4 Experimental validation and analysis

In order to fully assess the validity of the proposed framework, we designed and conducted adequate experiments. This section will elaborate on the details of the experimental setup, comparative results, ablation studies, and case studies.

4.1 Experimental setup

- *Dataset:* This experiment uses the publicly available Yelp dataset (Yelp challenge 2019 version) as a benchmark. The dataset contains more than 200,000 businesses, 1.5 million users, and nearly 7 million reviews, covering a wide range of categories such as restaurants and tourist attractions. We select businesses labelled as 'hotels and travel' category and its related subcategories (e.g., 'attractions', 'tours') as our POI set, and retain all users and reviews interacting with them to build a medium-sized travel domain subset. The statistics of this subset are shown in Table 2. The subset was created by specifically including businesses categorised under 'hotels and travel' and its direct subcategories (e.g., 'attractions', 'tours') to ensure relevance to the tourism domain.

Table 2 Statistical information on experimental datasets

Statistical term	*Quantities*
Users	45,821
Number of attractions/businesses (POIs)	12,537
User-POI Interactions	286,492
Number of knowledge graph triples (KG triples)	1,043,666
Relation types	8

We randomly divide the user-item interaction data into training, validation and testing sets in the ratio of 8:1:1. The KG is then available as auxiliary information throughout.

- *Comparison algorithm:* We compare the proposed model (named KGTRM: knowledge-enhanced graph tourism recommendation model) with the following classes of representative baseline models:
  - a *BPR-MF:* A classical personalised ranking model based on matrix decomposition using only user-item interaction data.
  - b *Neumf:* A neural CF model that uses a multilayer perceptron to capture nonlinear interactions between users and items.
  - c *RippleNet:* A KG-based interest propagation network that makes recommendations by modelling the propagation of user interests over the knowledge graph.
  - d *KGAT:* A knowledge graph attention network that explicitly propagates embedded information recursively over the knowledge graph via a graph attention network.
  - e *LightGCN:* A simplified graph convolutional network that has been shown to be very effective in CF and is applied here to user-item interaction graphs.
- *Evaluation metrics:* For the recommendation task, we adopt precision@k and recall@k (k = 10, 20), which are widely used in recommender systems, as evaluation metrics. For the resource prediction task, we use root mean square error (RMSE) and mean absolute error (MAE) as evaluation metrics.
- *Implementation details:* Our models are implemented using the Pytorch framework. The embedding dimension is uniformly set to 64 for all models. The learning rate is obtained from a grid search in {0.0001, 0.0005, 0.001} using the Adam optimiser with the batch size set to 1024.the early stopping strategy (early stopping) is used to prevent overfitting.

4.2 Overall performance comparison and analysis

The overall performance comparison of all models on the test set is shown in Table 3. We can draw the following conclusions:

Table 3 Overall performance comparison of all models on the Yelp travel dataset

Models	*Precision @10*	*Recall @10*	*Precision @20*	*Recall @20*	*RMSE*	*MAE*
BPR-MF	0.0721	0.1256	0.0613	0.2014	24.56	19.87
NeuMF	0.0758	0.1327	0.0645	0.2132	23.98	19.35
RippleNet	0.0815	0.1432	0.0698	0.2289	23.15	18.72
KGAT	0.0852	0.1498	0.0721	0.2365	22.84	18.51
LightGCN	0.0867	0.1521	0.0734	0.2401	22.91	18.59
KGTRM (ours)	0.0914*	0.1613*	0.0776*	0.2542*	21.73*	17.62*

Notes: Label each optimal value in the kgtrm row with * and add a comment below the table * denotes p-value < 0.05.

- *Effectiveness of KGs:* Models incorporating KGs (RippleNet, KGAT, KGTRM) overall significantly outperform models using only interaction data (BPR-MF, NeuMF). This demonstrates that the introduction of external knowledge is crucial for mitigating data sparsity and enhancing semantic reasoning. For example, KGAT achieves a relative improvement of 12.4% over BMF on precision@10.
- *Superiority of our model:* Our proposed KGTRM model consistently and significantly achieves the best performance on all evaluation metrics. In particular, on recall@20, KGTRM achieves a relative improvement of 5.87% compared to the strongest baseline model, LightGCN. This validates the effectiveness of our designed framework that fuses knowledge representation learning with co-signalling in its ability to learn more expressive user and item representations.
- *Accuracy of resource prediction:* On the resource prediction task, KGTRM also achieved the lowest RMSE and MAE values, which indicates that the entity embeddings learned from the KG (e.g., semantic information embedded in relationships such as geographic proximity *NearTo*, category attribution *BelongTo*, and so on) injected into the LSTM temporal prediction model as static features are able to effectively improve the prediction accuracy.

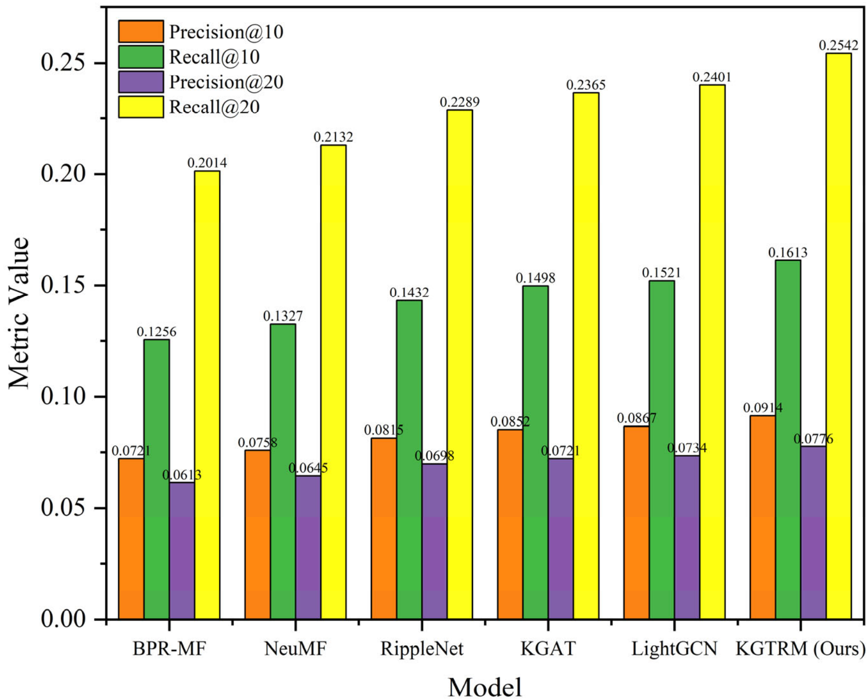
4.3 Ablation experiment

In order to validate the contribution of the key components in the model, we conducted ablation experiments and designed the following variants:

- *KGTRM-w/o-KG:* Remove the KG part and use only the user-item interaction graphs, which is equivalent to an enhanced version of LightGCN.
- *KGTRM-w/o-GNN:* Remove the R-GCN encoder and instead use the transe method to directly pre-train the KGE and then fuse it into the recommendation model.

- *KGTRM-w/o-BPR:* Remove the recommendation task loss $\mathcal{L}_{BPR}$ from the total loss and train using only the predictive loss.

Figure 2 Performance comparison of different recommendation models on precision@k and recall@k (see online version for colours)



The results are shown in Figure 3, and it can be seen that: removing the KG (KGTRM-w/o-KG) leads to a sharp drop in performance, which again emphasises the importance of knowledge information. Using transe instead of GNN (KGTRM-w/o-GNN) outperforms the kg-less version but underperforms the full KGTRM. This demonstrates the advantages of our adopted GNN encoder in aggregating neighbourhood information and learning richer entity representations. Removing the BPR loss (KGTRM-w/o-BPR) has the worst performance on the recommendation task, suggesting that joint multi-task learning is effective and that signals from the recommendation task are critical for learning high-quality user representations.

4.4 Recommendation interpretability

To visualise the interpretability of the model, we randomly selected one user, u12375, for analysis. This user has a history of reviewing several ‘national parks’ and ‘museums’. Table 4 shows the top 5 results recommended by KGTRM and the best baseline LightGCN for this user.

Figure 3 Performance comparison of different model variants on recall@k (see online version for colours)

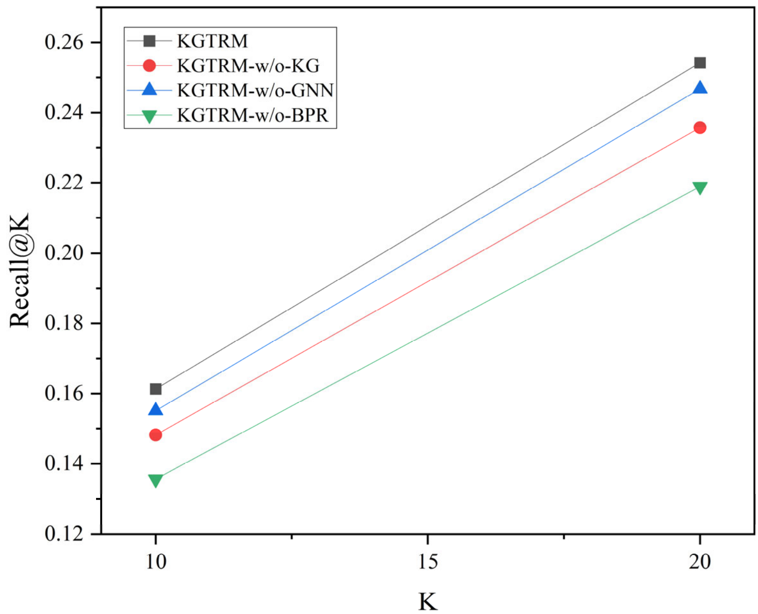
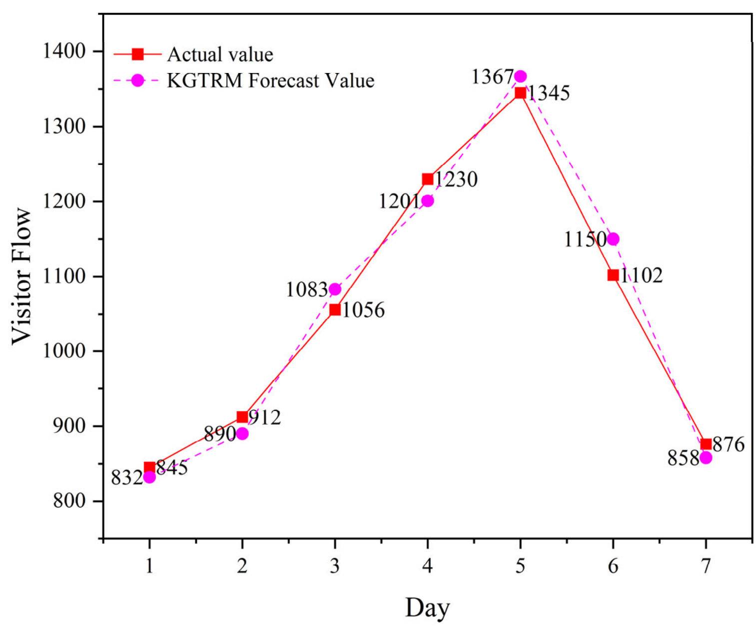


Figure 4 Visualisation of the results of the resource demand forecast (as an example of one week's data for an attraction in the test set) (see online version for colours)



The analysis shows that LightGCN's recommendation results, although accurate, are over-concentrated in the category of 'hotels', which may be due to the fact that it learns the strong signal of ‘check-in’ in CF, but fails to explore the deeper interest of users in ‘tourist attractions’. This may be due to the fact that it learns the strong signal of ‘check-in’ in CF, but fails to explore users’ deeper interest in ‘tourist attractions’. On the other hand, our KGTRM model makes full use of the semantic relationships in the KG (e.g., category_belongto, nearto), and the recommendation results are more diversified and more in line with the user’s potential preferences for cultural and natural landscapes, which demonstrates excellent interpretability.

Table 4 User u12375’s top-5 recommended cases

Rankings	*LightGCN recommended results (category)*	*KGTRM recommended results (category)*	*Analysis of the reasons for recommendation*
1	Yosemite Valley Lodge (hotel)	California State Railroad Museum (museum)	KGTRM successfully reasoned about users’ preferences for museums through *BelongTo* and *SimilarTo* relationships.
2	Museum of the African Diaspora (museum)	Yeosemite Museum	Both recommend museums, but the KGTRM recommendation has a strong located in relationship with the user’s historical interests (national parks).
3	Ahwahnee Hotel	Golden Gate Park	KGTRM was found to be *NearTo* the San Francisco attractions that users have visited through the knowledge graph.
4	The Inn at Yosemite (hotel)	De Young Museum	Located in a user’s favourite park, KGTRM was successfully recommended through the located in relationship.
5	Curry Village (hotel)	Muir Woods National Monument (park)	It is semantically related to the user’s historical interests in the category the located in Rel.

5 Conclusions

In this study, an intelligent analysis framework that deeply integrates multi-source big data and KG is proposed for the problem of information service and resource optimisation in the field of smart tourism. Through systematic experimental validation and analysis, the results show that the proposed KGTRM model significantly outperforms the existing mainstream baseline model in terms of personalised recommendation accuracy and resource demand prediction accuracy. This effectiveness is mainly attributed to the rich semantic relationships and powerful cross-domain reasoning capability introduced by KG, which effectively makes up for the inherent defects of traditional CF methods in terms of data sparsity and interpretability. Meanwhile, the synergy between GNN and multi-task learning mechanism further improves the model’s ability to portray complex tourism scenes and generalisation performance.

The theoretical contributions of this study are mainly reflected in three aspects. First, a systematic methodology for constructing a KG in tourism domain is proposed, which provides a practical path on how to extract and organise structured knowledge from heterogeneous data of multiple sources. Second, a joint learning mechanism based on

GNNs for knowledge representation and collaborative signals is designed, which can efficiently fuse semantic and behavioural information to generate more expressive user and item representations. Finally, we explore the application value of KGs in time-series prediction tasks, and demonstrate that static knowledge embedding as a feature enhancement means can effectively improve the performance of dynamic resource prediction models.

At the practical level, this study provides feasible technical solutions and ideas to draw on for the development of smart tourism systems. For tourism service platforms, this model can be used to build a more accurate, transparent, and exploratory personalised recommendation engine, thus enhancing user satisfaction and platform stickiness. For the management of scenic spots and government regulators, the resource prediction function based on this model can provide data support for decision-making such as passenger flow diversion, facility deployment, and emergency management, which can help realise the optimal allocation and efficient use of tourism resources and promote the sustainable development of the tourism industry.

Declarations

All authors declare that they have no conflicts of interest.

References

Askew, A.J. (1974) 'Chance-constrained dynamic programming and the optimization of water resource systems', *Water Resources Research*, Vol. 10, No. 7, p.159.

Attarzadeh, I. and Ow, S.H. (2010) 'A novel soft computing model to increase the accuracy of software development cost estimation', *IEEE Systems Journal*, Vol. 3, No. 24, p.689.

Brush, T.H. and Artz, K.W. (1999) 'Toward a contingent resource-based theory: the impact of information asymmetry on the value of capabilities in veterinary medicine', *Strategic Management Journal*, Vol. 20, No. 3, pp.223–250.

Bussolino, F., Camussi, G., Aglietta, M., Braquet, P. and Marchisio, P.C. (1987) 'Human endothelial cells are target for platelet-activating factor. I. platelet-activating factor induces changes in cytoskeleton structures', *Journal of Immunology*, Vol. 139, No. 7, pp.2439–2446.

Coppola, M. and Padricelli, G.M. (2024) 'Challenges in digital social research methods: algorithms, traces and footprints. A resume of the current debate', *Italian Sociological Review*, Vol. 14, No. 5, p.246.

Federmeier, K.D., Wlotko, E.W. and Meyer, A.M. (2008) 'What's 'right' in language comprehension: event-related potentials reveal right hemisphere language capabilities', *Language and Linguistics Compass*, Vol. 7, No. 2, p.816.

Ferragina, P., Giancarlo, R., Greco, V., Manzini, G. and Valiente, G. (2007) 'Compression-based classification of biological sequences and structures via the universal similarity metric: experimental assessment', *BMC Bioinformatics*, Vol. 8, No. 1, p.252.

Gacau, K.K., Kombe, Y., Ngayo, M.O. and Mwaniki, P. (2016) 'Depressive disorders among workers in the selected hotels in the capital city of Kenya: a cross sectional study on prevalence and correlates', *Italian Sociological Review*, Vol. 6, No. 8, p.318.

Herman, E. (2004) 'Research in progress. Part 2 – some preliminary insights into the information needs of the contemporary academic researcher', *Aslib Proceedings*, Vol. 56, No. 2, pp.118–131.

Herrmann, C., Blume, S., Kurle, D., Schmidt, C. and Thiede, S. (2015) 'The positive impact factory-transition from eco-efficiency to eco–effectiveness strategies in manufacturing', *Procedia CIRP*, Vol. 29, No. 7, pp.19–27.

Interian, A., Martinez, I., Rios, L.I., Krejci, J. and Guarnaccia, P.J. (2010) 'Adaptation of a motivational interviewing intervention to improve antidepressant adherence among Latinos', *Cultural Diversity & Ethnic Minority Psychology*, Vol. 16, No. 2, pp.215–225.

Joachims, T., Granka, L., Pan, B., Hembrooke, H., Radlinski, F. and Gay, G. (2007) 'Evaluating the accuracy of implicit feedback from clicks and query reformulations in web search', *ACM Transactions on Information Systems*, Vol. 25, No. 2, p.7.

Kennedy, C. and Theodoropoulos, G. (2005) 'Towards intelligent data-driven simulation for policy decision support in the social sciences', *School of Computer Science*, Vol. 12, No. 8, p.277.

Lindenbeck, C., Ulmer, H. and Schulz, M. (2007) 'Open source technology in Webgis application development: implementation of a borehole information system', *Zeitschrift Der Deutschen Gesellschaft Für Geowissenschaften (ZDGG)*, Vol. 15, No. 13, p.380.

Liu, Z., Chen, S., Chen, Y., Su, J., Zhong, J. and Dong, C. (2024) 'A knowledge-enhanced interest segment division attention network for click-through rate prediction', *Neural Computing and Applications*, Vol. 36, No. 34, pp.21817–21837.

Lu, J.F., Chi, M.J. and Chen, C.M. (2013) 'Advocacy of home telehealth care among consumers with chronic conditions', *Journal of Clinical Nursing*, Vol. 2, No. 7, p.379.

Lynne, E., Mark, H., David, R. and Low (2016) 'The role of social marketing, marine turtles and sustainable tourism in reducing plastic pollution', *Marine pollution Bulletin*, Vol. 9, No. 3, p.328.

McNee, S.M., Riedl, J. and Konstan, J.A. (2006) 'Making recommendations better: an analytic model for human-recommender interaction', *School of Computer Science*, Vol. 11, No. 6, p.227.

Mo, Y., Chen, J., Xie, X., Luo, C. and Yang, L.T. (1970) 'Cloud-based mobile multimedia recommendation system with user behavior information', *IEEE Systems Journal*, Vol. 8, No. 1, pp.184–193.

Pappas, C., Fatichi, S., Rimkus, S., Burlando, P., Bugmann, H. and Huber, M.O. (2013) 'The role of local scale heterogeneities in terrestrial ecosystem functioning', *Agu Fall Meeting*, Vol. 5, No. 17, p.259.

Pulkki-Brännström, A.M., Wolff, C., Brännström, N. and Skordis-Worrall, J. (2012) 'Cost and cost effectiveness of long-lasting insecticide-treated bed nets – a model-based analysis', *Cost Effectiveness and Resource Allocation*, Vol. 10, No. 1, p.5.

Ramli and Ainon (2013) 'Usage of, and satisfaction with, accounting information systems in the hotel industry: the case of Malaysia', *University of Hull*, Vol. 6, No. 9, p.591.

Steenkamp, E.T. and Wingfield, M.J. (2013) 'Global forest research, science education and community service positively impacted by a unique centre of excellence in tree health biotechnology', *Southern Forests: a Journal of Forest Science*, Vol. 3, No. 16, p.390.

Trotter, C.S., Winn, J.N. and Hewitt, J.N. (2000) 'A multipole-Taylor expansion for the potential of gravitational lens mg j0414+0534', *Astrophysical Journal*, Vol. 535, No. 2, pp.671–691.

Verhoef, P.C., Leeflang, P.S.H., Reiner, J., Natter, M. and Saunders, J. (2011) 'A cross-national investigation into the marketing department's influence within the firm: toward initial empirical generalizations', *Journal of International Marketing*, Vol. 19, No. 3, pp.59–86.

Victoria and Carty (2009) 'Capitalist measures within a socialist model: a commodity chains analysis of the emerging Cuban tourism industry', *Canadian Journal of Latin American & Caribbean Studies*, Vol. 17, No. 12, p.916.

Wenwen, L., Linna, L., Michael, G. and Luc, A. (2013) 'A geospatial cyberinfrastructure for urban economic analysis and spatial decision-making', *International Journal of Geo-Information*, Vol. 2, No. 2, pp.413–431.

Witt, K., Daniels, C., Reiff, J., Krack, P. and Deuschl, G. (2008) 'Neuropsychological and psychiatric changes after deep brain stimulation for Parkinson's disease: a randomised, multicentre study', *Lancet Neurology*, Vol. 7, No. 7, pp.605–614.

Yonghong, L. and Shuwen, Z. (2019) 'Path of big data driving the transformation and upgrading of traditional industry: based on big data value chain perspective', *Science and Technology Management Research*, Vol. 5, No. 11, p.492.