Article

DTCRSKG: A Deep Travel Conversational Recommender System Incorporating Knowledge Graph

Hui Fang [1,2,3], **Chongcheng Chen** [1,2,*], **Yunfei Long** [4], **Ge Xu** [3] **and Yongqiang Xiao** [3]

1   Key Laboratory of Spatial Data Mining and Information Sharing of MOE, Fuzhou University,
    Fuzhou 350108, China; fanghui_ce@mju.edu.cn
2   Academy of Digital China (Fujian), Fuzhou University, Fuzhou 350108, China
3   Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, College of Computer
    and Control Engineering, Minjiang University, Fuzhou 350108, China; xuge@pku.edu.cn (G.X.);
    xyqhappy12@sina.com (Y.X.)
4   School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, UK;
    yl20051@essex.ac.uk
*   Correspondence: chencc@fzu.edu.cn; Tel.: +86-0591-63179500

Abstract: In the era of information explosion, it is difficult for people to obtain their desired information effectively. In tourism, a travel recommender system based on big travel data has been developing rapidly over the last decade. However, most work focuses on click logs, visit history, or ratings, and dynamic prediction is absent. As a result, there are significant gaps in both dataset and recommender models. To address these gaps, in the first step of this study, we constructed two human-annotated datasets for the travel conversational recommender system. We provided two linked data sets, namely, interaction sequence and dialogue data sets. The usage of the former data set was done to fully explore the static preference characteristics of users based on it, while the latter identified the dynamics changes in user preference from it. Then, we proposed and evaluated BERT-based baseline models for the travel conversational recommender system and compared them with several representative non-conversational and conversational recommender system models. Extensive experiments demonstrated the effectiveness and robustness of our approach regarding conversational recommendation tasks. Our work can extend the scope of the travel conversational recommender system and our annotated data can also facilitate related research.

1. Introduction

With the information explosion, it is difficult for a user to find travel information that is in line with their interests and suitable for their travel plan to enjoy high-quality travel, which is driving an urgent need for a personalized travel recommender system to provide more ingenious travel suggestions and contribute to the success of the service provider. Traditional travel recommendation methods are mainly divided into two categories: collaborative filtering (CF) recommendations [1] and content-based recommendations [2]. The CF methods mostly use travel interaction records and the data sparsity limits their performance. On the other hand, the content-based recommendation methods can alleviate the data sparsity problem by using richer auxiliary information, such as textual descriptions, content tags, and social and geographical information. Recently, more attention has been paid to deep neural networks, such as DeepMF [3], NCF [4], WideDeep [5], and DeepFM [6].

However, existing travel recommender systems are mostly based on static recommendation models, which primarily predict a user's preference toward a travel service

by analyzing past behaviors offline, e.g., click logs, visit history, and ratings on services. Static recommendation models rely heavily on the quality of historical offline data, which may be noisy and sparse. An assumption for static models is that historical interactions entirely represent a user's preference. However, in many practical scenarios, a user's current preference cannot be reliably predicted from past interactions, which are significantly lacking or even blank since travel is a low-frequency activity. Furthermore, a user's preference would dynamically be impacted by contextual factors. From the standpoint of communication science, the dissemination of information is often accompanied by the flow of preferred attitudes or feelings. Finally, the user may not know their preference until the decision-making phase, e.g., when knowing the available options. The emergence of the conversational recommender system provides new insight to address many of these challenges. Unlike static models, conversational recommender system models can elicit the current and detailed preferences of the user, respond to the feedback by users on the suggestion, and provide explanations for the recommended item. Unfortunately, there has been little work done on travel conversational systems.

Our research aimed to propose a deep conversational recommender system incorporating travel knowledge graph (TKG) that can complete a dynamic context-based travel recommendation task. Dialogue is introduced as a supplement to an interaction sequence, alleviates data sparsity, captures current user preferences, or even finds user preferences. Our deep model represents user preferences by encoding historical conversations and historical interaction sequences. In addition, we incorporated knowledge to make the conversation recommendation process more fluid and fit the travel scenario with spatio-temporal constraints. Meanwhile, we built a travel conversational recommendation dataset in Chinese to facilitate our study since the rich semantic content of Chinese dialogues can provide many clues for feature extraction in a deep approach. Overall, we provide resources and baseline models for using NLP technology to travel recommendations. To our knowledge, our work is the first to produce a deep travel conversational recommendation system.

In summary, our work has the following contributions:

We constructed two human-annotated Chinese datasets for the travel conversational recommender system. We provided two linked data sets, namely, interaction sequence and dialogue data sets. The usage of the former data set was to fully explore the static preference characteristics of users based on it. At the same time, the latter identified the dynamic changes in user preference from it. Both datasets were collected in Chinese since the rich semantic content of Chinese dialogues can provide many clues for feature extraction in a deep approach. The dataset will be released to the public for research purposes. We fill the current research gap by incorporating travel knowledge graph into a deep conversational recommender system. In this area, our research is one of the first models to model user preferences by encoding historical conversations and historical interaction sequences. We conducted a detailed comparison and analysis between our model and the current SOTA models in terms of various challenges the travel recommender system faces. The discoveries from the study can promote the development of travel recommendations.

2. Related Work

2.1. Conversational Recommender System

A conversational recommender system (CRS) is defined as a system that can predict users' dynamic preferences through dialogue context to complete recommendation tasks. As expressed in [7,8], interactive recommender and task-oriented dialogue systems are regarded as simplified and homogeneous CRSs, respectively. Existing CRSs roughly fall into three main categories: system-driven [9–11], mixed-driven [12–15], and user-driven [16–18]. In a system-driven CRS, the system mainly asks questions or offers options about user preferences to recommend. This requires the user to appropriately point out the appropriate query candidates and be familiar with each item they want. Mixed-driven CRS allows both the system and user to lead the conversation by asking questions or via chit-chat. The system constantly interacts with the user in a multi-turn conversation while discussing

the different topics (e.g., greeting or philosophy) to lead to the final recommendation. The natural language-based response requires the generated language to be proper, correct (even fluent), and meaningful, involving helpful information about the recommended target. Unlike the previous two, a user-driven CRS focuses on scenarios that require query understanding where the user has explicit claims or query objectives.

2.2. Personalized Travel Recommender System

A personalized travel recommender system needs to capture individual travel preferences accurately. However, the data involved in personal travel preferences are sparse and vulnerable to change in the current context. Recently, in addition to traditional machine learning methods [19–21], personalized travel recommender systems have gradually moved to a deep learning approach [22–25]. However, these deep learning approaches are mostly static models that extract highly abstracted features from historical interaction data. Therefore, these approaches are still unable to solve the problems associated with data scarcity, cold starts, and dynamic preferences in travel recommendations.

Along with the breakthrough and rapid development of NLP technology, there is a trend in tourism that moves away from the earlier question and answer systems [26] to conversational systems [27,28]. CRS has the potential to become the new desired travel recommendation framework due to its ability to dynamically elicit or discover current user preferences, as was demonstrated early on by [29]. Unfortunately, they did not have the full advantage of advanced deep learning methods for travel conversational recommendations then.

2.3. Knowledge-Graph-Based Recommender System

In recent years, knowledge-graph-based recommender systems have attracted considerable interest since they can alleviate data sparsity, mitigate cold starts, and provide a better understanding of recommendations with the knowledge graph as side information. Existing methods can be roughly divided into path-based methods [30–32], embedding-based methods [33–37], and hybrid methods [38,39]. Path-based methods utilize the connectivity similarity defined in heterogeneous information networks to enrich user or item representation. Embedding-based methods obtain more precise entity representations by leveraging the information in a graph structure via KG embedding. Finally, hybrid methods integrate the semantic representation of entities and relations and the connectivity information learned by the graph neural network framework [38]. However, KG-based recommender systems are unable to infer real-time interests due to their static property. Therefore, integrating knowledge into the conversational recommender system is a straightforward solution.

3. The Proposed Approach

In this section, we first formulate the personalized travel conversational recommendation task. Then we introduce our solution to this task.

3.1. Problem Formulation

Given a user u we assume that they have a historical interaction sequence

$$P_u = \{poi_n\}_{n=1}^{m}, \tag{1}$$

which is a chronologically-ordered sequence of points of interests (POIs) that u has interacted with. Each POI may be a hotel, a restaurant, or an attraction. Each conversation consists of a list of utterances, which is denoted by

$$Uts = \{ut_n\}_{n=1}^{k}, \tag{2}$$

where ut_n is the utterance at the nth turn.

Based on these basic concepts and notations, the task of a personalized travel conversational recommendation is defined as: given a user u, user historical interaction sequence P_u, historical utterances $\{ut_1, \ldots, ut_{k-1}\}$, and associated entities from TKG, the target of the task is to predict the poi_k that satisfies user u given the conversation context and their past interaction records.

3.2. Model Architecture

We proposed a Deep Travel Conversational Recommender System, abbreviated as DTCRSKG, to complete the aforementioned task. The architecture of the proposed model is illustrated in Figure 1. Inspired by TGCRS [15], we used two BERT-based modules to encode historical utterances and historical interaction sequences, respectively. In addition, we integrated knowledge into BERT to make our model deeply understand the underlying semantic information about user preferences contained in interaction sequences and conversations in the travel domain. In the following sections, we introduce the details of our model.
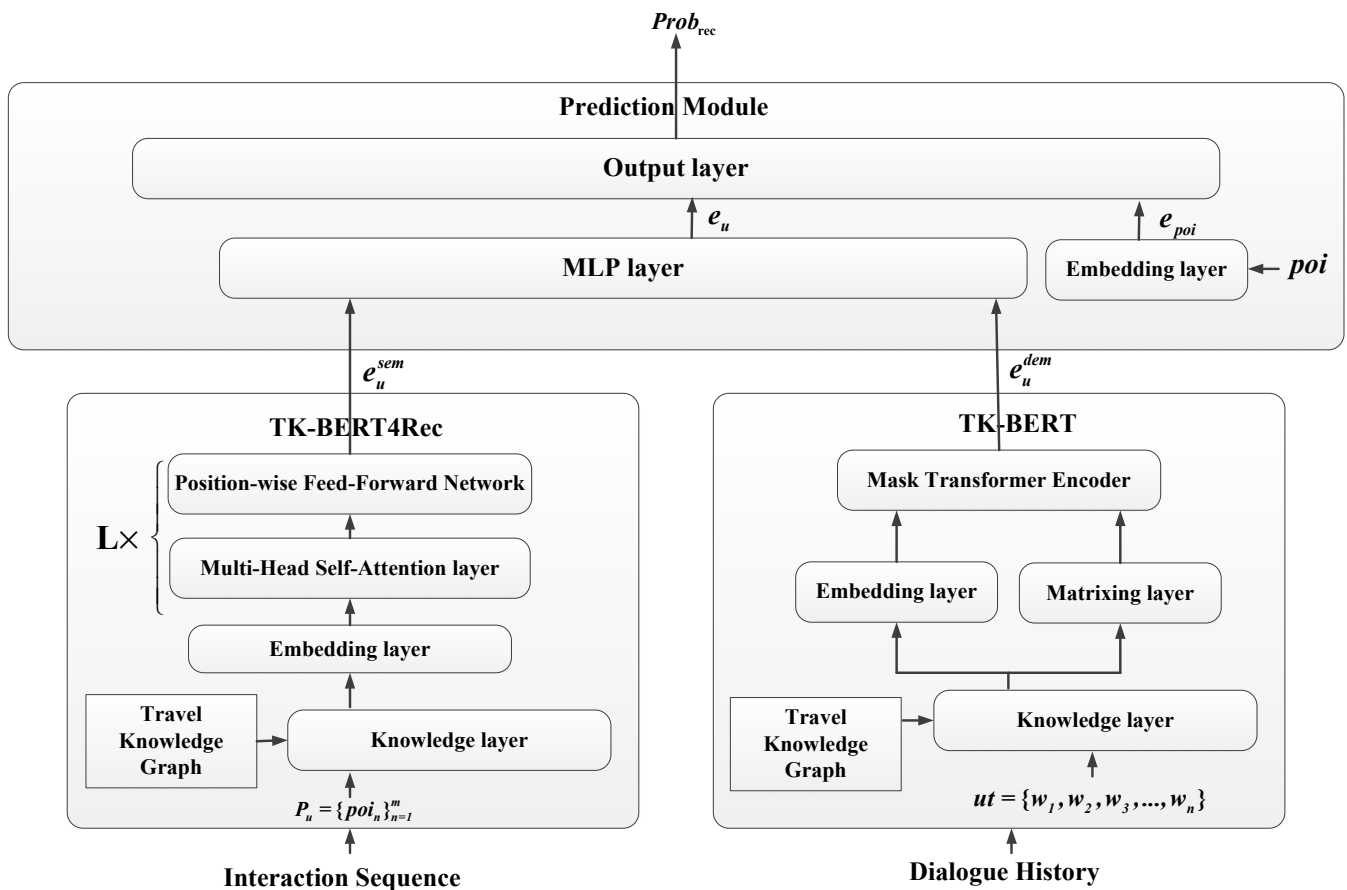

Figure 1. The architecture of the proposed DTCRSKG model. It contains two major parts, i.e., the TK-BERT4Rec part that learns user interaction representation and the TK-BERT part that learns dialogue representation.

3.2.1. Dialog Encoding

We utilized a travel-knowledge-infused BERT (TK-BERT) to encode historical utterances in the dialog encoding module to capture more information about user preferences in the interactive dialogue. TK-BERT was derived from K-BERT [40] and pre-trains BERT with TKG. As shown in Figure 1, TK-BERT consists of a knowledge layer, an embedding layer, a matrixing layer, a mask transformer encoder, and a TKG. First, unlike K-BERT, only key entities in the input utterance sentence $ut = \{w_1, w_2, w_3, \ldots, w_n\}$ are selected to query their corresponding triples from the travel knowledge graph in the knowledge layer. Here,

key entities are the ones corresponding to keywords in the sentence. A knowledge query can be formulated as:

$$T = KQuery(ut, TKG), \qquad (3)$$

where KQuery is an abbreviation for the knowledge query operation and $T = \{(w_i, r_{i0}, w_{i0}), \ldots, (w_i, r_{ij}, w_{ij})\}$ is a collection of queried triples. Then, these knowledge triples are injected into the input sentence by placing them in their corresponding positions and a sentence tree is generated. The sentence tree can have multiple branches and its depth is set to 1. The structure of a sentence tree is illustrated in Figure 2.
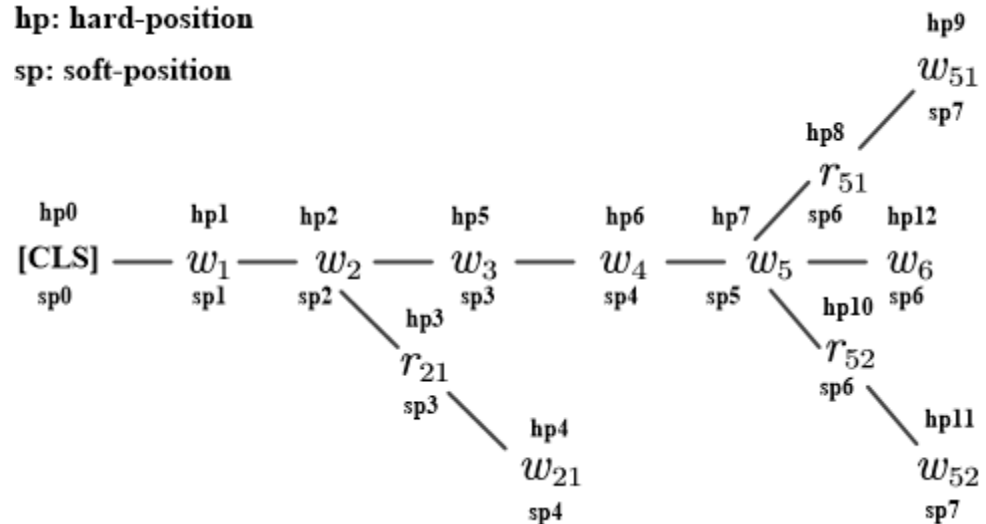

Figure 2. Structure of a sentence tree. The tree starts with [CLS] and is extended in the order of the words. Each branch is a triple that is queried.

Next, the embedding layer was designed to convert the sentence tree into an embedding representation, which is the sum of token embedding, position embedding, and segment embedding. Among them, the token embedding is obtained via a trainable lookup table. Furthermore, to maintain the sentence's structural information (i.e., the order of tokens), hard-positioning is replaced by soft-positioning [40] in the position embedding. Taking the sentence tree in Figure 2 as an example, r_{21} and w_{21} are inserted between w_2 and w_3, and r_{51} and w_{51} with r_{52} and w_{52} are inserted between w_5 and w_6. If the original hard-position sequential encoding scheme of BERT is followed, the input sentence is changed to $\{w_1, w_2, r_{21}, w_{21}, w_3, w_4, w_5, r_{51}, w_{51}, r_{52}, w_{52}, w_6\}$, resulting in a wrong semantic situation where w_{51} is the subject of r_{52}. With the soft-position sequential encoding scheme, the position ordinal number of r_{52} is replaced by six instead of ten and the position ordinal number of w_3 is replaced by three instead of five. In this way, the original semantic structure of the sentence information can be maintained. Finally, similar to BERT, segmentation embedding is used to identify different sentences when multiple sentences are included.

After using soft positioning, the position numbers of both r_{51} and r_{52} are six, which makes them close in the calculation of self-attentiveness, but in reality, they may be unrelated. w_{21} is only related to w_2 and not to w_{51} or w_{52}. Therefore, the representation of w_{21} should not be influenced by w_{51} or w_{52}. On the other hand, the [CLS] labels used for classification should not bypass w_2 to get the information of w_{21} because this will bring the risk of semantic change. To prevent false semantic changes, from the sentence tree, a matrixing layer (borrowing from the seeing layer [40]) is constructed to calculate a visible matrix (VM) that indicates whether there is a direct semantic association between each two hard-position-encoded symbols. The visibility matrix is shown in Figure 3, where, for example, w_2 is visible with w_{12} and both w_3 and w_4 are not visible with w_{12}. Last, VM is used to control the visible area of each token when implementing the mask self-attentive

mechanism in the transformer encoder. After the mask transformer encoder, the system produces the final embedded representations of the input utterance sentence *ut*.
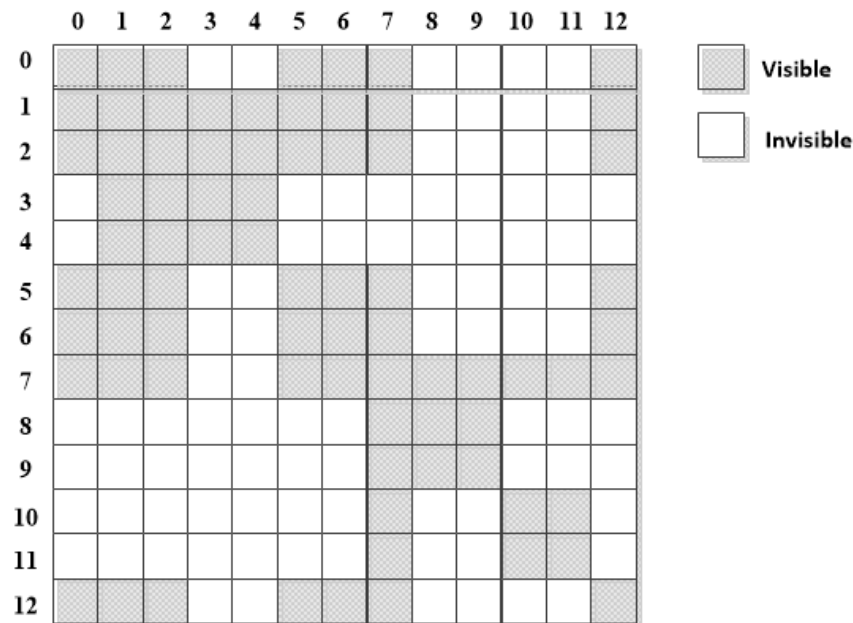

Figure 3. VM matrix.

3.2.2. Sequence Encoding

To fully exploit the representation of a user's preferences from a limited sequence of historical interactions, in the sequence encoding module, a sequential recommendation model named TK-BERT4Rec was adapted to encode the user interaction sequence. In detail, we introduced BERT4Rec [41] as a base model that is essentially a sequential recommendation model based on BERT. Hidden representations in a sequence can be fully explored without the strict constraints of sequence order due to its bidirectional encoding representation capability. However, the model uses only behavioral information, not information about items (e.g., category information for attractions), and the potential learned information remains relatively limited. Similar to TK-BERT, we added a new knowledge layer and modified the embedding layer of the traditional BERT4Rec to improve the knowledge representation ability further. The role of the knowledge layer is to select the triples corresponding to the item entities in the interaction sequence from the knowledge graph. Similar to the dialogue sentence tree, an extended interaction sequence tree is generated by populating these knowledge triples with the corresponding positions of the interaction sequence. Since there is no sentence-like semantic structure relationship between the entity referents of the interaction sequence, but only an order relationship between the words, each token of the interaction sequence tree is numbered sequentially.

As shown in Figure 1, K-BERT4Rec includes four parts: a knowledge layer, an embedding layer, transformer layers, and a TKG. Similar to a dialogue encoding module, given a historical sequence P_u and a TKG, the knowledge layer outputs a sequence tree. To make use of the sequential information of the input sequence, we summed the corresponding item embedding and positional embedding as the output of the embedding layer. In the transformer layer, we stacked hidden representations in L layers together into a representation matrix to simultaneously compute the attention function in all positions. Each transformer layer consists of a multi-head self-attention sub-layer and a position-wise feed-forward network. The former linearly projects the representation matrix into subspaces and then applies the attention function in parallel to produce the output representation. The latter handles non-linear projections through two affine transformations with a Gaussian error linear unit activation in between. After L layers that hierarchically exchange informa-

tion across all positions in the previous layers, the system produces the final embedded representations for all POIs in P_u.

3.2.3. Prediction

The representation e_u of user u is

$$e_u = MLP([e_u^{sem}; e_u^{dem}]),\qquad(4)$$

where e_u^{sem} is the embedding to represent the historical interaction sequence that is produced by the sequence encoding module and e_u^{dem} is the embedding to represent the historical utterance that is produced by the dialogue encoding module. Given the user representation, the probability that an item *poi* would be recommended to a user u is

$$Prob(poi) = soft\max(e_{poi} \cdot e_u),\qquad(5)$$

where e_{poi} is the item embedding for *poi* through the embedding layer. All the POIs are ranked according to the softmax value. The item *poi* with the largest probability value is selected for recommendation.

4. Experiments

4.1. Data Curation

Several datasets have been released to facilitate the study of conversational recommender systems in recent years. Among them, ReDial [12], GoReDial [42], DuRecDial [14], and TG-ReDial [15] were created by human annotation with pre-defined recommendation targets. Unfortunately, none of them are to do with tourism. It is worth noting that Multi-WOZ [43], CrossWOZ [44], and KdConv [45] are dialogue datasets that are related to the field of tourism. However, they all lack well-labeled user interaction sequences.

To fill the current research gap, we developed two conversational recommendation datasets for tourism named CwConvRec and KdConvRec. In our dataset, each conversation can belong to only the tourism domain. To generate the conversation, we obtained all single-domain conversations from CrossWoz and KdConv. To simulate the recommendation scenario, we also extracted POIs (e.g., attractions) from CrossWoz and KdConv to form a visiting record. The entire visiting record was split into several coherent visited subsequences, where each of the POIs was ensured to share at least one common feature (e.g., categories) with another. Since the original dataset lacks feature information of POIs, categories were introduced as features in this study. The categories included ancient ruins, historical buildings, museums, art galleries, parks and gardens, wildlife parks, theme parks, and natural landscapes. Each visited subsequence corresponds to a unique conversation, and each user participates in several conversations. To build knowledge-graph-driven datasets, CwConvRec and KdConvRec needed to be able to provide turn-level knowledge annotations. Although CrossWoz does not contain any ready-made knowledge annotations, it has a domain database with the POIs attribute fields and their attribute values. Therefore, we first constructed an attribute-graph-based knowledge graph for CwConvRec from the travel database in CrossWoz, by combining the poi, attribute, and attribute value into a triplet. Then, we matched entity mentions in a given conversation through a mention dictionary that can be represented as two-tuple:

$$D = (M, E),\qquad(6)$$

where $M = \{m_1, m_2, \ldots, m_k\}$ is the set of all mentions in the knowledge graph already obtained above and $E = \{E_{m_1}, E_{m_2}, \ldots, E_{m_k}\}$ is the set of entities corresponding to the mentions in M. If the term obtained by splitting the conversation utterance precisely matches a mention in the dictionary, we take it as a mention candidate. Each identified mention and its associated entities form a set of mention–entity pairs. Last, we evaluated the probability of each link from a mention to an entity by computing the weighted sum

of features of each possible mention–entity pair. The features include the entity's name length, the link's a priori probability, and the entity relatedness. Since the knowledge form in KdConv contained both unstructured text (e.g., information about the attraction) and structured graphs (e.g., Forbidden City—Surrounding Attraction—South Luogu Lane), the tourism knowledge graph in KdConvRec was directly inherited from KdConv.

In the quality control process for the human-annotated data, each utterance was assigned an annotator and an inspector. We developed a unified annotation specification before annotation to ensure the consistency of the data. Every annotator must perform a real-time inspection and every inspector must complete a full sample inspection and sampling inspection. The detailed statistics of CwConvRec and KdConvRec are shown in Table 1, and examples for the two datasets are illustrated in Figures 4 and 5.

Table 1. Data statistics of our CwConvRec dataset and KdConvRec dataset.

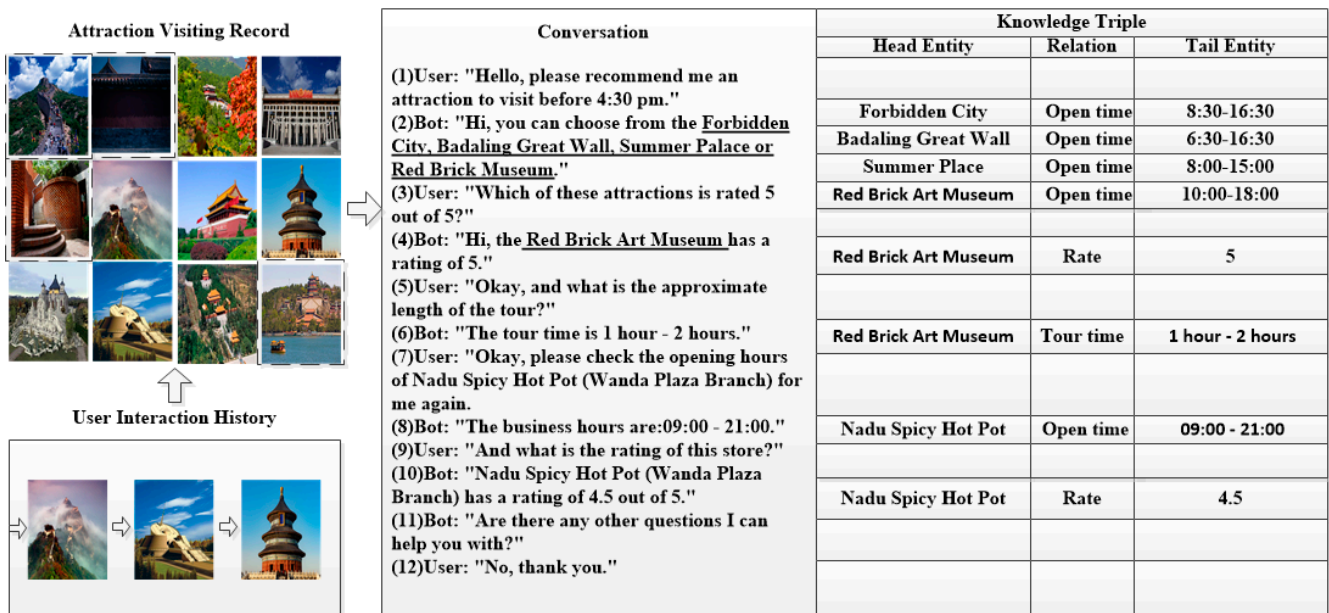| Dataset	CwConvRec	KdConvRec
#Users	82	300
#Dialogue	779	1650
#Utterances	10,082	21,349
#Attractions	465	1154
#Words per Utterance	22.5	20.0
#Attractions per Dialogue	3	4
#Visited Attractions per User	4.5	5.5
#Knowledge Type	Text and Graph	Text and Graph
#Annotation level	Sentence	Sentence

Figure 4. An illustrative example from the CwConvRec dataset. We selected three attractions (boxed by the dashed line) sharing the same category in the user interaction history from the visiting record that corresponded to the attractions (marked by underline) recommended by the conversation. The rightmost section shows the corresponding knowledge triplets.
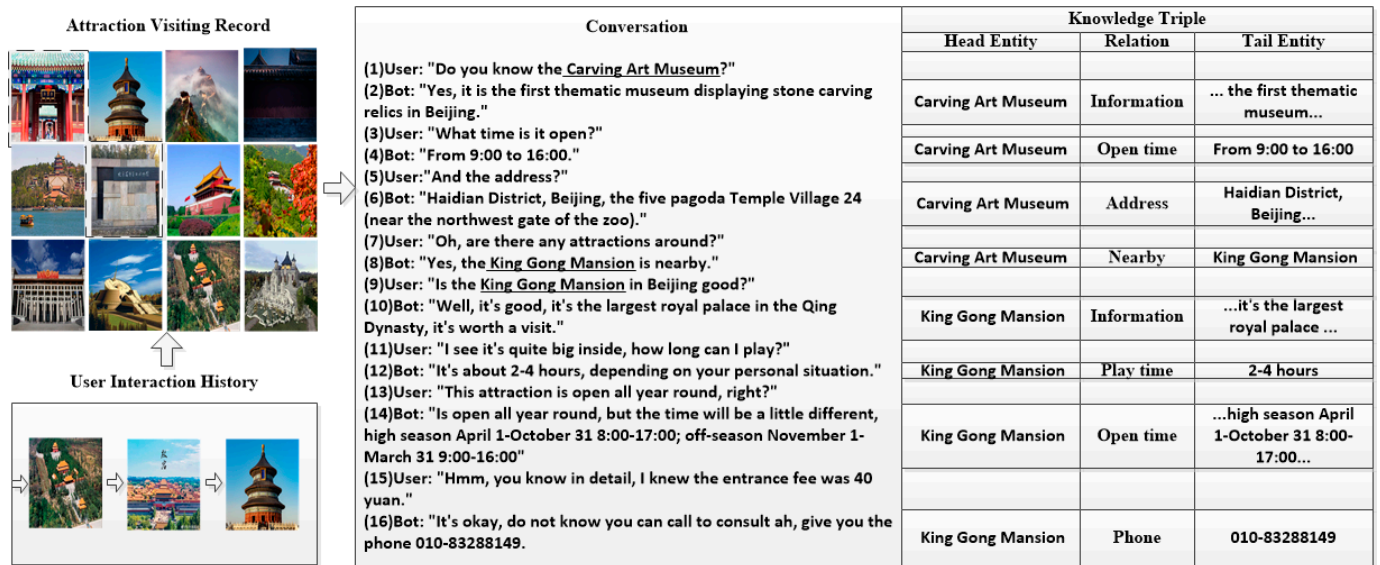
	Conversation	Knowledge Triple		
		Head Entity	Relation	Tail Entity
	(1)User: "Do you know the <u>Carving Art Museum</u>?"			
	(2)Bot: "Yes, it is the first thematic museum displaying stone carving relics in Beijing."	Carving Art Museum	Information	... the first thematic museum...
	(3)User: "What time is it open?"			
	(4)Bot: "From 9:00 to 16:00."	Carving Art Museum	Open time	From 9:00 to 16:00
	(5)User:"And the address?"			
	(6)Bot: "Haidian District, Beijing, the five pagoda Temple Village 24 (near the northwest gate of the zoo)."	Carving Art Museum	Address	Haidian District, Beijing...
	(7)User: "Oh, are there any attractions around?"			
	(8)Bot: "Yes, the <u>King Gong Mansion</u> is nearby."	Carving Art Museum	Nearby	King Gong Mansion
	(9)User: "Is the <u>King Gong Mansion</u> in Beijing good?"			
	(10)Bot: "Well, it's good, it's the largest royal palace in the Qing Dynasty, it's worth a visit."	King Gong Mansion	Information	...it's the largest royal palace ...
	(11)User: "I see it's quite big inside, how long can I play?"			
	(12)Bot: "It's about 2-4 hours, depending on your personal situation."	King Gong Mansion	Play time	2-4 hours
	(13)User: "This attraction is open all year round, right?"			
	(14)Bot: "Is open all year round, but the time will be a little different, high season April 1-October 31 8:00-17:00; off-season November 1-March 31 9:00-16:00"	King Gong Mansion	Open time	...high season April 1-October 31 8:00-17:00...
	(15)User: "Hmm, you know in detail, I knew the entrance fee was 40 yuan."			
	(16)Bot: "It's okay, do not know you can call to consult ah, give you the phone 010-83288149.	King Gong Mansion	Phone	010-83288149

Figure 5. An illustrative example from the KdConvRec dataset. We selected two attractions (boxed by the dashed line) sharing the same category in the user interaction history from the visiting record that corresponded to the attractions (marked by underline) recommended by the conversation. The rightmost section shows the corresponding knowledge triplets.

4.2. Baselines

To evaluate the effectiveness of the proposed approach, we compare it with the following state-of-the-art baselines.

- Popularity: It ranks items according to popularity measured by the number of interactions.
- timeSVD [46]: This model encodes both users and items with low-rank vectors using matrix decomposition and considers that user preferences may change over time as well. It is a dynamic matrix factorization-based recommendation model.
- SASRec [47]: This model adopts the transformer architecture to encode user interaction history without using conversation data. It is a transformer-based sequential recommendation model.
- BERT4Rec [41]: It adopts the deep bi-directional transformer architecture to encode the user interaction history without using conversation data. It is a BERT-based sequential recommendation model.
- TGCRS (SASRec+BERT) [15]: This model adopts SASRec to encode user interaction history and BERT to encode conversation data. This is currently the state-of-the-art conversational recommender system model.

4.3. Evaluation Metrics

In this study, we adopted NDCG and hit rate as evaluation metrics for recommendation performance.

- Normalized discounted cumulative gain (NDCG) [48]: This matrix is commonly used as an evaluation indicator of the ranking results to evaluate the accuracy of the ranking; the NDCG score has also been widely used in evaluating recommender systems. A recommender system usually returns a list of items for a user, and assuming the list length is K, the gap between the sorted list and the user's real interaction list can be evaluated with NDCG@K, where a higher score denotes better performance.
- Hit rate: Like [47], we calculated the hit rate, which represents the fraction of times that the ground-truth next item was among the top item list. The proportion of test cases that have the correctly recommended items in a top K position in a ranking list can be evaluated with HR@K. A higher score denotes better performance.

4.4. Performance Comparison and Analysis

In this study, we chose K = 1, 5, and 10 to illustrate the different metrics results at K. The result of the evaluation on the CwConvRec dataset and the KdConvRec dataset are presented in Tables 2 and 3, respectively. For each method, the results are obtained on the best model.

Table 2. Experimental results of performance comparison using the CwConvRec dataset.

Models	NDCG@1	NDCG@5	NDCG@10	HR@1	HR@5	HR@10
Popularity	0.001	0.004	0.003	0.002	0.005	0.011
timeSVD	0.006	0.011	0.012	0.008	0.015	0.028
SASRec	0.011	0.023	0.054	0.012	0.019	0.035
BERT4Rec	0.016	0.040	0.074	0.015	0.033	0.045
TGCRS	0.035	0.069	**0.089**	0.037	0.102	0.144
DTCRSKG	**0.041 ***	**0.070 ***	0.083 *	**0.043 ***	**0.106 ***	**0.146 ***

* Denotes significance at $p \leq 0.05$.

Table 3. Experimental results of performance comparison using the KdConvRec dataset.

Models	NDCG@1	NDCG@5	NDCG@10	HR@1	HR@5	HR@10
Popularity	0.002	0.005	0.006	0.003	0.004	0.010
timeSVD	0.009	0.018	0.016	0.010	0.019	0.033
SASRec	0.011	0.037	0.054	0.011	0.027	0.059
BERT4Rec	0.017	0.040	0.073	0.017	0.040	0.061
TGCRS	0.050	0.072	**0.137**	0.033	0.079	0.128
DTCRSKG	**0.053 ***	**0.074 ***	0.136 *	**0.034 ***	**0.080 ***	**0.130 ***

* Denotes significance at $p \leq 0.05$.

As shown in Tables 2 and 3, we can see that the performances of the conversational recommender system models significantly outperformed any other non-conversational models, especially regarding the NDCG and hit rate of the recommendation task. This was thanks to the ability of these models to take full advantage of both the historical interaction sequence and historical utterances by combining the merits of the BERT part and the sequential recommendation part. Our proposed DTCRSKG model achieved better performance compared with the TGCRS model on almost all measurements. On the one hand, we used the BERT4Rec model with better performance on sequence recommendation to mine deeper behavioral relationships, and on the other hand, we incorporated the tourism knowledge graph into the BERT model so that the encoding of the conversation contained more knowledge in the travel domain and improved the understanding and representation of the conversation.

However, the improvement degree of our model was lower than expected and there was even a slight deterioration in performance in NDCG@10. Maybe this was due to the relatively short length of the interaction sequence in the tourism dataset, and the implicit information contained in the historical interaction sequence was also minimal. It is not easy to guarantee the correctness of long sequence ordering. Meanwhile, the knowledge graph may still bring noise that interferes with the user features represented by conversation encoding. We take the sentence tree in Figure 6 as an example. We can find that the triple for the [西湖 (West Lake)] entity is {西湖 (West Lake), 位于 (locate), 杭州 (Hanzhou)}. However, if the West Lake mentioned in the conversation is located in Fuzhou instead of Hangzhou, then the knowledge introduced becomes noise.
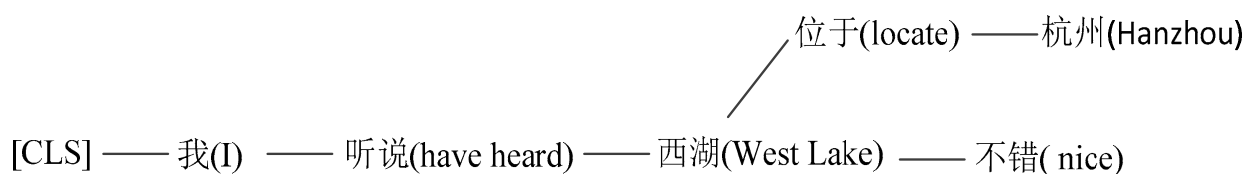
位于(locate) —— 杭州(Hanzhou)

[CLS] —— 我(I) —— 听说(have heard) —— 西湖(West Lake) —— 不错(nice)

Figure 6. A sentence tree example. The triple with "West Lake" in this sentence tree is: "West Lake"—"locate"—"Hangzhou". However, in fact, we are talking about the West Lake in Fuzhou.

4.5. Ablation Experiments

To better understand the proposed DTCRSKG model, namely, TK-BERT+TK-BERT4Rec, we did several sets of ablation experiments using our datasets with the metrics mentioned above. The results are shown in Tables 4 and 5. We found that the model BERT+BERT4Rec was improved regardless of which encoding module incorporated the knowledge. Our proposed TK-BERT+BERT4Rec enhanced the performance to a greater extent than the model BERT+TK-BERT4Rec due to there being more noise in the latter. Among the tested models, the DTCRSKG produced the most improvement.

Table 4. Experimental results of ablation experiments using the CwConvRec dataset.

Models	NDCG@1	NDCG@5	NDCG@10	HR@1	HR@5	HR@10
BERT+BERT4Rec	0.009	0.039	0.054	0.008	0.019	0.035
BERT+TK-BERT4Rec	0.011	0.043	0.055	0.012	0.023	0.037
Tk-BERT+BERT4Rec	0.031	0.063	0.078	0.036	0.101	0.145
TK-BERT+TK-BERT4Rec	**0.041 ***	**0.070 ***	**0.083 ***	**0.043 ***	**0.106 ***	**0.146 ***

* Denotes significance at $p \leq 0.05$.

Table 5. Experimental results of ablation experiments using the KdConvRec dataset.

Models	NDCG@1	NDCG@5	NDCG@10	HR@1	HR@5	HR@10
BERT+BERT4Rec	0.014	0.021	0.045	0.017	0.027	0.061
BERT+TK-BERT4Rec	0.015	0.022	0.045	0.019	0.030	0.063
Tk-BERT+BERT4Rec	0.044	0.067	0.133	0.029	0.076	0.128
TK-BERT+TK-BERT4Rec	**0.053 ***	**0.074 ***	**0.136 ***	**0.034 ***	**0.080 ***	**0.130 ***

* Denotes significance at $p \leq 0.05$.

4.6. Case Analysis

We used the RASA chatbot as a carrier for the case analysis. The chatbot receives input sent by the user and will display a recommendation in the form of text. The system is executed in several steps: (1) the user sends text to the chatbot; (2) after receiving the text from the user, the NLU (natural language understanding) component identifies the user's intention and transfers the data for processing; (3) if the user intends on requesting for an attraction recommendation, the DM (dialog management) component will run the recommendation model and the NLG (natural language generation) component displays the recommendation result; (4) if there is no intent to ask for an attraction but only intends to ask for, say, time or location, the NLG component will generate an informational result for an attraction; (5) if the result satisfies the user or they have no other needs, the conversation is over. The system interface is shown in Figure 7.

In Figure 8, we present a sample to illustrate how our model and TGCRS work in practice. For both models, the user ends up with satisfactory recommended attractions during the conversation given the user interaction sequence, dialogue history, and related knowledge. The last recommended attractions "Longqing Gorge (龙庆峡)" or "Yudu Mountain (玉渡山)" in S7 not only have the same category label "natural landscapes" with attraction "Hantuo Mountain (海坨山)" that is mentioned in S5 but also with the attraction in the user interaction sequence, such as "Hundred Flowers Mountain (百花山)" and "Horn Gorge Primeval forest park (喇叭沟森林公园)". It can be seen that both knowledge-aware

conversational recommendation models could use more correct knowledge to meet the requirements of travel recommendations in a spatio-temporal constrained environment, such as categories, opening time, tour time, and address. Since the category information of the attractions in the interaction sequence is not well utilized, the TGCRs model easily misidentifies Horn Gorge Primeval Forest Park as a park-like attraction and, therefore, fails to quickly predict the user's preference for natural scenery in the absence of user preference information in the current historical dialog. It was not until the user explicitly expressed his preference in S6 that the first correct recommendation for the natural scenery category was given. In contrast, our model incorporated category knowledge into the encoding of the interaction sequence to better understand the user's preference information in the interaction sequence. It predicted that the user may prefer natural attractions, such as Haitou Mountain in S5, before they explicitly expressed their preference.

Figure 7. The system interface, where the dialogue is in Chinese.
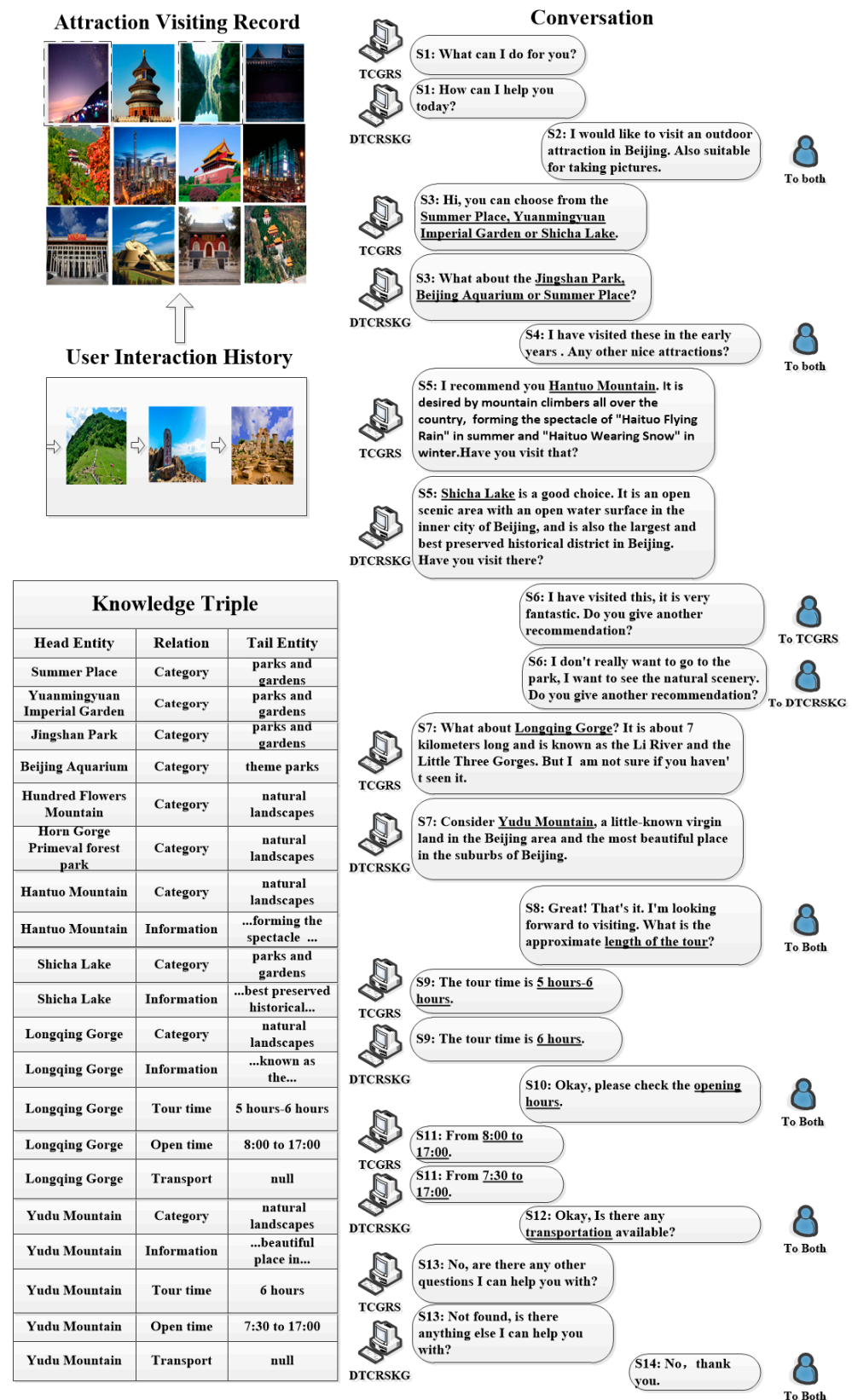
Figure 8. Sample travel conversational recommendation. We underlined the entities and relations in the knowledge triple mentioned therein. We selected two attractions, which is boxed by the dashed line, sharing the same category in the user interaction history from the visiting record that corresponded to the attractions recommended successfully by the conversation. The rightmost section shows the corresponding knowledge triplets.

However, there are still some unsatisfactory points. First, since the limited sequence of user interactions tended to provide little information about user preferences, the conversational recommender system did not initially predict user preferences for natural landscapes in S1–S3. The system did not start giving the correct recommendations until S6 when the user directly stated that they liked the attraction mentioned in S5 and needed additional recommendations. This invariably increases the time for conversational recommendations. The sample also indicated that the accuracy of our proposed conversational recommender system needed to be further improved. Second, when the user mentioned an entity or relation unknown to the system, the system was unable to reply correctly, as in S12 and S13. Even though the proposed conversational recommender system could also produce a knowledge-grounded recommendation, the used knowledge was relatively limited and inappropriate. The knowledge incorporated in the dialogue recommendation system is yet to be resolved to allow for a complement or update.

5. Conclusions

This study constructed two Chinese conversational recommendation datasets, Cw-ConvRec and KdConvRec, for the travel recommender system. We also proposed a deep travel conversational recommender system model DTCRKG as a benchmark for model comparisons. Since both historical dialogues and interaction sequences were well encoded with tourism domain knowledge, the learned user preference representation features were more in-depth and the performance of the model was the best among those tested, which was also verified by our case analysis. In addition, we also found that dialogue data, as a complement to the sequence data, could alleviate data sparsity and provide a reason for the recommendation. Dialogue mode is effective for travel recommender systems, especially regarding cold starts and dynamics. Our work can expand and promote the development of a travel recommender system. In the future, we will explore the semi-supervised methods to amplify the size of dataset annotation and complement TKG or other knowledge graphs to support the proposed model better. Meanwhile, we will investigate semi-supervised learning and GNN-related technologies for a unified model, hopefully solving the noise and sequence length problem.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Schafer, J.B.; Frankowski, D.; Herlocker, J.; Sen, S. Collaborative filtering recommender systems. In *The Adaptive Web*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 291–324. [CrossRef]
2. Pazzani, M.J.; Billsus, D. Content-based recommendation systems. In *The Adaptive Web*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 325–341. [CrossRef]
3. Xue, H.J.; Dai, X.; Zhang, J.; Huang, S.; Chen, J. Deep matrix factorization models for recommender systems. *IJCAI* **2017**, *17*, 3203–3209. [CrossRef]
4. He, X.; Liao, L.; Zhang, H.; Nie, L.; Hu, X.; Chua, T.S. Neural collaborative filtering. In Proceedings of the 26th International Conference on World Wide Web, Perth, Australia, 3–7 April 2017. [CrossRef]
5. Cheng, H.T.; Koc, L.; Harmsen, J.; Shaked, T.; Chandra, T.; Aradhye, H.; Shah, H. Wide & deep learning for recommender systems. In Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, Boston, MA, USA, 15 September 2016. [CrossRef]
6. Guo, H.; Tang, R.; Ye, Y.; Li, Z.; He, X. Deepfm: A factorizationmachine based neural network for ctr prediction. *arXiv* **2017**, arXiv:1703.04247. [CrossRef]
7. Gao, C.; Lei, W.; He, X.; de Rijke, M.; Chua, T.S. Advances and challenges in conversational recommender systems: A survey. *arXiv* **2021**, arXiv:2101.09459. [CrossRef]
8. Jannach, D.; Manzoor, A.; Cai, W.; Chen, L. A survey on conversational recommender systems. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–36. [CrossRef]
9. Sun, Y.; Zhang, Y. Conversational recommender system. In Proceedings of the 41st International Acm Sigir Conference on Research & Development in Information Retrieval, Ann Arbor, MI, USA, 8–12 July 2018. [CrossRef]
10. Lei, W.; Zhang, G.; He, X.; Miao, Y.; Wang, X.; Chen, L.; Chua, T.S. Interactive path reasoning on graph for conversational recommendation. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event, 6–10 July 2020. [CrossRef]
11. Lei, W.; He, X.; Miao, Y.; Wu, Q.; Hong, R.; Kan, M.Y.; Chua, T.S. Estimation-action-reflection: Towards deep interaction between conversational and recommender systems. In Proceedings of the 13th International Conference on Web Search and Data Mining, Houston, TX, USA, 3–7 February 2020. [CrossRef]
12. Li, R.; Ebrahimi Kahou, S.; Schulz, H.; Michalski, V.; Charlin, L.; Pal, C. Towards deep conversational recommendations. *arXiv* **2018**, arXiv:1812.07617. [CrossRef]
13. Chen, Z.; Wang, X.; Xie, X.; Parsana, M.; Soni, A.; Ao, X.; Chen, E. Towards Explainable Conversational Recommendation. *IJCAI* **2020**, *414*, 2994–3000. [CrossRef]
14. Liu, Z.; Wang, H.; Niu, Z.Y.; Wu, H.; Che, W.; Liu, T. Towards conversational recommendation over multi-type dialogs. *arXiv* **2020**, arXiv:2005.03954. [CrossRef]
15. Zhou, K.; Zhou, Y.; Zhao, W.X.; Wang, X.; Wen, J.R. Towards topic-guided conversational recommender system. *arXiv* **2020**, arXiv:2010.04125. [CrossRef]
16. Hoeve, M.; Sim, R.; Nouri, E.; Fourney, A.; de Rijke, M.; White, R.W. Conversations with documents: An exploration of document-centered assistance. In Proceedings of the 2020 Conference on Human Information Interaction and Retrieval, Vancouver, BC, Canada, 14–18 March 2020. [CrossRef]
17. Vakulenko, S.; Kanoulas, E.; De Rijke, M. A Large-Scale Analysis of Mixed Initiative in Information-Seeking Dialogues for Conversational Search. *arXiv* **2021**, arXiv:2104.07096. [CrossRef]
18. Ren, P.; Liu, Z.; Song, X.; Tian, H.; Chen, Z.; Ren, Z.; de Rijke, M. Wizard of Search Engine: Access to Information Through Conversations with Search Engines. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, 11–15 July 2021. [CrossRef]
19. Lyu, D.; Chen, L.; Xu, Z.; Yu, S. Weighted multi-information constrained matrix factorization for personalized travel location recommendation based on geo-tagged photos. *Appl. Intell.* **2020**, *50*, 924–938. [CrossRef]
20. Abbasi-Moud, Z.; Vahdat-Nejad, H.; Sadri, J. Tourism recommendation system based on semantic clustering and sentiment analysis. *Expert Syst. Appl.* **2021**, *167*, 114324. [CrossRef]
21. Pan, H.; Zhang, Z. Research on context-awareness mobile tourism e-commerce personalized recommendation model. *J. Signal Process. Syst.* **2021**, *93*, 147–154. [CrossRef]
22. Zhang, Y.; Han, B.; Gao, X.; Li, H. Personalized travel recommendation via multi-view representation learning. In Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision (PRCV), Beijing, China, 8–11 November 2019. [CrossRef]
23. Bin, C.; Gu, T.; Jia, Z.; Zhu, G.; Xiao, C. A neural multi-context modeling framework for personalized attraction recommendation. *Multimed. Tools Appl.* **2020**, *79*, 14951–14979. [CrossRef]
24. Duan, Z.; Gao, Y.; Feng, J.; Zhang, X.; Wang, J. Personalized tourism route recommendation based on user's active interests. In Proceedings of the 2020 21st IEEE International Conference on Mobile Data Management (MDM), Online, 30 June–3 July 2020. [CrossRef]

25. Zhang, S.; Yochum, P.; Bin, C.; Chang, L. Travel attractions recommendation based on max-negative the gated recurrent unit trajectory mining representation. *J. Phys.* **2020**, *1437*, 012047. [CrossRef]

26. Janarthanam, S.; Lemon, O.; Liu, X.; Bartie, P.; Mackaness, W.; Dalmas, T.; Goetze, J. A spoken dialogue interface for pedestrian city exploration: Integrating navigation, visibility, and question-answering. In Proceedings of the SemDial 2012 (SeineDial): The 16th Workshop on the Semantics and Pragmatics of Dialogue, Paris, France, 19–21 September 2012.

27. Jannach, D.; Zanker, M.; Jessenitschnig, M.; Seidler, O. Developing a conversational travel advisor with advisor suite. *ENTER* **2007**, *7*, 43–52. [CrossRef]

28. Mahmood, T.; Ricci, F.; Venturini, A. Improving recommendation effectiveness: Adapting a dialogue strategy in online travel planning. *Inf. Technol. Tour.* **2009**, *11*, 285–302. [CrossRef]

29. Mahmood, T.; Ricci, F. Improving recommender systems with adaptive conversational strategies. In Proceedings of the 20th ACM Conference on Hypertext and Hypermedia, Torino, Italy, 29 June–1 July 2009. [CrossRef]

30. Shi, C.; Hu, B.; Zhao, W.X.; Philip, S.Y. Heterogeneous information network embedding for recommendation. *IEEE Trans. Knowl. Data Eng.* **2018**, *31*, 357–370. [CrossRef]

31. Song, W.; Duan, Z.; Yang, Z.; Zhu, H.; Zhang, M.; Tang, J. Explainable knowledge graph-based recommendation via deep reinforcement learning. *arXiv* **2019**, arXiv:1906.09506. [CrossRef]

32. Huang, X.; Fang, Q.; Qian, S.; Sang, J.; Li, Y.; Xu, C. Explainable interaction-driven user modeling over knowledge graph for sequential recommendation. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019. [CrossRef]

33. Yang, D.; Guo, Z.; Wang, Z.; Jiang, J.; Xiao, Y.; Wang, W. A knowledge-enhanced deep recommendation framework incorporating gan-based models. In Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM), Singapore, 17–20 November 2018. [CrossRef]

34. Ye, Y.; Wang, X.; Yao, J.; Jia, K.; Zhou, J.; Xiao, Y.; Yang, H. Bayes embedding (bem) refining representation by integrating knowledge graphs and behavior-specific networks. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019. [CrossRef]

35. Sarkar, R.; Goswami, K.; Arcan, M.; McCrae, J.P. Suggest me a movie for tonight: Leveraging Knowledge Graphs for Conversational Recommendation. In Proceedings of the 28th International Conference on Computational Linguistics, Online, 8–13 December 2020. [CrossRef]

36. Fu, Z.; Xian, Y.; Zhu, Y.; Xu, S.; Li, Z.; De Melo, G.; Zhang, Y. HOOPS: Human-in-the-Loop Graph Reasoning for Conversational Recommendation. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Online, 11–15 July 2021. [CrossRef]

37. Liu, Z.; Wang, H.; Niu, Z.Y.; Wu, H.; Che, W. Durecdial 2.0: A bilingual parallel corpus for conversational recommendation. *arXiv* **2021**, arXiv:2109.08877. [CrossRef]

38. Cao, Y.; Wang, X.; He, X.; Hu, Z.; Chua, T.S. Unifying knowledge graph learning and recommendation: Towards a better understanding of user preferences. In Proceedings of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019. [CrossRef]

39. Wong, C.M.; Feng, F.; Zhang, W.; Vong, C.M.; Chen, H.; Zhang, Y.; Chen, H. Improving Conversational Recommendation System by Pretraining on Billions Scale of Knowledge Graph. *arXiv* **2021**, arXiv:2104.14899. [CrossRef]

40. Liu, W.; Zhou, P.; Zhao, Z.; Wang, Z.; Ju, Q.; Deng, H.; Wang, P. K-bert: Enabling language representation with knowledge graph. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020. [CrossRef]

41. Sun, F.; Liu, J.; Wu, J.; Pei, C.; Lin, X.; Ou, W.; Jiang, P. Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019. [CrossRef]

42. Kang, D.; Balakrishnan, A.; Shah, P.; Crook, P.; Boureau, Y.L.; Weston, J. Recommendation as a communication game: Self-supervised bot-play for goal-oriented dialogue. *arXiv* **2019**, arXiv:1909.03922. [CrossRef]

43. Budzianowski, P.; Wen, T.H.; Tseng, B.H.; Casanueva, I.; Ultes, S.; Ramadan, O.; Gašić, M. Multiwoz–a large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling. *arXiv* **2018**, arXiv:1810.00278. [CrossRef]

44. Zhu, Q.; Huang, K.; Zhang, Z.; Zhu, X.; Huang, M. Crosswoz: A largescale chinese cross-domain task-oriented dialogue dataset. *Trans. Assoc. Comput. Linguist.* **2020**, *8*, 281–295. [CrossRef]

45. Zhou, H.; Zheng, C.; Huang, K.; Huang, M.; Zhu, X. Kdconv: A Chinese multi-domain dialogue dataset towards multi-turn knowledge-driven conversation. *arXiv* **2020**, arXiv:2004.04100. [CrossRef]

46. Koren, Y. Collaborative filtering with temporal dynamics. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, 28 June–1 July 2009. [CrossRef]

47. Kang, W.C.; McAuley, J. Self-attentive sequential recommendation. In Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM), Singapore, 17–20 November 2018. [CrossRef]

48. Yilmaz, E.; Kanoulas, E.; Aslam, J.A. A simple and efficient sampling method for estimating AP and NDCG. In Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Singapore, 20–24 July 2008. [CrossRef]