KDDC: Knowledge-Driven Disentangled Causal Metric Learning for Pre-Travel Out-of-Town Recommendation

Yinghui Liu[1] , **Guojiang Shen**[1] , **Chengyong Cui**[1] , **Zhenzhen Zhao**[1] , **Xiao Han**[2] , **Jiaxin Du**[1] , **Xiangyu Zhao**[2] and **Xiangjie Kong**[1*]
[1]Zhejiang University of Technology
[2]City University of Hong Kong
{LiuYingHui240, zhenzhenzhao_97}@outlook.com, gjshen1975@zjut.edu.cn, {hahayunc3c3y25, hahahenha, jiaxin.joyce.du}@gmail.com, xianzhao@cityu.edu.hk, xjkong@ieee.org

Abstract

Pre-travel recommendation is developed to provide a variety of out-of-town Point-of-Interests (POIs) for users planning to travel away from their hometowns but have not yet decided on their destination. Existing out-of-town recommender systems work on constructing users' latent preferences and inferring travel intentions from their check-in sequences. However, there are still two challenges that hamper the performance of these approaches: i) Users' interactive data (including hometown and out-of-town check-ins) tend to be rare, and while candidate POIs that come from different regions contain various semantic information; ii) The causes for user check-in include not only interest but also conformity, which are easily entangled and overlooked. To fill these gaps, we propose a **K**nowledge-**D**riven **D**isentangled **C**ausal metric learning framework (KDDC) that mitigates interaction data sparsity by enhancing POI semantic representation and considers the distributions of two causes (i.e., conformity and interest) for pre-travel recommendation. Specifically, we pretrain a constructed POI attribute knowledge graph through a segmented interaction method and POI semantic information is aggregated via relational heterogeneity. In addition, we devise a disentangled causal metric learning to model and infer user-related representations. Extensive experiments on two real-world nationwide datasets display the consistent superiority of our KDDC over state-of-the-art baselines.

1 Introduction

With the rapid growth of location-based social networks such as Foursquare and Yelp, Point-of-Interest (POI) recommendation [Wang *et al.*, 2023a] has become increasingly prevalent in our daily lives. Compared to conventional region-specific POI recommendation, out-of-town POI recommendation serves users who want to travel out of their hometowns. Most out-of-town POI recommender systems [Wang *et al.*, 2017; Ding *et al.*, 2019; Li and Gong, 2020; Sun *et al.*, 2021; Xin *et al.*, 2021] aim to solve issues such as data sparsity and interest drift when users have ascertained the travel locations. Recently, a new task (i.e. pre-travel out-of-town recommendation) [Xin *et al.*, 2022] has been proposed for providing POIs to the users who plan to out-of-town travel yet have not decided on the destination region. To deal with user interest drift, the pre-travel recommendation requires modeling users' behavior preferences and inferring their travel intentions to out-of-town POIs spanning diverse regions. However, existing out-of-town recommendation methods still face two significant challenges, which may fail to provide satisfying results for pre-travel recommendation:

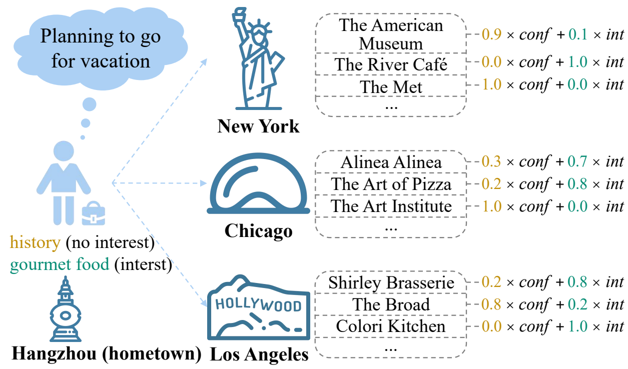


Figure 1: An illustrative example of a user's travel intention to different out-of-town POIs. *conf* and *int* indicate the conformity and interest, respectively.

i) **The sparsity of check-in data and diversity of POIs.** Due to the sparsity of interactive data (whether hometown check-ins or out-of-town check-ins) and the diversity of candidate POIs, relying solely on the sequence of visited POIs and their spatial relationships for modeling user preference is insufficient. The semantic information of a POI, such as its category, type, rating, and other details, similarly plays a prominent role in shaping user preference [Yang *et al.*, 2022; Zhang *et al.*, 2023]. For instance, users are likely to consult the evaluations of a POI before making a visitation decision. Thus, it is wise to learn semantic knowledge of POIs.

ii) **The causes of a check-in are easily bundled together.** In terms of cause, user interest is undoubtedly a significant

[*]Corresponding author.

driver for out-of-town checking-in, but users may also visit popular POIs owing to crowd behavior [Zheng *et al.*, 2021; Zhao *et al.*, 2023]. As shown in Figure 1, a user residing in Hangzhou, whose primary interest lies in gourmet food rather than history, also intends to visit the popular American Museum of Natural History. Therefore, observed out-of-town check-ins can be attributed to both interest and conformity. Despite [Xin *et al.*, 2022] utilizes other travelers' behavior memory to take conformity into account, the crowd memory is generally similar for all users, which does not reflect conformity's personalization. Figure 1 also exhibits distinct conformity and interest distributions in a user's each travel intention. This demonstrates the necessity of disentangling the *causes* (conformity and interest) from diverse POIs for the *effect* (hometown and out-of-town check-ins).

In light of the above limitations and challenges, we propose a **K**nowledge-**D**riven **D**isentangled **C**ausal metric learning framework (KDDC) for pre-travel out-of-town recommendation. Specifically, to learn the semantic features of POIs, we first construct an attribute knowledge graph encompassing all POIs and employ a segment-wise pre-training strategy to enhance information interaction between knowledge graph embedding. Then, given the relation heterogeneity in knowledge graph, we introduce a relation-aware knowledge aggregation mechanism to capture dependent contextual signals from both entities and relations. In this way, the users' hometown and out-of-town POI embeddings can be enhanced. By using designed disentangled causal metric learning that considers both interest and conformity, we improve the representation of users' behavioral preferences. Meanwhile, we develop a travel intent inference network (TIIN) to facilitate the inferring of travel intentions from the perspectives of both interest and conformity. Ultimately, we conduct recommendations within the detailed metric space, partitioned into various subspaces w.r.t. the users' hometowns using dynamic mapping.

The following are our summarized contributions:

- We propose a knowledge-driven approach, namely KDDC, to jointly consider the semantic knowledge of POIs and two different reasons why users check-in (i.e. conformity and interest) for pre-travel recommendation.
- To alleviate the sparsity in check-in data while exploring potential similarities among diverse POIs, we develop a segmented interaction pre-training method for enhancing knowledge semantic representation and utilize relation heterogeneity to aggregate knowledge for enriching POI embedding. Besides, we design a disentangled causal metric learning to improve users' behavior preferences modeling and travel intentions inference.
- To demonstrate the effectiveness of KDDC, we conduct diverse experiments on two real-world datasets showing that the proposed model consistently outperforms others.

2 Preliminary

Definition 1 (POI Attribute Knowledge Graph). *The POI attribute knowledge graph is defined as $\mathcal{G}_k = \{(v, r, e)\}$, which organizes external POI attributes with diverse types of entities and corresponding relationships. Specifically, each POI-relation-entity triplet (v, r, e) characterizes the semantic relatedness between the POI and entity v and e with the relation r, such as (The Great Wall, Located in, Beijing).*

Definition 2 (User Out-of-town Travel Behavior). *Given a user u, his/her out-of-town travel behavior is denoted by a five-tuple $\tau = (u, \vec{c}_h, \vec{c}_o, a_h, a_o)$, indicating that the user u travels from his/her hometown a_h to out-of-town region a_o and leaves check-in records in both regions, which are represented by $\vec{c}_h$ and $\vec{c}_o$, respectively.*

Formally, given a set of users $\mathcal{U} = \{u_i\}_1^{|\mathcal{U}|}$, a set of POIs $\mathcal{V} = \{v_i\}_1^{|\mathcal{V}|}$, a POI attribute knowledge graph $\mathcal{G}_k$, a set of regions $\mathcal{A} = \{a_i\}_1^{|\mathcal{A}|}$, and users' out-of-town travel behavior records $\mathcal{O} = \{\tau_i\}_1^{|\mathcal{O}|}$, our studied problem is stated as follow.

Problem Formulation. Given the historical out-of-town travel behavior records $\mathcal{O}$ left by $\mathcal{U}$ and the POI attribute knowledge graph $\mathcal{G}_k$, learn a recommender $\mathcal{F}$. Then, a new user $u^* \notin \mathcal{U}$ shows up with his/her hometown check-in $\vec{c}_h$ observed in a_h^*, $\mathcal{F}$ recommends a list of out-of-town POIs $\mathcal{V}^* \subset \mathcal{V}$ which not in a_h to u^*.

3 Methodology

We present the overall framework of KDDC in Figure 2. This framework comprises three main stages: *Knowledge Graph Segmented Pre-Training, POI Semantical Knowledge Aggregation, and Disentangled Causal Metric Learning for Pre-Travel Recommendation*:

1) We perform fine-grained partitioning of multiple embeddings in a constructed attribute knowledge graph for pre-training to achieve sufficient information interaction. **2)** To consolidate semantic knowledge across various relationships, a relation-aware knowledge aggregation layer is introduced to update POI embeddings. **3)** We utilize a disentangled causal representation manner to attribute users' behavioral preferences to interest and conformity and hence infer the travel intentions of users by TIIN. Moreover, metric learning is adopted to assist model training for recommendation. Technical details are introduced in the following subsections.

3.1 Knowledge Graph Segmented Pre-Training

Knowledge graph pre-training aims to enhance models' understanding and inference capabilities by improving entity and relation embeddings. Inspired by [Liu *et al.*, 2022], low-dimensional embeddings can be partitioned into several segments and the interactions among these segments foster the exchange of richer information. Thereby, we first performed a fine-grained embedding segmentation of POIs, relations and entities. The number of their segments is equal and the set of segmented embedding vectors is defined as:

$$\mathbf{m} = \left\{\mathbf{m}^{(0)}, \mathbf{m}^{(1)}, ..., \mathbf{m}^{(n-1)}\right\}, \mathbf{m}^{(i)} \in \mathbb{R}^{d/n}, \quad (1)$$

where n is the number of segments, $\mathbf{m} \in \{\mathbf{v}, \mathbf{r}, \mathbf{e}\}$ represents POI, relation and entity respectively. Then, in accordance with [Xu *et al.*, 2020], we introduce a scoring function $f_S(\cdot)$ to implement embedding chiasma interactions among distinct segments and evaluate the association strength among them:

$$f_S(\mathbf{v}, \mathbf{r}, \mathbf{e}) = \sum_{0 \leq x, y < n} s_{x,y} \left\langle \mathbf{r}^x, \mathbf{v}^y, \mathbf{e}^{z_{x,y}} \right\rangle, \quad (2)$$

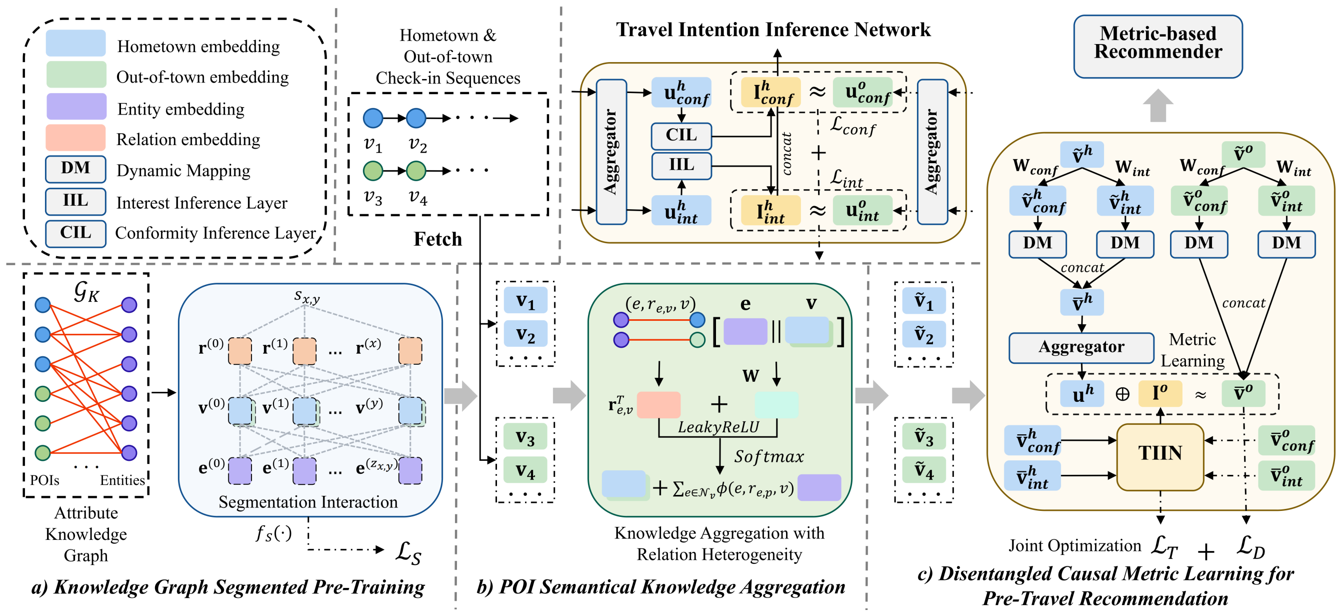


Figure 2: The framework of the proposed KDDC.

where $\langle\cdot\rangle$ means the dot product sum of vector elements. $\{x, y, z\}$ represent segmented ordinal numbers and

$$s_{x,y} = \begin{cases} -1, & \text{if } x + y \geq n \text{ and } x \text{ is odd} \\ 1, & \text{otherwise} \end{cases} \tag{3}$$

controls its sign to support both symmetric and antisymmetric relations [Xu *et al.*, 2020]. Simultaneously, we employ another variable $z_{x,y}$ to prevent over-embedding for computing overhead reduction:

$$z_{x,y} = \begin{cases} y, & \text{if } x \text{ is even} \\ (x + y)\%n, & \text{if } x \text{ is odd} \end{cases} \tag{4}$$

As a result, the even-numbered segments can capture the symmetry of the relationship and the odd-numbered segments capture antisymmetry information. Formally, the segmentation-based pre-training loss $\mathcal{L}_S$ is shown below:

$$\mathcal{L}_S = \sum_{(v,r,e,e')\in\mathcal{G}_K} -\ln\sigma(f_S(\mathbf{v}, \mathbf{r}, \mathbf{e}') - f_S(\mathbf{v}, \mathbf{r}, \mathbf{e})), \tag{5}$$

where e' is the negative sample generated by randomly replacing the entity e for the observed triplets (v, r, e).

3.2 POI Semantical Knowledge Aggregation

The semantics conveyed by different relationships are distinct. The embeddings of POIs that share similar semantics (e.g., POIs that belong to the same category) should be in close proximity to each other in the latent space. Given the success of graph attention mechanisms in [Wang *et al.*, 2019a; Xia *et al.*, 2021; Yang *et al.*, 2022], we utilize the relation heterogeneity between POI and entities over the knowledge graph to update POI embeddings as illustrated in Figure 2***b***. In order to avoid the manual design of path generation on knowledge graphs, entity- and relation-dependent context is projected into specific representations by parameterized attention matrix. To this end, we construct our relation-aware message aggregation mechanism between the POI and its connected entities in $\mathcal{G}_K$. This mechanism is used to generate knowledge-aware POI embeddings, relying on the heterogeneous attentive aggregator described as follows:

$$\begin{aligned} \tilde{\mathbf{v}} &= \mathbf{v} + \sum_{e\in\mathcal{N}_v} \phi(e, r_{e,v}, v)\mathbf{e}, \\ \phi(e, r_{e,v}, v) &= \frac{\exp(\sigma(\mathbf{r}_{e,v}^T \mathrm{W}[\mathbf{e} \parallel \mathbf{v}]))}{\sum_{e\in\mathcal{N}_v}\exp(\sigma(\mathbf{r}_{e,v}^T \mathrm{W}[\mathbf{e} \parallel \mathbf{v}]))}, \end{aligned} \tag{6}$$

where $\mathcal{N}_v$ is the neighboring entities of POI v under different types of relations $r_{e,v}$ in $\mathcal{G}_K$, $\mathbf{v} \in \mathbb{R}^d$ and $\mathbf{e} \in \mathbb{R}^d$ represent the embedding of POI and entity, respectively. The assessed entity- and relation-specific attentive relevance during the knowledge aggregation process is denoted as $\phi(e, r_{e,v}, v)$, which encodes the distinct semantics of relationships between POI v and entity e. $\parallel$ means the concatenation of two embeddings. $\mathrm{W} \in \mathbb{R}^{d\times 2d}$ denotes the weight matrix that is tailored to the input POI and entity representations. σ is the activation function *LeakyReLU* for non-linear transformation. Through the aforementioned mechanism, the underlying distances between embeddings of diverse POIs infused with similar semantic knowledge will be brought closer together.

3.3 Disentangled Causal Metric Learning for Pre-Travel Recommendation

User behavior preference modeling. With the aim of strengthening recommender systems, several causal-related methods such as inverse propensity scoring (IPS) [Gruson *et al.*, 2019], counterfactual inference [Wang *et al.*, 2021a] and causal embedding [Zheng *et al.*, 2021] have been proposed. In this work, we use causal embedding in the pre-travel recommendation because of the following advantages: 1) from a check-in generation perspective, causal embedding accurately models users' hometown and out-of-town preferences for different reasons; 2) causal modeling helps to improve model generalization capabilities. Following previous works [Zheng *et al.*, 2021; Zhao *et al.*, 2023], we focus on the two main causes: interest and conformity to model users' behavior preferences. One can see the process of disentangling each cause into causal embedding in Figure 2*c*. We first partition the users' hometown and out-of-town context embeddings $\tilde{\mathbf{v}}^h$ and $\tilde{\mathbf{v}}^o$ into corresponding interest and conformity embeddings with two distinct linear transformations:

$$\begin{aligned} \tilde{\mathbf{v}}^h_{conf} &= \mathrm{W}_{conf}\tilde{\mathbf{v}}^h + \mathrm{b}_{conf},\ \tilde{\mathbf{v}}^h_{int} = \mathrm{W}_{int}\tilde{\mathbf{v}}^h + \mathrm{b}_{int}, \\ \tilde{\mathbf{v}}^o_{conf} &= \mathrm{W}_{conf}\tilde{\mathbf{v}}^o + \mathrm{b}_{conf},\ \tilde{\mathbf{v}}^o_{int} = \mathrm{W}_{int}\tilde{\mathbf{v}}^o + \mathrm{b}_{int}, \end{aligned} \tag{7}$$

where $\mathrm{W}_{\{conf,int\}} \in \mathbb{R}^{d\times d}$ and $\mathrm{b}_{\{conf,int\}} \in \mathbb{R}^d$ are trainable parameters. In addition, we partition the embedding space by users' hometowns with a dynamic mapping mechanism for making the following metric more distinguishable [Xin *et al.*, 2022]. In particular, given a user u, his/her related check-in embeddings are encoded as

$$\begin{aligned} &\bar{\mathbf{v}}^h = \bar{\mathbf{v}}^h_{conf} \parallel \bar{\mathbf{v}}^h_{int}, \bar{\mathbf{v}}^o = \bar{\mathbf{v}}^o_{conf} \parallel \bar{\mathbf{v}}^o_{int}, \\ &\bar{\mathbf{v}}^h_{conf} = \mathrm{R}_{a_h}(\mathrm{W}_d\tilde{\mathbf{v}}^h_{conf} + \mathrm{b}_d)^T, \bar{\mathbf{v}}^h_{int} = \mathrm{R}_{a_h}(\mathrm{W}_d\tilde{\mathbf{v}}^h_{int} + \mathrm{b}_d)^T, \\ &\bar{\mathbf{v}}^o_{conf} = \mathrm{R}_{a_h}(\mathrm{W}_d\tilde{\mathbf{v}}^o_{conf} + \mathrm{b}_d)^T, \bar{\mathbf{v}}^o_{int} = \mathrm{R}_{a_h}(\mathrm{W}_d\tilde{\mathbf{v}}^o_{int} + \mathrm{b}_d)^T, \end{aligned} \tag{8}$$

where $\mathrm{W}_d \in \mathbb{R}^{d\times d}$ and $\mathrm{b}_d \in \mathbb{R}^d$ are trainable parameters. $\mathrm{R} \in \mathbb{R}^{d\times|\mathcal{A}|}$ is a learnable partition matrix and R_{a_h} represents the column vector in R w.r.t. region a_h. The user ultimate hometown and out-of-town check-in embeddings $\bar{\mathbf{v}}^h$ and $\bar{\mathbf{v}}^o$, considering both interest and conformity are separately generated through the concatenation operation. Since the user's hometown behavior is relatively long-term and could reveal the user's inherent preference, we model his/her behavior preference based on his/her hometown check-in embeddings:

$$\mathbf{u}^h = \mathrm{AGG}([\bar{\mathbf{v}}^h_i]^{|\vec{c}_h|}_{i=1}), \tag{9}$$

where $\mathrm{AGG}(\cdot)$ denotes an arbitrary aggregation operation, such as max pooling, average pooling or attention pooling.

User travel intention inference. As a user's hometown behavior is different from that outside the city, comprehending travel intention plays a crucial role in out-of-town recommendation. Toward that, we develop a travel intention inference network (TIIN) to infer users' intentions from different reasons. In particular, given a user's mapped check-in embeddings in Eq. (8), we first employ the aggregator to summarize his/her hometown and out-of-town preferences under two causes:

$$\begin{aligned} \mathbf{u}^h_{conf} &= \mathrm{AGG}([\bar{\mathbf{v}}^h_{conf,i}]^{|\vec{c}_h|}_{i=1}), \mathbf{u}^h_{int} = \mathrm{AGG}([\bar{\mathbf{v}}^h_{int,i}]^{|\vec{c}_h|}_{i=1}), \\ \mathbf{u}^o_{conf} &= \mathrm{AGG}([\bar{\mathbf{v}}^o_{conf,i}]^{|\vec{c}_o|}_{i=1}), \mathbf{u}^o_{int} = \mathrm{AGG}([\bar{\mathbf{v}}^o_{int,i}]^{|\vec{c}_o|}_{i=1}). \end{aligned} \tag{10}$$

Then, two nonlinear transformation layers: *conformity inference layer* and *interest inference layer* are adopted to infer the user's conformity and interest travel intentions:

$$\begin{aligned} \mathbf{I}^o_{conf} = \sigma(\mathrm{W}_C\mathbf{u}^h_{conf} + \mathrm{b}_C)&, \mathbf{I}^o_{int} = \sigma(\mathrm{W}_I\mathbf{u}^h_{int} + \mathrm{b}_I), \\ \mathbf{I}^o =& \mathbf{I}^o_{conf} \parallel \mathbf{I}^o_{int}, \end{aligned} \tag{11}$$

where $\mathrm{W}_C \in \mathbb{R}^{d\times d}$, $\mathrm{W}_I \in \mathbb{R}^{d\times d}$, $\mathrm{b}_C \in \mathbb{R}^d$ and $\mathrm{b}_I \in \mathbb{R}^d$ are trainable parameters. σ denotes the activation function *SiLU*.

Metric Learning. Inspired by the great performance of metric learning in recommendation [Hsieh *et al.*, 2017; Zhou *et al.*, 2019; Xin *et al.*, 2022], we apply it in disentangled causal intention inference and implement a metric-based out-of-town recommender. Aiming at the inference task, we first devise two supervised popularity-based causal metric losses to bridge the user's inferred embedding and factual out-of-town embedding from two causes as follows,

$$\begin{aligned} \mathcal{L}_{conf} &= \sum_{u\in\mathcal{U}} \exp(-P_{pop}) \left\|\mathbf{I}^o_{conf} - \mathbf{u}^o_{conf}\right\|^2_2, \\ \mathcal{L}_{int} &= \sum_{u\in\mathcal{U}} (1 - \exp(-P_{pop})) \left\|\mathbf{I}^o_{int} - \mathbf{u}^o_{int}\right\|^2_2, \\ \mathcal{L}_T &= \mathcal{L}_{conf} + \mathcal{L}_{int}, \end{aligned} \tag{12}$$

where P_{pop} represents the normalized popularity of the interacted POIs, which is defined as the ratio of the number of the POIs the user interacted with vs. the max POI interacted number [Zhao *et al.*, 2023]. For the highly popular POIs, the popularity weight $\exp(-P_{pop})$ is close to $\exp(-1)$ rather than 0, which allows the task to learn interest preferences to a certain extent. Oppositely, $1 - \exp(-P_{pop})$ ensures that interactions with more popular POIs are more likely to be attributed to conformity. The total loss of the task is expressed as $\mathcal{L}_T$.

Then we integrate the user's inferred intention $\mathbf{I}^o$ into his/her preference and employ the triangle inequality in the embedding space to calculate the distance between such fortified preference and a POI v_i as follows,

$$d(u, v_i) = \left\|\left[\mathbf{u}^h \oplus \mathbf{I}^o\right] - \bar{\mathbf{v}}_i\right\|^2_2. \tag{13}$$

Given a positive out-of-town POI $v_i \in \vec{c}_o$ visited by u, we sample a negative POI $v_j \notin \vec{c}_o$. From pairwise comparisons, the main disentangled causal metric loss under the Bayesian personalized ranking (BPR) criterion for pre-travel out-of-town recommendation is formally defined as,

$$\mathcal{L}_D = \sum_{u\in\mathcal{U}}\sum_{i\in\vec{c}_o}\sum_{j\notin\vec{c}_o} \left[\delta + d(u, v_i) - d(u, v_j)\right]_+, \tag{14}$$

where δ is the margin size and $[\cdot, 0]_+ = max(\cdot, 0)$ represents the standard hinge loss.

Joint Optimization and Recommendation. According to Eqs. (12) and (14), the integrative optimization loss of our KDDC is:

$$\mathcal{L} = \lambda_1\mathcal{L}_D + \lambda_2\mathcal{L}_T, \tag{15}$$

where λ_1, λ_2 are the hyper-parameters to balance the effects of the two losses in KDDC.

Here, we formally give the process of the pre-travel out-of-town recommendation. Given a new coming user u^*, we first gain his/her hometown behavior preference embedding $\mathbf{u}^h_*$ and inferred travel intention embedding $\mathbf{I}^o_*$ through Eqs. (7), (8), (9) and (11). Then we select top-k out-of-town POIs whose mapped embeddings $\bar{\mathbf{v}}^o_i$ are closest to the concatenation of $\mathbf{u}^h_*$ and $\mathbf{I}^o_*$ as the recommendation list for u^*, where we formulate the distance as below:

$$d(u^*, v^o_i) = \left\|\left[\mathbf{u}^h_* \oplus \mathbf{I}^o_*\right] - \bar{\mathbf{v}}^o_i\right\|^2_2. \tag{16}$$

Statistics	Dataset	
	Foursquare	Yelp
# Users	3,889	34,222
# Regions	21	469
# POIs	24,389	83,155
# Check-ins	167,580	597,888
# Hometown check-ins	147,357	495,049
# Out-of-town check-ins	20,223	102,839
	Knowledge Graph	
# Relations	2	8
# Entities	410	137,312
# Triples	48,778	961,295

Table 1: Statistics of two datasets.

4 Experiments

4.1 Experimental Settings

Datasets. We chose two nationwide travel behavior datasets, Foursquare[1] and Yelp[2], to evaluate our framework. The statistical information of our experimented datasets with different check-in records and knowledge graph characteristics is given in Table 1. For check-in records, following [Xin *et al.*, 2022], we picked out the users who left check-ins in both hometown and out-of-town regions from these two datasets and the check-in sequence of each user was reformed as the homologous out-of-town travel record $\tau = (u, \vec{c}_h, \vec{c}_o, a_h, a_o)$ (Definition 2). Besides, for the quality of two datasets, the POIs visited less than 2 times were filtered out and a user was dropped if his/her travel behavior $\tau = (u, \vec{c}_h, \vec{c}_o, a_h, a_o)$ fails to fulfill the following conditions: 1) $\vec{c}_h \geq 4$; 2) $\vec{c}_o \geq 2$; 3) the frequency of $(a_h, a_o) \geq 10$. The two datasets are randomly partitioned based on users into three sets for training, validation, and testing following the proportions: 80%, 10% and 10%. To ensure the fairness of the assessments, all of the users on the two datasets are anonymization. For the knowledge graphs, we generate entity-dependent relations with various types of entities (e.g., category, star, and location).

Baseline. We compare KDDC with: 1) Three general recommendation methods, BPR [Rendle *et al.*, 2012], TEMN [Zhou *et al.*, 2019] and STAN [Luo *et al.*, 2021]; 2) A knowledge graph-enhance method KGCL [Yang *et al.*, 2022]; 3) Two disentangled recommendation methods, DCCL [Zhao *et al.*, 2023] and DisenPOI [Qin *et al.*, 2023]; 4) Two out-of-town recommendation methods, TRAINOR [Xin *et al.*, 2021] and CAPTOR [Xin *et al.*, 2022]. Note that since there is no explicit user-item interaction graph in the task of pre-travel recommendation, for fairness, we model implicit user preference through the same sequence aggregation operation as Eq. (9) for some methods like DCCL and KGCL.

Evaluation Metrics. Three widely used metrics for recommendation are adopted to appraise method performances, i.e., Hit Rate, Precision, and NDCG at a cutoff k, abbreviated as HR@K, Pre@K, and NDCG@K respectively. Following previous works [Xin *et al.*, 2021; Xin *et al.*, 2022], a fixed number of negative POIs, randomly sampled from the candidate set, are taken together with the positive out-of-town POIs as final candidates for each user. In the experiments, the number of negative POIs is set to 100.

[1]https://sites.google.com/site/yangdingqi/home/foursquare-dataset

[2]https://www.yelp.com.tw/dataset

Implementation details. The baselines and their parameter settings were implemented based on the original papers and official public code. In addition, the parameters of the KDDC were sampled as follows: the number of dimensions of all latent representations was set to 128. In the *Knowledge Graph Segmented Pre-Training*, n was 8 for Foursquare and 4 was for Yelp. In the optimization stage, λ_1 and λ_2 were set as 1, the optimizer was chosen as Adam with an initial learning rate of 0.001 and an L2 regularization with a weight of 10^{-5}. We implemented our KDDC and experimented with Pytorch. Note that the evaluation was conducted three times with varying initial parameters, and the reported results represent the average. The code for the implementation of KDDC is available for reproducibility[3].

4.2 Performance Comparison

Table 2 summarizes the performance of KDDC and baselines on the two datasets. We carry out the following observations: 1) Compared with all baselines, KDDC achieves optimal performance in terms of all metrics across two datasets. This illustrates that our method considering POI semantic knowledge and analyzing two main causes of user interaction is effective and generalizable for pre-travel recommendation. In particular, the improvements in contrast to the corresponding runner-up baselines, 13.53% $\sim$ 19.11% in HR@K, 11.33% $\sim$ 16.57% in Pre@K and 15.36% $\sim$ 16.18% in NDCG@K on Foursquare, and 5.56% $\sim$ 12.88% in HR@K, 6.71% $\sim$ 10.41% in Pre@K and 3.91% $\sim$ 9.18% in NDCG@K on Yelp for K = 5, 10, and 15; 2) Among these baselines, the two methods including DCCL and DisenPOI achieve relatively satisfactory results in most cases, which suggests the dominance of disentangling on recommendation. Meanwhile, KGCL and CAPTOR also produce sub-optimal performances in some cases, which proves the necessity of paying attention to semantic knowledge and spatial correlation. The spatial correlation will be discussed in Subsection 4.5.

4.3 Ablation Study

To evaluate the effectiveness of each component of KDDC on two datasets w.r.t all metrics, we compare it with its variants, including three large modules: 1) -PT: KDDC removes *Knowledge Graph Segmented Pre-Training*; 2) -KA: KDDC without *POI Semantical Knowledge Aggregation*; 3) -DC: KDDC discards *Disentangled Causal Metric Learning*, and two small modules: 4) -DM: KDDC removes the dynamic mapping mechanism and performs recommendation in a universal metric space; 5) -II: KDDC removes the travel intention inference network (TIIN). The histogram in Figure 3 illustrates the comparative results. Evidently, KDDC demonstrates superior performance compared to -KA, -DC, and -DM across all metrics on both datasets, which proves the critical roles of the aggregation of semantic knowledge and the application of metric learning within separated mapping spaces in disentangled causal embedding training. Besides,

[3]https://github.com/Yinghui-Liu/KDDC

Dataset	Metric	BPR	TEMN	STAN	KGCL	DCCL	DisenPOI	TRAINOR	CAPTOR	KDDC	Improv.
Foursquare	HR@5	0.0806	0.1701	0.2830	0.2983	0.3129	0.2971	0.2605	0.2944	**0.3727**	+19.11%
	Pre@5	0.0752	0.1371	0.2418	0.2652	0.2728	0.2566	0.2050	0.2522	**0.3180**	+16.57%
	NDCG@5	0.0733	0.1488	0.2591	0.3100	0.3151	0.2736	0.2264	0.2604	**0.3635**	+15.36%
	HR@10	0.1228	0.2475	0.4134	0.4218	0.4246	0.4338	0.3860	0.4364	**0.5118**	+17.28%
	Pre@10	0.0598	0.1042	0.1853	0.1920	0.1920	0.1959	0.1714	0.1940	**0.2261**	+15.42%
	NDCG@10	0.0665	0.1301	0.2265	0.2433	0.2441	0.2256	0.2003	0.2287	**0.2836**	+16.18%
	HR@15	0.1782	0.3244	0.5067	0.4981	0.5040	0.5255	0.5141	0.5268	**0.5981**	+13.53%
	Pre@15	0.0585	0.0903	0.1530	0.1537	0.1515	0.1606	0.1554	0.1606	**0.1788**	+11.33%
	NDCG@15	0.0635	0.1150	0.1956	0.2061	0.2051	0.1952	0.1829	0.1980	**0.2387**	+15.82%
Yelp	HR@5	0.0432	0.2100	0.4007	0.4452	0.4056	0.4533	0.3921	0.4552	**0.4805**	+5.56%
	Pre@5	0.0194	0.1233	0.2178	0.2381	0.2198	0.2377	0.2121	0.2390	**0.2550**	+6.71%
	NDCG@5	0.0145	0.1380	0.2455	0.2728	0.2489	0.2636	0.2424	0.2630	**0.2835**	+3.91%
	HR@10	0.1580	0.3046	0.4959	0.5455	0.4937	0.5632	0.4777	0.5704	**0.6288**	+10.23%
	Pre@10	0.0359	0.0902	0.1394	0.1521	0.1393	0.1569	0.1335	0.1591	**0.1726**	+8.49%
	NDCG@10	0.0278	0.1099	0.1811	0.2006	0.1825	0.1990	0.1771	0.1984	**0.2161**	+7.73%
	HR@15	0.2034	0.3599	0.5396	0.5988	0.5370	0.6258	0.5158	0.6154	**0.7064**	+12.88%
	Pre@15	0.0310	0.0712	0.1026	0.1130	0.1026	0.1188	0.0974	0.1170	**0.1312**	+10.41%
	NDCG@15	0.0265	0.0926	0.1470	0.1634	0.1480	0.1629	0.1430	0.1622	**0.1784**	+9.18%

Table 2: Comparison of methods on Foursquare and Yelp datasets. Bold: Best, underline: Second best.

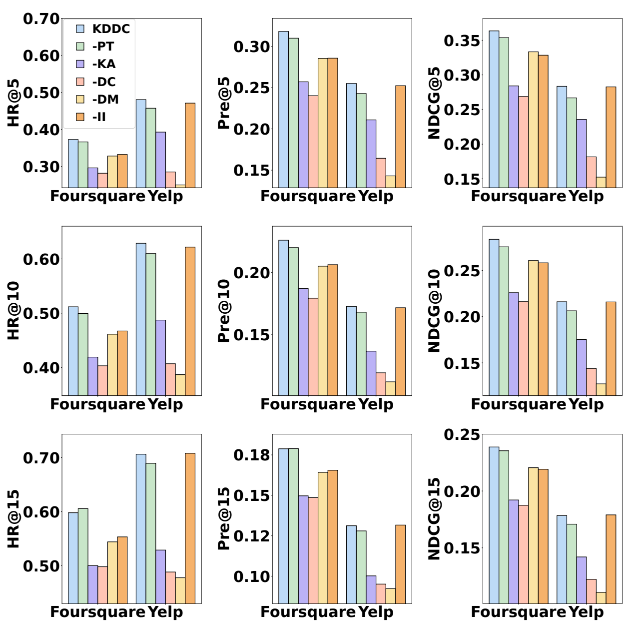
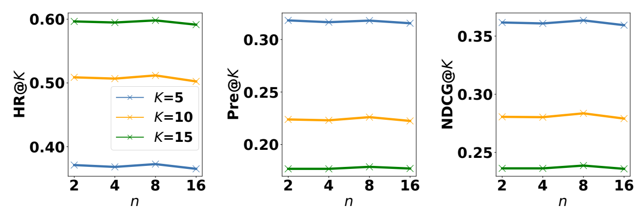
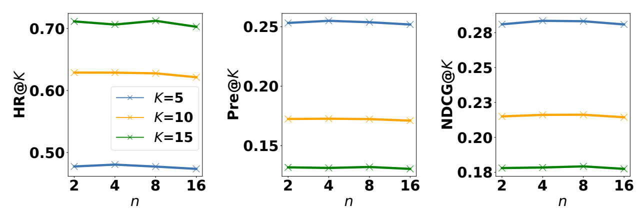


Figure 3: Contributions of various components in KDDC.

KDDC consistently outperforms -PT on Yelp but has a minor decrease on Foursquare w.r.t HR@15 and Pre@15, while KDDC persistently beats -II on Foursquare and achieves similar results on Yelp w.r.t all metrics. This is probably related to the size of the knowledge graph of the two datasets. When the scale is relatively small, pre-training has negligible impact on long sequence recommendation. However, as the scale increases, modeling users' out-of-town preferences can be achieved solely based on the enhanced user hometown preferences derived from a substantial amount of knowledge. But still, the segmented pre-training and travel intention inference are beneficial to the recommendation when there is no sufficient data.



(a) Effect on Foursquare



(b) Effect on Yelp

Figure 4: Effects of the number of embedding segments.

Dataset	Method	HR@10	Pre@10	NDCG@10	HR@15	Pre@15	NDCG@15
Foursquare	TransE	0.4955	0.2183	0.2750	0.5931	0.1768	0.2343
	TransR	0.4902	0.2163	0.2729	**0.5986**	0.1774	0.2340
	Ours(8)	**0.5118**	**0.2261**	**0.2836**	0.5981	**0.1788**	**0.2387**
Yelp	TransE	0.6162	0.1700	0.2142	0.6949	0.1295	0.1770
	TransR	0.6246	0.1724	0.2139	0.7062	0.1309	0.1766
	Ours(4)	**0.6288**	**0.1726**	**0.2161**	**0.7064**	**0.1312**	**0.1784**

Table 3: Comparison of knowledge graph embedding methods.

4.4 Parameter Sensitivity

We examine the sensitivity results across all metrics concerning the number of embedding segments, denoted as n in Eq. (1). The search for an optimal value is conducted within the range of $\{2, 4, 8, 16\}$. As displayed in Figure 4, the perfor-

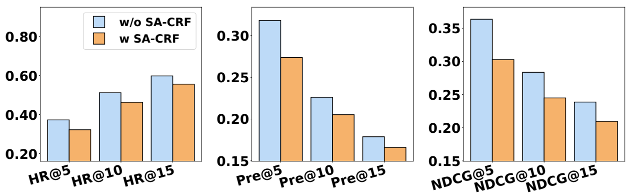


Figure 5: The evaluations of w and w/o SA-CRF on Foursquare.

mance is the best when n equals 8 on Foursquare and when n is 4 or 8 on Yelp, but the result in the case of n = 16 is not as good as n = 2. It may be due to the overfitting of the model with excess interaction information. Besides, to justify the validity of our embedding method separately, we compare it with other classical methods including TransE and TransR, and report the result under K = 10 and 15 in Table 3. One can see its performance improvement.

4.5 Discussion on Spatial Correlation

In most recent POI recommender systems, the spatial correlation among POIs is undoubtedly a research hotspot. For instance, CAPTOR [Xin *et al.*, 2022] captured the spatial affinity by a structured spatial affined conditional random field (SA-CRF) and DisenPOI specially designed a geographical graph to explore spatial dependence. For that purpose, we also try to consider the spatial correlation of POIs. Specifically, we adopt the SA-CRF same as [Xin *et al.*, 2022] in our framework, and it is directly inserted after Figure 2b and before Figure 2c. The comparison with the original framework w.r.t all metrics is reported in Figure 5, which presents a certain degree of performance degradation. We argue that incorporating spatial relation learning directly after knowledge aggregation will lead to contamination of the learned semantic features. There may be a demand to disentangle the two types of learning and we will consider this in our future work.

5 Related Work

Out-of-Town Recommendation. Out-of-town recommendation attempts to provide a list of new POIs for out-of-town users to visit. It is more intractable than the general POI recommendation due to problems like cold start and interest drift. [Ference *et al.*, 2013] was the first work to solve the problem with social influence. Next, some works based on Latent Dirichlet Allocation (LDA) [Wang *et al.*, 2017; Yin *et al.*, 2016] have been proposed to model the interest drift by user preferences, spatiotemporal effects and POI contents. In addition, there have been deep-learning methods for addressing this problem. [Li and Gong, 2020] learned the user preference from both hometown and out-of-town cities with transfer learning. [Xin *et al.*, 2021] used the neural topic model to analyze the user's out-of-town travel intentions. [Xin *et al.*, 2022] settled the more intractable pre-travel recommendation by capturing the crowd behavior memory. However, the above methods still face the challenge of data scarcity, especially in the pre-travel recommendation scenarios. In this paper, we are committed to alleviating this problem through the proposed knowledge-driven method.

Knowledge Graph-enhanced Recommendation. Existing Knowledge Graph (KG)-enhanced methods in the general recommendation task can be broadly grouped into three categories: embedding-based, path-based and graph neural network (GNN)-based. Embedding-based methods [Zhang *et al.*, 2016; Wang *et al.*, 2018] adopted the transition-based entity embedding schemes (e.g., TransE, TransR) to generate prior item embeddings. Path-based methods [Wang *et al.*, 2019b; Xia *et al.*, 2021] aim to build the user-item meta path to improve capturing high-order KG connectivity but they highly depend on the design of meta-paths needing domain knowledge and human efforts. Recent research directions have focused on the usage of GNNs to recursively perform information propagation between multi-hop nodes and incorporate long-range relational structures, such as KGIN [Wang *et al.*, 2021b] and KHGT [Xia *et al.*, 2021]. KGCL [Yang *et al.*, 2022] employed a joint self-supervised learning paradigm to alleviate the data noise issue but will cause extra time consumption. Accordingly, we combine a devised knowledge graph embedding segmented pre-training with a lightweight knowledge graph relational aggregation mechanism for pre-travel recommendation.

Disentangled Representation Learning for Recommendation. Disentangled representation learning aims to learn independent explanatory representations from disparate underlying factors behind data. For the temporal correlation, CLSR [Zheng *et al.*, 2022] tried to differentiate users' long-term and short-term interests from their clicking sequences. There also have been some GNN-based models to learn disentangled representation from various graph structures, including IDCL [Wang *et al.*, 2023b] (user-item graph), DisenHAN [Wang *et al.*, 2020] (heterogeneous graph) and DisenPOI [Zhao *et al.*, 2023] (multitype graph). Furthermore, for causality, DICE [Zheng *et al.*, 2021] and DCCL [Zhao *et al.*, 2023] disentangled conformity and interest influences by learning causal embedding. Analogously, we focus on such two causes to construct users' behavior preferences and infer their travel intentions with metric learning for pre-travel recommendation.

6 Conclusion

In this paper, our proposed KDDC framework performs pre-travel out-of-town recommendation by taking into account the semantic information of POI and the two causes of user interactions: conformity and interest. Three phases including *Knowledge Graph Segmented Pre-Training*, *POI Semantical Knowledge Aggregation*, and *Disentangled Causal Metric Learning for Pre-Travel Recommendation* make up our framework. The first phase uses a segmented pre-training strategy to improve the representations of knowledge graph elements. Then, the second phase aggregates semantics from different relationships to further ameliorate POI representation. After that, the third phase applies metric learning in disentangled causal representations for the more accurate construction of a user's preference and intention. Extensive experiments on two real-world datasets have demonstrated the superiority of KDDC over state-of-the-art methods for pre-travel recommendation.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62073295 and Grant 62072409.

References

[Ding *et al.*, 2019] Jingtao Ding, Guanghui Yu, Yong Li, Depeng Jin, and Hui Gao. Learning from hometown and current city: Cross-city poi recommendation via interest drift and transfer learning. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(4):1–28, 2019.

[Ference *et al.*, 2013] Gregory Ference, Mao Ye, and Wang-Chien Lee. Location recommendation for out-of-town users in location-based social networks. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, pages 721–726, 2013.

[Gruson *et al.*, 2019] Alois Gruson, Praveen Chandar, Christophe Charbuillet, James McInerney, Samantha Hansen, Damien Tardieu, and Ben Carterette. Offline evaluation to make decisions about playlistrecommendation algorithms. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, pages 420–428, 2019.

[Hsieh *et al.*, 2017] Cheng-Kang Hsieh, Longqi Yang, Yin Cui, Tsung-Yi Lin, Serge Belongie, and Deborah Estrin. Collaborative metric learning. In *Proceedings of the 26th International Conference on World Wide Web*, pages 193–201, 2017.

[Li and Gong, 2020] Dichao Li and Zhiguo Gong. A deep neural network for crossing-city poi recommendations. *IEEE Transactions on Knowledge and Data Engineering*, 34(8):3536–3548, 2020.

[Liu *et al.*, 2022] Jia Liu, Wei Huang, Tianrui Li, Shenggong Ji, and Junbo Zhang. Cross-domain knowledge graph chiasmal embedding for multi-domain item-item recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):4621–4633, 2022.

[Luo *et al.*, 2021] Yingtao Luo, Qiang Liu, and Zhaocheng Liu. Stan: Spatio-temporal attention network for next location recommendation. In *Proceedings of the web conference 2021*, pages 2177–2185, 2021.

[Qin *et al.*, 2023] Yifang Qin, Yifan Wang, Fang Sun, Wei Ju, Xuyang Hou, Zhe Wang, Jia Cheng, Jun Lei, and Ming Zhang. Disenpoi: Disentangling sequential and geographical influence for point-of-interest recommendation. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, pages 508–516, 2023.

[Rendle *et al.*, 2012] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. Bpr: Bayesian personalized ranking from implicit feedback. *arXiv preprint arXiv:1205.2618*, 2012.

[Sun *et al.*, 2021] Huimin Sun, Jiajie Xu, Rui Zhou, Wei Chen, Lei Zhao, and Chengfei Liu. Hope: a hybrid deep neural model for out-of-town next poi recommendation. *World Wide Web*, 24(5):1749–1768, 2021.

[Wang *et al.*, 2017] Hao Wang, Yanmei Fu, Qinyong Wang, Hongzhi Yin, Changying Du, and Hui Xiong. A location-sentiment-aware recommender system for both home-town and out-of-town users. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1143, 2017.

[Wang *et al.*, 2018] Hongwei Wang, Fuzheng Zhang, Xing Xie, and Minyi Guo. Dkn: Deep knowledge-aware network for news recommendation. In *Proceedings of the 2018 World Wide Web Conference*, pages 1835–1844, 2018.

[Wang *et al.*, 2019a] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. Kgat: Knowledge graph attention network for recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 950–958, 2019.

[Wang *et al.*, 2019b] Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, and Tat-Seng Chua. Explainable reasoning over knowledge graphs for recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 5329–5336, 2019.

[Wang *et al.*, 2020] Yifan Wang, Suyao Tang, Yuntong Lei, Weiping Song, Sheng Wang, and Ming Zhang. Disenhan: Disentangled heterogeneous graph attention network for recommendation. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 1605–1614, 2020.

[Wang *et al.*, 2021a] Wenjie Wang, Fuli Feng, Xiangnan He, Hanwang Zhang, and Tat-Seng Chua. Clicks can be cheating: Counterfactual recommendation for mitigating clickbait issue. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1288–1297, 2021.

[Wang *et al.*, 2021b] Xiang Wang, Tinglin Huang, Dingxian Wang, Yancheng Yuan, Zhenguang Liu, Xiangnan He, and Tat-Seng Chua. Learning intents behind interactions with knowledge graph for recommendation. In *Proceedings of the Web Conference 2021*, pages 878–887, 2021.

[Wang *et al.*, 2023a] Xinfeng Wang, Fumiyo Fukumoto, Jin Cui, Yoshimi Suzuki, Jiyi Li, and Dongjin Yu. Eedn: Enhanced encoder-decoder network with local and global context learning for poi recommendation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 383–392, 2023.

[Wang *et al.*, 2023b] Yuling Wang, Xiao Wang, Xiangzhou Huang, Yanhua Yu, Haoyang Li, Mengdi Zhang, Zirui Guo, and Wei Wu. Intent-aware recommendation via disentangled graph contrastive learning. In *Proceedings of the 32th international joint conference on artificial intelligence*, pages 2343–2351, 2023.

[Xia *et al.*, 2021] Lianghao Xia, Chao Huang, Yong Xu, Peng Dai, Xiyue Zhang, Hongsheng Yang, Jian Pei, and

Liefeng Bo. Knowledge-enhanced hierarchical graph transformer network for multi-behavior recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4486–4493, 2021.

[Xin *et al.*, 2021] Haoran Xin, Xinjiang Lu, Tong Xu, Hao Liu, Jingjing Gu, Dejing Dou, and Hui Xiong. Out-of-town recommendation with travel intention modeling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4529–4536, 2021.

[Xin *et al.*, 2022] Haoran Xin, Xinjiang Lu, Nengjun Zhu, Tong Xu, Dejing Dou, and Hui Xiong. Captor: A crowd-aware pre-travel recommender system for out-of-town users. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1174–1184, 2022.

[Xu *et al.*, 2020] Wentao Xu, Shun Zheng, Liang He, Bin Shao, Jian Yin, and Tie-Yan Liu. Seek: Segmented embedding of knowledge graphs. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3888–3897, 2020.

[Yang *et al.*, 2022] Yuhao Yang, Chao Huang, Lianghao Xia, and Chenliang Li. Knowledge graph contrastive learning for recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1434–1443, 2022.

[Yin *et al.*, 2016] Hongzhi Yin, Xiaofang Zhou, Bin Cui, Hao Wang, Kai Zheng, and Quoc Viet Hung Nguyen. Adapting to user interest drift for poi recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 28(10):2566–2581, 2016.

[Zhang *et al.*, 2016] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. Collaborative knowledge base embedding for recommender systems. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 353–362, 2016.

[Zhang *et al.*, 2023] Dabin Zhang, Ronghui Xu, Weiming Huang, Kai Zhao, and Meng Chen. Towards an integrated view of semantic annotation for pois with spatial and textual information. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 2441–2449, 2023.

[Zhao *et al.*, 2023] Weiqi Zhao, Dian Tang, Xin Chen, Dawei Lv, Daoli Ou, Biao Li, Peng Jiang, and Kun Gai. Disentangled causal embedding with contrastive learning for recommender system. In *Companion Proceedings of the ACM Web Conference 2023*, pages 406–410, 2023.

[Zheng *et al.*, 2021] Yu Zheng, Chen Gao, Xiang Li, Xiangnan He, Yong Li, and Depeng Jin. Disentangling user interest and conformity for recommendation with causal embedding. In *Proceedings of the Web Conference 2021*, pages 2980–2991, 2021.

[Zheng *et al.*, 2022] Yu Zheng, Chen Gao, Jianxin Chang, Yanan Niu, Yang Song, Depeng Jin, and Yong Li. Disentangling long and short-term interests for recommendation. In *Proceedings of the ACM Web Conference 2022*, pages 2256–2267, 2022.

[Zhou *et al.*, 2019] Xiao Zhou, Cecilia Mascolo, and Zhongxiang Zhao. Topic-enhanced memory networks for personalised point-of-interest recommendation. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 3018–3028, 2019.