Multi-granularity Contrastive Learning For Tourism Recommendation

Xiaojun Wang

College of Management, Dalian University of Finance and Economics, Dalian, 116622, Liaoning, China

Corresponding author. E-mail: xiaojwang2024@163.com

Received: April 29, 2024; Accepted: June 05, 2024

Tourism recommendation is a valuable and captivating location-based offering that aids novice travelers in crafting highly personalized travel itineraries. However, existing approaches fall short in capturing the breadth of human preferences and transition patterns. In instances of limited travel data, these methods may even offer recommendations that stray from the genuine travel intentions of tourists. To this end, a multi-granularity contrastive learning within the self-supervised framework is devised for tourism recommendation (MCL-TR), consisting of contrastive POI learning and contrastive tourism learning. Through the joint optimization of dual contrastive learning, MCL-TR achieves a holistic approach to tourism recommendation. By considering both POIs and tourism factors simultaneously, the system can provide comprehensive recommendations that cater to the individual preferences and needs of users, thereby enhancing the overall recommendation quality and user satisfaction. Finally, experiments conducted on three datasets demonstrate that MCL-TR sets a new benchmark in tourism recommendation tasks.

Keywords: Tourism recommendation; multi-granularity contrastive learning, representation learning

http://dx.doi.org/10.6180/jase.202504_28(4).0020

1. Introduction

Tourism recommendation refers to the provision of personalized and accurate travel advice and recommendations for travelers using information technology and data analysis techniques [1–4]. With the rapid development of information technology and the increasing demand for personalized and customized services [5–8], travel recommendation has become an important component of the tourism industry. Traditional travel recommendations often rely on manual experience and simple rules, which cannot meet the diverse and personalized needs of users. Therefore, leveraging technologies such as artificial intelligence and big data analysis to deeply explore and analyze user behavior, preferences, and historical data can provide users with more personalized and accurate travel recommendations. The background of travel recommendation is the increasing integration of information technology and travel demand, providing travelers with a more convenient and thoughtful travel experience.

Current tourism recommendations are divided into two schools of thought: traditional methods and deep methods [9, 10]. The former typically relies on rules and heuristic algorithms, such as using a greedy algorithm to recommend popular attractions, or optimizing an objective function that takes into account cost, time, and user preferences through integer linear programming. The latter is capable of capturing more complex user preferences and behavioral patterns. Deep representation learning employs neural networks to learn low-dimensional representations of users and attractions, thereby revealing intricate nonlinear relationships. These deep learning methods have demonstrated superior performance on real-world datasets, bringing new perspectives and possibilities to the field of travel recommendation.

Despite significant advancements in tourism recommendation systems, several challenges remain. One major issue is the diversity of user needs, as the wide variety of travel plans requested by users makes it difficult for systems to fully grasp and meet all requirements, especially for uncommon or previously unrecorded queries. Additionally, the uncertainty in travel plans poses a problem, as limited data and the vast range of individual preferences make it challenging to accurately understand and predict users' true travel intentions. Furthermore, the complexity of access patterns adds another layer of difficulty; exploring higher-order transition patterns and long-term dependencies is complicated by data sparsity, hindering effective analysis and recommendation.

To address the challenges in tourism recommendation systems, a novel approach called Multi-Granularity Contrastive Learning within a self-supervised framework (MCL-TR) has been developed. MCL-TR enhances the interaction between user queries and points of interest (POIs) by implementing a contrastive POI learning paradigm that leverages a POI graph and simple interaction augmentation. Additionally, it introduces a contrastive tourism learning paradigm within the frameworks of a time-aware query extractor and a tourism extractor, which strengthens the connection between user queries and tourism information. By jointly optimizing these dual contrastive learning paradigms, MCL-TR provides a holistic approach to tourism recommendations. This method simultaneously considers both POIs and broader tourism factors, leading to comprehensive recommendations that better match individual user preferences and needs. Experiments on three datasets have shown that MCL-TR sets a new benchmark in tourism recommendation tasks, significantly improving overall recommendation quality and user satisfaction.

2. Related works

2.1. Deep-based Tourism Recommendation

In recent years, major internet companies' app stores release a vast number of software products daily, leading to a massive amount of heterogeneous data from multiple sources. The sources of this data are diverse, and its structure is complex, exhibiting high-dimensional nonlinear characteristics. These features cannot be accurately captured through conventional machine learning methods, leading to suboptimal performance in tasks such as classification and clustering. However, the emergence of Deep Learning (DL) has alleviated these challenges [11]. DL excels in nonlinear transformations, deep feature learning, and high availability. Increasingly, scholars are exploring the use of deep neural network models for feature learning tasks, aiding in better understanding latent feature vectors from users' and projects' historical behaviors [12].

For example, He et al. integrated the multilayer perceptron of DL with matrix factorization models, proposing a novel neural matrix factorization model [13]. This model inputs both user and item latent vectors into a multilayer perceptron, effectively capturing the nonlinear relationships between users and items. CNN has advantages in extracting local features and text analysis. Many scholars have combined CNN with recommendation algorithms. Yan et al. utilized CNN to learn higher-order user-location interaction features, fully considering the interaction between different feature dimensions, resulting in more accurate location recommendations [14]. Wang et al., on the basis of utilizing CNN to extract users' historical and course information, incorporated attention mechanisms to better capture the varying degrees of influence of different courses on users [15]. Graph Neural Networks (GNNs) have been proven effective in handling non-Euclidean spatial data, enabling effective exploration of potential associations between entities. In the field of recommendation, as users' ratings of items can form a graph structure, GNNs can be leveraged to learn on the interaction graph between users and items. Zhang et al. utilized multiple graph convolutional networks to learn hidden features of user and item nodes, masking some nodes during training [16]. By reconstructing the masked node information using a decoder, they further enhanced the effectiveness of recommendations. Wang et al. constructed a knowledge graph based on the interaction between users and items, selectively aggregating neighborhood information on the knowledge graph using graph convolutional networks. This method incorporates semantic and structural information covering various entities and relationships within the knowledge graph into the model, accurately representing users' latent intentions and increasing the diversity of recommended items [17]. Chen et al. built two bipartite interaction graphs based on the relationships between users and item labels and between item labels themselves [18]. By utilizing multi-layer GNNs, they could better capture collaborative information between nodes, enriching the representation of each node and ultimately completing the recommendation of item labels.

2.2. Session-based Tourism Recommendation

In certain application scenarios, it's not feasible to collect user identity information or historical records of actions like browsing, liking, and sharing. Many users browse websites without logging in, necessitating the modeling of users' current behavior. In session-based recommendation, a session sequence can be understood as multiple sequential item click behaviors occurring in a user's current

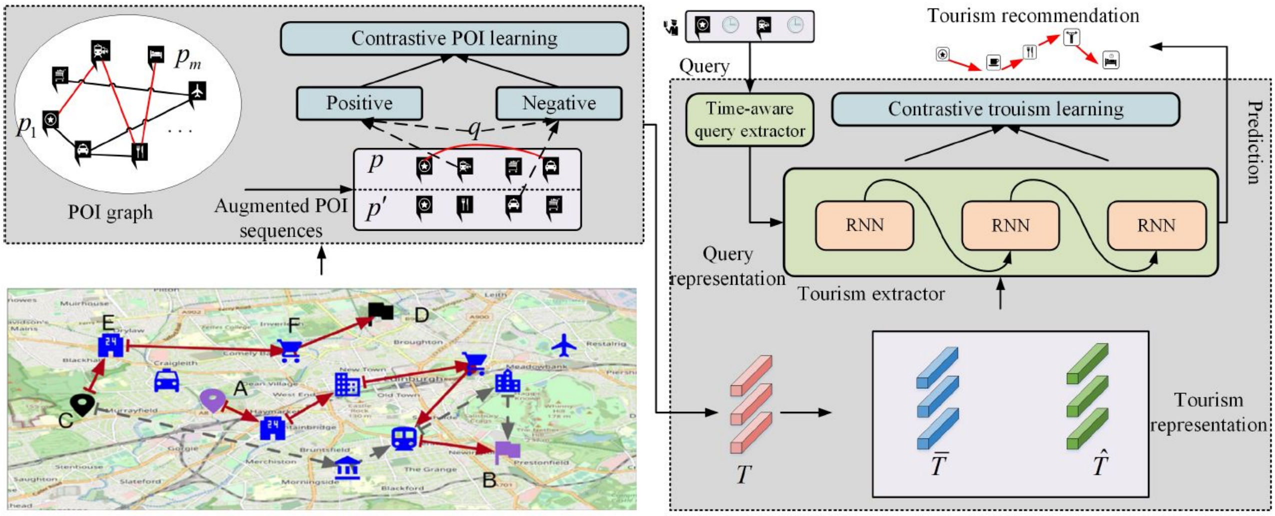

Fig. 1. The illustration of MCL-TR. It comprises self-supervised representation learning of point-of-interests and tourism. The former endeavors to map each POI into a reduced-dimensional space. This pivotal stage in trip recommendation not only tackles the issue of dimensionality but also fosters the exploration of semantic connections between queries and POIs. The later starts by formulating a context-sensitive query vector, followed by the random application of curated trip augmentation techniques to generate variations of the original trip. Subsequently, a query-based RNN is introduced as the foundational encoder for contrastive learning. Furthermore, this encoder functions as the trip generator during the formal training process.

session. The primary task of session-based recommendation is to predict the next item a user is most likely to click based on the current session sequence, and then recommend the top few items with the highest probability of being clicked by the user. Session-based recommendation algorithms largely address the limitations of traditional recommendation algorithms, which overly rely on interaction behavior between users and items, especially in scenarios where the number of items clicked by the user is relatively limited, making it challenging to provide precise project recommendation lists for users. The occurrence of anonymous user visits to websites is common, sparking significant interest among researchers who have designed various session-based recommendation [19, 20].

For example, Xu et al. utilized Graph Neural Networks (GNNs) and self-attention networks to construct a session-based recommendation model. This model captures local dependency relationships among items in the session graph using GNNs and captures long-term dependency relationships of sessions using self-attention networks [21]. Qiu et al. proposed a weighted graph attention network layer, which constructs user session sequences into a weighted directed graph. The weight values are calculated based on the frequency of appearance of the two items connected by the edge in the session [22]. Using the graph attention network layer, they obtain feature embeddings of each item in the session graph and use a Readout function to obtain the representation of the entire session graph based on the features of nodes. This model simultaneously considers specific sequence orders and potential orders in the session graph. Yu et al. introduced the concept of Target-aware GNN for session recommendation [23]. They define the predicted candidate item as the target item and utilize a target-aware attention network to capture the relevance of historical behaviors under specific target items. This approach automatically captures changes in user interests. They then combine the obtained user target intent embeddings with local intent embeddings and global intent embeddings to further improve the embedding representation of the session. Song et al. argued that the short-term and long-term preferences of friends can influence a user's interests. They constructed a user and friend network graph and modeled the user's dynamic preferences and social influences using a dynamic graph attention network. When a user's friend's interests change, the information of the user node is updated accordingly [24].

3. Multi-granularity contrastive learning for tourism recommendation

As shown in Fig. 1, a novel deep dual contrastive representation learning within the self-supervised framework is devised for tourism recommendation (MCL-TR), comprising self-supervised representation learning of point-of-interests and self-supervised representation learning of tourism.

3.1. Problem Definition

We first introduce the definitions in the recommendation: (1) Tourism: Considering P and T stand for sets of POIs and tourism, a triplet $p =< id, x, y >$ is used to denote a POI, where x and Y are longitude and latitude of POI. A tourism is comprised of s a POI sequence following time, i.e., $< p_1, t_1 > \rightarrow < p_2, t_2 > \rightarrow \cdots \rightarrow < p_m, t_m >$, in which $< p_*, t_* >$ represents p_* interacted at the t_* time. (2) Query: It is defined as a quintuple, i.e., $< p_s, t_s, p_e, t_e, m >$, where p_s, t_s, p_e, t_e, m denotes the start POI, the start time, the end POI, the end time, the visited POI number, respectively. Based on above definitions, the tourism recommendation can generate an expected journey via users, i.e., $T_e = \{p_i\}_{i=1}^{m}$, when users provide a historical tourism T_h and a query $q_e =< p_s, t_s, p_e, t_e, m >$, which is articulated in a formal description as follows:

$$T = \arg\max_{\theta} P(T_e|T_h, q) \tag{1}$$

where θ denotes the model parameter.

3.2. Self-supervised representation learning of point-of-interests

Based on the basic concept behind tourism recommendation aims to leverage a user-provided tourism query to generate a series of points of interest, a self-supervised representation learning of point-of-interests is devised to enhance interactions between the query and POIs. Specifically, MCL-TR aims to construct a POI graph to depict their interrelations and employs an interaction-enhanced strategy to integrate tourists' geographical preferences. To tackle the challenges posed by large-scale datasets and sparsity in query records, researchers have proposed a query-based sampling approach to generate more plausible travel routes. Additionally, a comparative POI learning method is introduced to specifically explore the potential interactions between user queries and points of interest, aiming to enhance the performance of the tourism recommendation.

POI graph structure: We conduct a POI graph, i.e., $G_o(P, E)$ based on user's historical tourism for modelling interactions between POIs and user query, in which E represents a collection of edges that depict the sequential tourism of users. Meanwhile, we attach each POI with a distance threshold ϖ for considering possible geographic inclinations of users. That is, if the distance between POIs is less than ϖ, then an edge is constructed between them. Then, we get a improvement POI graph $G_i(P, E_i)$ with an improvement adjacency matrix A_i whose elements depict the likelihood of interaction between source POI and target POI. Thus, the transition matrix A can be gained via computing the transition probability between the POIs based on the improvement adjacency matrix A_i:

$$P(p_j|p_i) = f_{ij}/f_i \tag{2}$$

where p_i and p_j deneote the i-th and j-th POIs in P. f_i stands for the frequency of p_i in E_i

Simple interaction augmentation: For query candidates $Q = \{< p_i, p_j > | i \neq j\}$, due to data sparsity, $|P| \ll |Q|$ always holds true. Thus, after obtaining G_i and A, a random walk mechanism is introduced to produce POI sequences for eliminating the influence of the shortage of queries. Specifically, considering a query $< p_i, p_j >$, we model r_k as the k-th node within a random walk mechanism in which $p_i = r_0$. Then, r_k is obtained via the distribution $P(r_k|r_{k-1})$ ($P(r_k|r_{k-1}) \in A$). When $p_j = r_k$ and $k \leq \vartheta$, we save the POI sequence $\{r_0, r_1, ..., r_k\}$. ϑ is a length budget that is used to restrict the generation of a long-distance sequence. Such a manner can generate POI sequences that better match real travel behaviors. Leveraging these sequences as training data further improves the performance of POI representation learning.

Contrastive representation learning: To capture correlations between query and POIs, the contrastive learning is used to learn representations of POIs. Specifically, we maximize the mutual information between the query and the POI belonging to the same sequence and minimize the mutual information between the query and the POI belonging to the different sequence. More specifically, for an any POI sequence $P^s = \{p_i^s\}_{i=1}^{m}$, let $z \in R^{m \times d}$ denote corresponding representations where d stands for representation dimension and q_s denote a query representation obtained via $q^s = \frac{z_1 + z_m}{2}$ where z_* stands for corresponding representations of the POI p_*^s. Then, the query representation q_s is viewed as anchor, forming a positive pair with POI $p_i \in p^s$ and $k-1$ negative pairs with POIs in other POI sequence which is related with other queries. Mathematically,

$$L_p = \mathrm{E}_{p_i^s \in P^s, q^s}[\exp(sim(z_i^s, q^s)) - \log \sum_{j=1}^{k-1} \exp(sim(z_j^{*\neq s}, q^s))] \tag{3}$$

where $sim(\cdot)$ denotes the similarity function, e.g., cosine similarity.

3.3. Self-supervised representation learning of tourism

Self-supervised representation learning of tourism consists of the time-aware query extractor, the tourism extractor, and the contrastive tourism learning.

Time-aware query extractor: Given a query $< p_s, t_s, p_e, t_e, m >$, we obtain representations z_s^p and z_e^p of the start POI p_s and the end POI p_e via the self-supervised representation learning of POI. To capture the time correlation hidden in the query, an hour-granularity encoding engine is utilized to extract representations z_s^t and z_e^t of the start time t_s and the end time t_e, respectively. Finally, a fusion function is designed to enhance information aggregations between representations instead of simple concatenations in previous works, Mathematically, the time-aware query representation is conducted via:

$$z_q = F_{fusion}([z_e^p||z_e^t] \sum([z_s^p||z_s^t]K_q) + [z_s^p||z_s^t||]z_e^p||z_e^t]W_q) \tag{4}$$

where $||$ stands for the concatenation. W_q stands for the network parameter. K_q stands for the third-order tensor.

Tourism extractor: We opt for a query-based recurrent neural network (RNN) as the tourism extractor for modeling tourism representations. This choice stems from the RNN's particular suitability in capturing and learning the temporal dependencies between various POIs along travel routes. The sequence in which users visit POIs is crucial for understanding their travel preferences. Additionally, this network architecture can recursively integrate user query information, thereby more accurately simulating and understanding user travel intentions.

Specifically, after obtaining the time-aware query representation z_q of the query q and the initial tourism representation $z_t = \{z_i\}_{i=1}^m$, we utilize the tourism extractor to mine correlations between POIs. Mathematically, the hidden state h_i of p_i in the tourism T is:

$$h_i = ext([z_i||z_q], h_{i-1}), i = 1, 2, ..., m \tag{5}$$

Contrastive tourism learning: In the self-supervised learning, to train the model to effectively learn tourism representations, it is necessary to utilize various aspects of real travel data as positive samples for training. This aids the model in capturing rich features of travel data, thereby enabling more accurate recommendations for users in the future. Specifically, given n queries $\{q_i\}_{i=1}^n$ with corresponding tourism $\{T_i\}_{i=1}^n$, we utilize traditional data augmentation paradigms to generate augmentation tourism $\{\bar{T}_i\}_{i=1}^n$ and $\{\hat{T}_i\}_{i=1}^n$. Then, we obtain corresponding query representations z_q and final hidden states $\bar{h}_i$ and $\hat{h}_i$ of $\bar{T}_i$ and $\hat{T}_i$ via the time-aware query extractor and tourism extractor, respectively. Thus, contrastive tourism learning loss is designed as:

$$L_t = -\sum_{i=1}^{n} \log \frac{\exp(sim(\bar{h}_i, \hat{h}_i))}{\exp(sim(\bar{h}_i, \hat{h}_i)) + \sum_{j=1, j\neq i}^{n} \exp(sim(\bar{h}_i, \hat{h}_j))} \tag{6}$$

3.4. The objective loss of MCL-TR

In MCL-TR, representation learning of point-of-interests and tourism are jointly learned via integrating into a unified self-supervised architecture, whose training objective loss is:

$$L = L_p + \lambda L_t \tag{7}$$

where λ is a parameter for balancing L_p and L_t. Multi-granularity contrastive learning in MCL-TR exist two advantages for tourism recommendations. (1) Enhanced User Preference Capture: Contrastive learning across various granularities facilitates a nuanced understanding of user preferences for different tourism elements. This refined capture aids in providing more tailored recommendations that closely match user interests. (2) Improved Recommendation System Robustness: Conducting contrastive learning at both POI and tourism levels strengthens feature representations, enabling the system to generate more reliable recommendations even with sparse or incomplete user data. This enhances the stability and dependability of recommendation outcomes.

3.5. Theoretical analysis of MCL-TR

In this section, the advantages of employing multi-granularity contrastive learning in tourism recommendations from a theoretical standpoint.

Contrastive POIs learning: In tourism recommendations, the representations of POIs are expected to depict interactions between the query and POIs to the greatest extent possible. In other words, the goal of POI representation is mutual information maximization between the query q and each POI P belonging to the same tourism sequnece:

$$I(q; p) = \mathrm{E}_{\mathcal{P}(p,q)} \log(\frac{\mathcal{P}(p|q)}{\mathcal{P}(p)}) \tag{8}$$

Then, the posterior distribute $\mathcal{R}$ is used to approximate true $\mathcal{P}$, having:

$$I(z_q; z_p) = \mathrm{E}_{\mathcal{R}(z_q,z_p)} \log(\frac{\mathcal{R}(z_p|z_q)}{\mathcal{R}(z_p)}) \tag{9}$$

where z_q and z_p are learned via deep neural networks. Due to the hardship in estimating $\mathcal{R}(\mathrm{z_p}|\mathrm{z_q})$ and $\mathcal{R}(\mathrm{z_p})$, it is suitable via data sampling to appropriate. That is, we choose POIs belonging to the same tourism sequence with z_p as positive samples and POIs belonging to the different tourism sequence with z_p, as negative samples.Next, we partition positive and negative samples within the binary classifier:

$$L_p = -\mathop{\mathbb{E}}_{p} \log[\frac{\frac{\mathcal{R}(z_p|z_q)}{\mathcal{R}(z_p)}}{\frac{\mathcal{R}(z_p|z_q)}{\mathcal{R}(z_p)} + \sum_{j=1}^{k-1} \frac{\mathcal{R}(z_j|z_q)}{\mathcal{R}(z_j)}}] \tag{10}$$

Based on a density ratio $g(z_q; z_p) \propto \frac{\mathcal{R}(z_p|z_q)}{\mathcal{R}(z_p)}$ that retains correlations between z_p and z_q, having

$$\begin{aligned} L_p &= -\mathop{\mathbb{E}}_{p} \log[\frac{g(z_q;z_p)}{g(z_q;z_p)+\sum_{j=1}^{k-1} g(z_j;z_q)}] \\ &= \mathop{\mathbb{E}}_{p} \log[1 + \frac{\sum_{j=1}^{k-1} g(z_j;z_q)}{g(z_q;z_p)}] \\ &\approx \mathop{\mathbb{E}}_{p} \log[1 + \frac{k-1}{g(z_q;z_p)}] \\ &\geq -I(z_q; z_p) + \log(k) \end{aligned} \tag{11}$$

we can observe that:

$$I(z_q; z_p) \geq -L_p + \log(k) \tag{12}$$

Thus, minimizing the loss L_p is equivalent to maximizing mutual information $I(z_q; z_p)$, thus demonstrating the rationality of contrastive learning of POI representations.

Contrastive tourism learning: Generally, tourism demands of users are diverse, meaning that even if the system provides a trip plan T, users might plan a trip that is similar to T based on their personal preferences. To enrich the diversity of T further, MCL-TR uses two similar trips $\bar{T}$ and $\hat{T}$ to maximize the mutual information between them within a deep neural network:

$$I(\bar{T}; \hat{T}) = \mathbb{E}_{\mathcal{P}(\bar{T};\hat{T})} \log(\frac{\mathcal{P}(\bar{T}|\hat{T})}{\mathcal{P}(\bar{T})}) \tag{13}$$

Similarity, we can obtain the loss of contrastive tourism learning:

$$L_t = -\mathop{\mathbb{E}}_{T} \log[\frac{\frac{\mathcal{R}(\bar{T}|\hat{T})}{\mathcal{R}(\bar{T})}}{\frac{\mathcal{R}(\bar{T}|\hat{T})}{\mathcal{R}(\bar{T})} + \sum_{j=1}^{k-1} \frac{\mathcal{R}(\bar{T}_j|\hat{T})}{\mathcal{R}(\bar{T}_j)}}] \tag{14}$$

Correspondingly, we have:

$$I(\bar{T}; \hat{T}) \geq -L_t + \log(n) \tag{15}$$

Contrastive tourism learning streamlines training and enriches itinerary diversity, capturing human travel intentions. It establishes comparative relationships, enabling nuanced feature acquisition for personalized recommendations. Additionally, it enhances generalization for high-quality suggestions, adapting to diverse user needs effectively.

Table 1. The three tourism datasets.

Tourism city	POIs	Trajectory	User	Query
Edinburgh	33944	5028	1454	812
Glasgow	11434	2227	601	812
Toronto	39419	6057	1395	870

4. Experiments

4.1. Setup

Dataset and metric: Three common tourism datasets, i.e., Toronto, Glasgow, and Edinburgh, are utilized to verify the performance of MCL-TR, whose statistical information are shown in Table 1. Following previous works, F_1 score and $pair - F_1$ score are used for performance comparison between methods. Their values closer to 1 indicate greater in the experiments.

Implementation Details: In the model training, the Adam optimizer is chosen to learn and optimize the parameter selection, with the ReLU activation function being employed. The initial learning rate and decay factor are set to 0.001 and 0.1, respectively, the batch size is 256, and the number of training epochs is 20 rounds. The L_2 penalty coefficient is 10^{-5}, and the dimensions for the point of interest (POI), POI category, and user embedding vectors are 100. The Dropout strategy is used to prevent overfitting in the training process of the MCL-TR. The parameter settings for the baseline methods are consistent with those described in the original papers.

4.2. Comparison with baselines

Comparison baselines: Eight baselines are compared to MCL-TR, including CATHI [1], C-ILP [2], TRED [3], TRPE [4], PTRD [7], CatDM [8], DeepTrip [9], and NASR+ [10].

Comparison results: Table 2 presents a comparison of MCL-TR to prior works on three datasets in terms of F_1 score and $pair - F_1$ score. Specifically, our proposed MCL-TR model achieves the highest scores across all evaluated metrics on the Edinburgh, Glasgow, and Toronto datasets, indicating its superior performance in trip recommendation tasks. The results demonstrate the effectiveness of our model's ability to capture the complex patterns and preferences inherent in trip planning, leading to more accurate and personalized trip recommendations. The reasons are two-fold: (1) Detailed User Preference Understanding: MCL-TR's multi-granularity contrastive learning enables a nuanced grasp of user preferences across various tourism aspects, leading to more precise recommendations aligned with individual expectations. (2) Improved System Robustness: Learning at both the point of interest (POI) and broader tourism levels equips the model with robust

Table 2. Comparison of MCL-TR to prior works on three datasets in terms of F_1 score and $pair - F_1$ score.

Method	Edinburgh		Glasgow		Toronto	
	F_1 score	$pair - F_1$ score	F_1 score	$pair - F_1$ score	F_1 score	$pair - F_1$ score
C-ILP	0.752	0.535	0.852	0.709	0.811	0.623
CATHI	0.760	0.731	0.710	0.659	0.820	0.782
TRED	0.745	0.608	0.813	0.680	0.827	0.684
PTRD	0.713	0.498	0.804	0.657	0.801	0.680
TRPE	0.704	0.479	0.811	0.653	0.812	0.667
CatDM	0.756	0.590	0.835	0.691	0.813	0.680
DeepTrip	0.765	0.660	0.831	0.782	0.808	0.748
NASR+	0.755	0.734	0.849	0.756	0.829	0.803
MCL-RT (ours)	**0.781**	**0.768**	**0.855**	**0.802**	**0.837**	**0.819**

Table 3. Ablation experiments of each component in MCL-TR.

Method	Edinburgh		Glasgow		Toronto	
	F_1 score	$pair - F_1$ score	F_1 score	$pair - F_1$ score	F_1 score	$pair - F_1$ score
MCL-TR_1	0.593	0.418	0.729	0.489	0.665	0.428
MCL-TR_2	0.613	0.599	0.755	0.658	0.707	0.771
MCL-TR_3	0.777	0.767	0.849	0.800	0.815	0.818
MCL-RT (ours)	**0.781**	**0.768**	**0.855**	**0.802**	**0.837**	**0.819**

feature representations, especially beneficial for sparse or incomplete user data scenarios, ensuring relevant recommendations through contrastive learning insights.

4.3. Ablation Study

This section presents ablation experiments of MCL-TR to prior works on three datasets in terms of F_1 score and $pair - F_1$ score. Specifically, MCL-RT_1 only uses self-supervised representation learning of point-of-interests to train the model for generating tourism recommendations. MCL-RT_2 only uses self-supervised representation learning of tourism to train the model for generating tourism recommendations. MCL-RT_3 uses multi-granularity contrastive learning without simple interaction augmentation to train the model for generating tourism recommendations.

As shown in Table 3, there are three observations: (1) The results indicate that as we move from MCL-TR_1 to MCL-TR_3 and finally to MCL-RT (ours), there is a consistent increase in performance across all datasets for both F_1 score and $pair - F_1$ score. This suggests that incorporating multi-granularity contrastive learning into the model leads to incremental improvements in capturing user preferences and generating effective tourism recommendations. (2) The comparison between MCL-TR_1 and MCL-TR_2 shows the individual contributions of self-supervised representation learning at the point-of-interest level and the tourism level, respectively. The higher scores achieved by MCL-TR_3, which combines both, underscore the importance of leveraging both granularities of information for enhancing the model's understanding and recommendation quality. (3) The ablation study reveals that the version of the model without simple interaction augmentation (MCL-TR_3) still outperforms the single-granularity models (MCL-TR_1 and MCL-TR_2), but the inclusion of interaction augmentation in MCL-RT (ours) leads to the best results. This suggests that interaction augmentation, possibly through the modeling of transitions between different types of tourism elements, plays a crucial role in improving the accuracy and reliability of the recommendations.

5. Conclusion

In this paper, we introduce MCL-TR, a multi-granularity contrastive learning framework for tourism recommendation. MCL-TR includes self-supervised representation learning for both point-of-interests and tourism, enriching interaction dynamics between user queries and POIs across hierarchical levels. This comprehensive understanding enhances recommendation accuracy and robustness, leading to personalized travel suggestions. Future directions include integrating advanced machine learning techniques like deep learning and reinforcement learning, as well as incorporating real-time contextual data for more dynamic recommendations.

References

[1] F. Zhou, X. Yue, G. Trajcevski, T. Zhong, and K. Zhang. "Context-aware variational trajectory encoding and human mobility inference". In: *The world wide web conference*. 2019, 3469–3475. DOI: 10.1145/3308558.3313608.

[2] J. He, J. Qi, and K. Ramamohanarao. "A joint context-aware embedding for trip recommendations". In: *International Conference on Data Engineering*. 2019, 292–303. DOI: 10.1109/ICDE.2019.00034.

[3] F. Zhou, H. Wu, G. Trajcevski, A. Khokhar, and K. Zhang, (2020) *"Semi-supervised trajectory understanding with poi attention for end-to-end trip recommendation"* **Transactions on Spatial Algorithms and Systems** ***6***(2): 1–25. DOI: 10.1145/3378890.

[4] N. L. Ho and K. H. Lim. "User preferential tour recommendation based on poi-embedding methods". In: *International Conference on Intelligent User Interfaces-Companion*. 2021, 46–48. DOI: 10.1145/3397482.3450717.

[5] P. Li, J. Gao, J. Zhang, S. Jin, and Z. Chen, (2022) *"Deep Reinforcement Clustering"* **Transactions on Multimedia**: DOI: 10.1109/TMM.2022.3233249.

[6] J. Gao, M. Liu, P. Li, J. Zhang, and Z. Chen, (2023) *"Deep Multiview Adaptive Clustering With Semantic Invariance"* **Transactions on Neural Networks and Learning Systems**: DOI: 10.1109/TNNLS.2023.3265699.

[7] Z. Duan, Y. Gao, J. Feng, X. Zhang, and J. Wang. "Personalized tourism route recommendation based on user's active interests". In: *International Conference on Mobile Data Management*. 2020, 729–734. DOI: 10.1109/MDM48529.2020.00071.

[8] F. Yu, L. Cui, W. Guo, X. Lu, Q. Li, and H. Lu. "A category-aware deep model for successive POI recommendation on sparse check-in data". In: *the web conference*. 2020, 1264–1274. DOI: 10.1145/3366423.3380202.

[9] Q. Gao, F. Zhou, K. Zhang, F. Zhang, and G. Trajcevski, (2021) *"Adversarial human trajectory learning for trip recommendation"* **Transactions on Neural Networks and Learning Systems** ***34***(4): 1764–1776. DOI: 10.1109/TNNLS.2021.3058102.

[10] J. Wang, N. Wu, and W. X. Zhao, (2021) *"Personalized route recommendation with neural network enhanced search algorithm"* **Transactions on Knowledge and Data Engineering** ***34***(12): 5910–5924. DOI: 10.1109/TKDE.2021.3068479.

[11] P. Li, Z. Chen, L. T. Yang, J. Gao, Q. Zhang, and M. J. Deen, (2018) *"An incremental deep convolutional computation model for feature learning on industrial big data"* **Transactions on Industrial Informatics** ***15***(3): 1341–1349. DOI: 10.1109/TII.2018.2871084.

[12] J. Latrech, Z. Kodia, and N. Ben Azzouna, (2024) *"CoDFi-DL: a hybrid recommender system combining enhanced collaborative and demographic filtering based on deep learning"* **Journal of Supercomputing** ***80***(1): 1160–1182. DOI: s11227-023-05519-2.

[13] X. He and T.-S. Chua. "Neural factorization machines for sparse predictive analytics". In: *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*. 2017, 355–364. DOI: 10.1145/3077136.3080777.

[14] C. Yan and Y. Shi. "A personalized location recommendation based on convolutional neural network". In: *Information Technology and Mechatronics Engineering Conference (ITOEC)*. 2020, 1516–1519. DOI: 10.1109/TSMC.2019.2931723.

[15] J. Wang, H. Xie, O. T. S. Au, D. Zou, and F. L. Wang. "Attention-based CNN for personalized course recommendations for MOOC learners". In: *International symposium on educational technology*. 2020, 180–184. DOI: 10.1109/ISET49818.2020.00047.

[16] J. Zhang, X. Shi, S. Zhao, and I. King. "STAR-GCN: Stacked and Reconstructed Graph Convolutional Networks for Recommender Systems". In: *International Joint Conference on Artificial Intelligence*. 2019, 4264. DOI: 1783.1/100998.

[17] H. Wang, M. Zhao, X. Xie, W. Li, and M. Guo. "Knowledge graph convolutional networks for recommender systems". In: *The world wide web conference*. 2019, 3307–3313. DOI: 10.1145/3308558.3313417.

[18] Y.-H. Chen, L. Huang, C.-D. Wang, and J.-H. Lai, (2021) *"Hybrid-order gated graph neural network for session-based recommendation"* **Transactions on Industrial Informatics** ***18***(3): 1458–1467. DOI: 10.1109/TII.2021.3091435.

[19] X. Xin, L. Yang, Z. Zhao, P. Ren, Z. Chen, J. Ma, and Z. Ren. "On the Effectiveness of Unlearning in Session-Based Recommendation". In: *International Conference on Web Search and Data Mining*. 2024, 855–863. DOI: 10.1145/3616855.3635823.

[20] Y. Wu, J. Cao, and G. Xu, (2023) *"FASTER: A dynamic fairness-assurance strategy for session-based recommender systems"* **Transactions on Information Systems** ***42***(1): 1–26. DOI: 10.1145/3586993.

[21] C. Xu, P. Zhao, Y. Liu, V. S. Sheng, J. Xu, F. Zhuang, J. Fang, and X. Zhou. "Graph contextualized self-attention network for session-based recommendation." In: *IJCAI*. ***19***. 2019, 3940–3946. DOI: 10.24963/ijcai.2019/547.

[22] R. Qiu, J. Li, Z. Huang, and H. Yin. "Rethinking the item order in session-based recommendation with graph neural networks". In: *Proceedings of the 28th ACM international conference on information and knowledge management*. 2019, 579–588. DOI: 10.1145/3357384.3358010.

[23] F. Yu, Y. Zhu, Q. Liu, S. Wu, L. Wang, and T. Tan. "TAGNN: Target attentive graph neural networks for session-based recommendation". In: *international ACM SIGIR conference on research and development in information retrieval*. 2020, 1921–1924. DOI: 10.1145/3397271.3401319.

[24] W. Song, Z. Xiao, Y. Wang, L. Charlin, M. Zhang, and J. Tang. "Session-based social recommendation via dynamic graph attention networks". In: *international conference on web search and data mining*. 2019, 555–563. DOI: 10.1145/3289600.3290989.