

# CRFR: Improving Conversational Recommender Systems via Flexible Fragments Reasoning on Knowledge Graphs

Jinfeng Zhou<sup>1,2</sup>, Bo Wang<sup>2,1</sup>, Ruifang He<sup>2,1</sup>, Yuexian Hou<sup>1</sup>

<sup>1</sup>State Key Laboratory of Communication Content Cognition,  
People's Daily Online, Beijing, China

<sup>2</sup>College of Intelligence and Computing, Tianjin University, Tianjin, China  
{jfzhou, bo\_wang, rfhe, yxhou}@tju.edu.cn

## Abstract

Although paths of user interests shift in knowledge graphs (KGs) can benefit conversational recommender systems (CRS), explicit reasoning on KGs has not been well considered in CRS, due to the complex of high-order and incomplete paths. We propose CRFR, which effectively does explicit multi-hop reasoning on KGs with a conversational context-based reinforcement learning model. Considering the incompleteness of KGs, instead of learning single complete reasoning path, CRFR flexibly learns multiple reasoning fragments which are likely contained in the complete paths of interests shift. A fragments-aware unified model is then designed to fuse the fragments information from item-oriented and concept-oriented KGs to enhance the CRS response with entities and words from the fragments. Extensive experiments demonstrate CRFR's SOTA performance on recommendation, conversation and conversation interpretability.

## 1 Introduction

Different from traditional one-shot recommendation systems (Jannach et al., 2020), conversational recommender systems (CRS) obtain users' interests through multi-turn conversation, and make recommendations with responses. Typical CRS consists of two parts: recommender and response generation. The recommender aims to understand users' dynamic preference from contextual utterances to find the items matching the preference best. Subsequently, the response generation aims to generate appropriate sentences asking for more information or exhibiting the recommended items and related explanation. Recommender and response generation are expected to be mutually beneficial.

However, contextual utterances are usually insufficient to understand users' preference (Zhou et al., 2020a). External knowledge, especially knowledge graphs (KGs), helps CRS to alleviate the problem

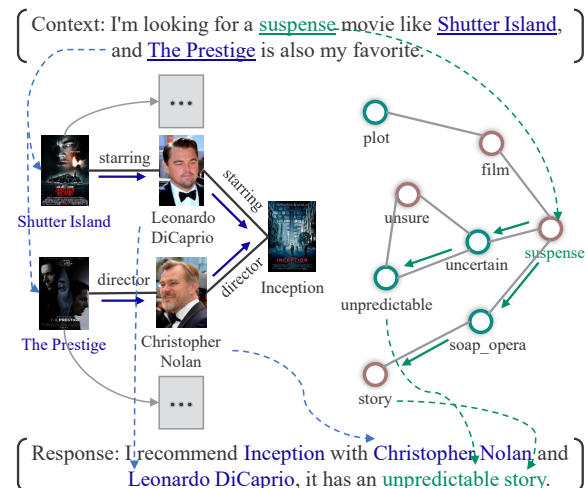


Figure 1: An example of conversational movie recommendation. In item-oriented and concept-oriented KGs, multiple high-order reasoning paths contain fragments, from which items and concepts support the recommendation and the interpretability of response.

(Chen et al., 2019; Zhou et al., 2020a). Existing KG-based CRS still have some issues.

The first issue is the lack of explicit reasoning, especially high-order explicit reasoning, in KG to track the deep shift of user interest in conversations. For example, in Figure 1, the user prefers "Shutter Island" and "The Prestige" due to certain attributes, i.e., actor "DiCaprio" and director "Nolan". In this case, both "DiCaprio" and "Nolan" lead to a 2-order explicit reasoning of the user's interest shift to "Inception" in the KG. Such multiple high-order explicit reasoning paths are strong evidence to recommend "Inception". Besides item KG, explicit high-order reasoning in concept KG can further describe the shift of concept words of interest, e.g., from "suspense" to "unpredictable" in Figure 1, and enhance the interpretability of response. However, because KG nodes usually have many neighbors, effective high-order explicit reasoning is challengeable. Therefore, instead of explicit reasoning, one alternate in existing KG-based CRS (Chen et al.,

2019; Zhou et al., 2020a) is to aggregate neighborhood of interest entities in context, which can implicitly gather high-order relations by multi-turn aggregation. Another alternate strategy (Ma et al., 2020) is to explicitly grow a high-order tree in KG starting from interest entities to cover the potential preference, without aiming to the best recommended items. Both strategies can improve CRS, but have not full appreciation of the advantage of tracking interest shift in KG.

The second issue is that KGs are usually incomplete to track the path of interest shift. For example, in Figure 1, the user mentioned the two movies probably because he/she likes "Nolan" and "DiCaprio". Multiple paths in KG help to locate the subspace of the user's interest and generate interpretive utterances in line with people's dialogue behavior, e.g., "...Inception with Christopher Nolan and Leonardo DiCaprio...". However, KGs can not record all relations of interest entities involved in real-world diverse dialogues (Ma et al., 2020; Hayati et al., 2020; Sarkar et al., 2020). Therefore, it is often difficult to achieve a complete reasoning path of interest shift within limited number of hops.

To address these issues, we improve Conversational Recommender Systems with Flexible Fragments Reasoning on KGs (CRFR). CRFR uses an explicit multi-hop reasoning method to model the user's interest shift in a conversation with respect to an item-oriented KG (DBpedia (Bizer et al., 2009)) and a concept-oriented KG (ConceptNet (Speer et al., 2017)). We formalize the interest shift as a Markov Decision Process and propose an explicit policy-guided reasoning model based on reinforcement learning. Due to the possible absence of complete reasoning path in limited hop, instead of finding single best reasoning path, the learned policy flexibly obtains a set of high-order optimal path fragments. The obtained path fragments point to the destination of the user's interest shift and are likely to be contained in complete interest shift paths.

To make recommendation with multiple reasoning path fragments, we further design a fragments-aware recommendation module to fuse fragments information to learn the final representation of user preference. Subsequently, in a reasoning information enhanced dialog module, the user preference representation is integrated into the response at token level. In this way, both the entities and words in reasoning path fragments can be flexibly involved

and thus enhance the informativeness and interpretability of generated response.

Our contributions are summarized as follows:

(1) To explicitly model the multiple shift paths of user's interest in a conversation, we propose a multi-hop policy reasoning model based on reinforcement learning with respect to item-oriented and concept-oriented KGs.

(2) To avoid the difficulty of exact reasoning in incomplete KG, CRFR flexibly obtains an optimal set of path fragments from heterogeneous KGs.

(3) To effectively use obtained fragments, CRFR dynamically encodes the fragments to facilitate the recommendation and dialog response.

(4) Extensive experiments demonstrate that CRFR exceeds the state-of-the-art baselines in recommendation accuracy and generates high quality response with more interpretability.

## 2 Related Work

Conversational recommender systems (CRS) can be divided into two types. One is agent-driven CRS known as "system ask, user respond" (Gao et al., 2021; Zhang et al., 2018; Zou et al., 2020). This strategy obtains user preference by asking for predefined attributes (Sun and Zhang, 2018; Lei et al., 2020a) and makes response by the user's feedback (Luo et al., 2020; Xu et al., 2021; Li et al., 2020). Graph reasoning (Lei et al., 2020b) and comment information (Chen et al., 2020) can make the recommendation process of agent-driven CRS interpretable. Although agent-driven CRS is popular in specific field, predefined template limits its generalization ability.

The other is user-oriented CRS known as "user talk, system understand" (Kang et al., 2019; Liao et al., 2019; Liu et al., 2020; Zhou et al., 2020b). Due to the insufficient context, user-oriented CRS often need the help of external knowledge, e.g., knowledge graphs (KGs). Closely related to our work, Chen et al. (2019) integrates KGs to connect recommender and response generation. Zhou et al. (2020a) further uses item-level and word-level KGs to align two different semantic spaces. However, these two approaches only implicitly aggregating the neighbor information with GCN-based methods, instead of conducting explicit reasoning in KGs to track the deep user preference shift.

An ideal explicit KG reasoning for CRS is supposed to accurately describe the shift process of user interests using a single complete path. How-

ever, it is usually challengeable due to the incompleteness of the KGs, the large search space and the diversity of expressing user interests. Different from the approaches looking for single path (Moon et al., 2019), the reasoning tree (Ma et al., 2020) executes relatively undirected reasoning and lacks concept relations in the reasoning.

### 3 Preliminaries

In general, a knowledge graph  $\mathcal{G}$  with entity set  $\mathcal{E}$  and relation set  $\mathcal{R}$  is defined as  $\mathcal{G} = \{(e, r, e') \mid e, e' \in \mathcal{E}, r \in \mathcal{R}\}$ , where each triplet  $(e, r, e')$  represents a relation  $r$  from entity  $e$  to entity  $e'$ . In this paper, we denote the item-oriented knowledge graph DBpedia (Bizer et al., 2009) as  $\mathcal{G}_D$  and the concept-oriented knowledge graph ConceptNet (Speer et al., 2017) as  $\mathcal{G}_C$ . Correspondingly, their entity sets are  $\mathcal{E}_D$  and  $\mathcal{E}_C$  respectively. A multi-hop complete reasoning path  $p_{e_0, e_t}$ , which connects a starting entity  $e_0$  and a target entity  $e_t$ :  $p_{e_0, e_t} = \{e_0 \xrightarrow{r_1} e_1 \xrightarrow{r_2} \dots \xrightarrow{r_t} e_t\}$ . A subset of a complete reasoning path is a path fragment  $p$ .

In this work, given an n-turn conversation history  $\mathcal{H}$ , which contains the utterances of each turn, CRFR takes conversation history  $\mathcal{H}$  with the user  $u$ , graphs  $\mathcal{G}_D$  and  $\mathcal{G}_C$  as input, utilizes context information to perform flexible policy reasoning on  $\mathcal{G}_D$  and  $\mathcal{G}_C$  to obtain high-order path fragments set  $\mathcal{P}$  that point to the destination of the user’s interest shift, and further uses  $\mathcal{P}$  as a guide to output the response containing the recommended item and the explanation of the recommendation.

### 4 Methodology

In this section, we introduce our proposed CRFR model. The overall framework is shown in Figure 2. CRFR executes flexible multi-hop policy reasoning on KGs  $\mathcal{G}_D$  and  $\mathcal{G}_C$  to obtain high-order path fragments that point to the destination of the user’s interest shift on the two KGs respectively. Path fragments from  $\mathcal{G}_D$  help the fragments-aware recommendation module to obtain a more accurate user preference representation. Further, the reasoning information enhanced dialog module integrates the user’s preference representation at the token level to flexibly select two types of path fragment information to guide response generation.

#### 4.1 Flexible Policy Reasoning

Following Chen et al. (2019) and Zhou et al. (2020a), we first use R-GCN (Schlichtkrull et al.,

2018) and GCN (Edwards and Xie, 2016; Kipf and Welling, 2017) to encode knowledge graphs  $\mathcal{G}_D$  and  $\mathcal{G}_C$ , respectively. A representation model of entities is pre-trained by adjusting two KGs to identical semantic space. We formalize the user interest shift as Markov Decision Process (MDP), and learn the shift policy by performing multi-hop path reasoning in KGs. This section uses  $\mathcal{G}_D$  as a MDP environment to learn path reasoning policy. Similar method is applied to  $\mathcal{G}_C$ .

**State.** Initial representation  $p_{u_0}$  of current user  $u$  is the aggregate embedding of common entities in  $\mathcal{G}_D$  and  $\mathcal{H}$ . The aggregate embedding is obtained by the pre-trained representation model.  $p_{u_0}$  is used as the starting state  $s_0 \in \mathcal{S}$  of path reasoning, i.e.,  $s_0 = p_{u_0}$ .  $s_t \in \mathcal{S}$  is the state representation of  $t$ -th step.  $s_t$  is obtained by concatenating  $p_{u_0}$ , the embeddings of the reasoning history entities  $h_t = (e_0, e_1, \dots, e_{t-1})$  and the entity  $e_t$  reached at  $t$ -th step. Formally,  $s_t = [p_{u_0}; h_t; e_t]$ .

**Action.** For state  $s_t$ , its action space  $\mathcal{A}_t$  is all neighbors of  $e_t$  in KG, except for the entities that already appeared in the path.  $\mathcal{A}_t = \{e \mid (e_t, r, e) \in \mathcal{G}_D, e \in \mathcal{E}_D \setminus \{e_0, e_1, \dots, e_{t-1}\}\}$ . For the nodes with a large number of neighbors, we design a pruning strategy function  $g(u, a)$  to select important neighbors.  $g(u, a) = p_{u_0}^T \cdot e_a$ , where  $e_a$  is the embedding of entity  $a$ . The trimmed action space is  $\mathcal{A}'_t = \{a \mid \text{topk}(g(u, a)), a \in \mathcal{A}_t\}$ , where the size of the action space is a hyperparameter separately set for each KG. Specially, action space  $\mathcal{A}_0$  of the initial state  $s_0$  is the set of entities mentioned in  $\mathcal{H}$ .

**Transition.** Given current state  $s_t = [p_{u_0}; h_t; e_t]$ , the agent chooses  $a_t \in \mathcal{A}'_t$  as the next action. The next state  $s_{t+1}$  is transitioned to by  $s_{t+1} = \mathcal{T}(s_t, a_t) = [p_{u_0}; h_t; e_t; e_{t+1}]$ ,  $\mathcal{T} : \mathcal{S} \times \mathcal{T} \rightarrow \mathcal{S}$ .

**Reward.** Intuitively, we reward the decision made at each step by how well it matches the user’s preference. Inspired by (Lv et al., 2020; Xu et al., 2020), we adopt a soft reward, which calculates cosine similarity  $\text{sim}_{e_T, a_t}$  between  $e_T$  and the action entity  $a_t$  as the reward  $R_t$  at step  $t$ , where  $e_T$  is the embedding of target entity or target concepts,  $\alpha$ ,  $\beta$ ,  $\varepsilon_1$  and  $\varepsilon_2$  are hyperparameters.

$$R_t = \begin{cases} \alpha, & \text{sim}_{e_T, a_t} \geq \varepsilon_1 \\ \text{sim}_{e_T, a_t} + \beta, & \varepsilon_1 > \text{sim}_{e_T, a_t} \geq \varepsilon_2 \\ \max(0, \text{sim}_{e_T, a_t}), & \text{otherwise.} \end{cases} \quad (1)$$

**Policy Optimization.** We employ the A2C method (Sutton and Barto, 2018) for optimization.

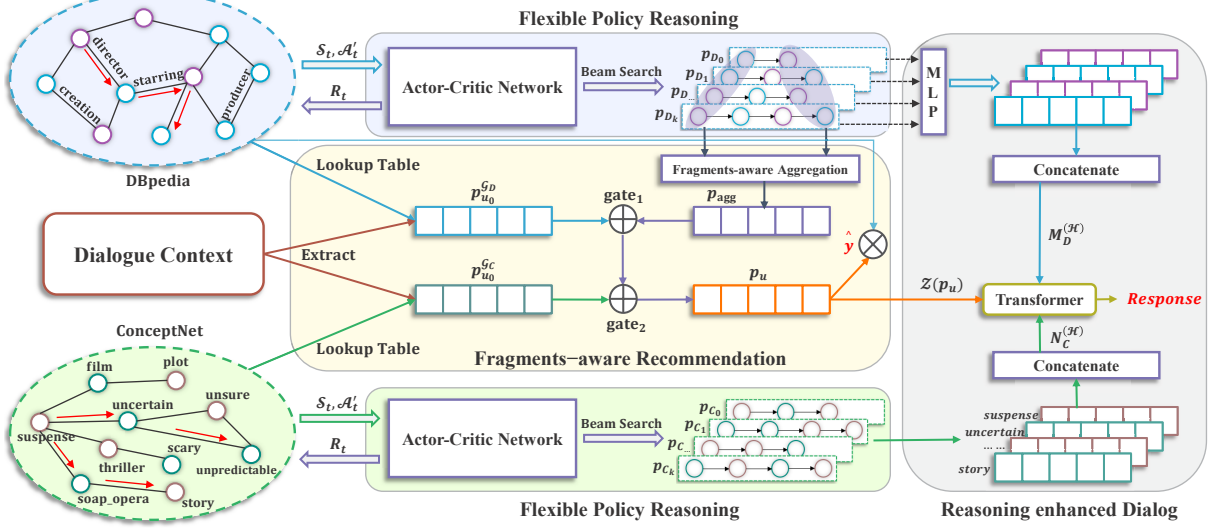


Figure 2: **The Overview of CRFR.** Policy reasoning flexibly obtains high-order path fragments in DBpedia ( $\mathcal{G}_D$ ) and ConceptNet ( $\mathcal{G}_C$ ), which are beneficial to recommendation and dialogue. Here,  $p_{D_i}$  and  $p_{C_i}$  are reasoning path fragments in  $\mathcal{G}_D$  and  $\mathcal{G}_C$ , respectively.  $p_u$  is the final user preference representation.

The goal of the agent is to learn a path finding policy  $\pi(a_t, s_t, \mathcal{A}'_t)$ , which calculates the probability distribution of each action  $a_t$  based on current state  $s_t$  and trimmed action space  $\mathcal{A}'_t$ . At the same time, the agent learns a critic network, which calculates the value of  $a_t$  according to  $s_t$ . We use two fully connected layers as the policy network, and each layer has an ELU activation function and a dropout layer. The output of the network is further sent to a softmax layer and a fully connected layer to obtain the probability  $p(a_t | s_t, \mathcal{A}'_t) = \pi(a_t, s_t, \mathcal{A}'_t)$  and value  $q(a_t)$  of each action. Specifically, the learning goal is to maximize the expected cumulative reward for all users:

$$\mathcal{J}_{\theta_{RL}} = \mathbb{E}_{\pi} \left[ \sum_{t=0}^T \gamma^t R_t \right], \quad (2)$$

where  $\theta_{RL}$  are the parameters to be learned,  $\gamma$  is the discount parameter. Following reinforce algorithm, the gradient of the learning object becomes:

$$\nabla \mathcal{J}_{\theta_{RL}} = \mathbb{E}_{\pi} \left[ \nabla_{\theta} \log \pi_{\theta}(a_t, s_t, \mathcal{A}'_t) (G_t - q(a_t)) \right], \quad (3)$$

where  $G_t$  is the discounted cumulative reward starting from state  $s_t$  to the final time step  $T$ .

**Flexible Fragments Reasoning.** Due to the incompleteness of KG (Sarkar et al., 2020), as a key idea of this work, instead of only modeling user’s interest with the destinations of complete reasoning paths, we prefer to model user’s interest shift with partial reasoning path, i.e., reasoning fragments.

After training the two policy networks, guided by the probability of each action made by the policy network, we employ beam search to explore the candidate path fragments  $\mathcal{P}_{\text{candidate}}$  on two KGs  $\mathcal{G}_D$  and  $\mathcal{G}_C$ , respectively. We select the fragments with top generating probabilities. Selected fragments are supposed to be most consistent with the user’s interest shift process, and will be used in the following fragments-aware recommendation and reasoning information enhanced dialog.

## 4.2 Fragments-aware recommendation

Given fragments obtained by reasoning in item-oriented KG  $\mathcal{G}_D$ , we design a fragments-aware recommendation module to improve the representation of user preference.

First, we collect the beginning nodes of all path fragments, that is, the more important entities selected from the entities mentioned by  $\mathcal{H}$ . The embeddings of these entities are obtained by a simple lookup operation. We connect the embedding of the beginning entities to the matrix  $B^{(\mathcal{H})}$ . Similarly, we also combine the embeddings of the destination of all path fragments to get the matrix  $D^{(\mathcal{H})}$ . Next, we apply the self-attention mechanism to obtain a single aggregate representation of the two matrices. Specifically, we use self-attention to learn the representation of  $b^{(\mathcal{H})}$  for  $B^{(\mathcal{H})}$ :

$$b^{(\mathcal{H})} = \alpha * B^{(\mathcal{H})}, \quad (4)$$

$$\alpha = \text{softmax} (W_{\alpha}^1 \tanh (W_{\alpha}^2 B^{(\mathcal{H})})),$$

where  $W_{\alpha}^1$  and  $W_{\alpha}^2$  are learnable parameters. The

aggregated representation  $d^{(\mathcal{H})}$  of  $D^{(\mathcal{H})}$  is also obtained in the same way. Then, we design an interactive aggregation method to obtain path fragments' fusion information  $p_{\text{agg}}$ :

$$p_{\text{agg}} = \text{ReLU} \left( W_{\text{agg}}^1 \left( b^{(\mathcal{H})} \odot d^{(\mathcal{H})} \right) \right) + \text{ReLU} \left( W_{\text{agg}}^2 \left( b^{(\mathcal{H})} + d^{(\mathcal{H})} \right) \right), \quad (5)$$

where  $W_{\text{agg}}^1$  and  $W_{\text{agg}}^2$  are learnable parameters and  $\odot$  means element-wise product. Next, we use two gate networks to fuse the information of entities ( $p_{u_0}$  in  $\mathcal{G}_D$ ) and the information of concepts ( $p_{u_0}$  in  $\mathcal{G}_C$ ) in  $\mathcal{H}$ , respectively, to strengthen the user's preference representation:

$$\eta = \sigma \left( W_g \left( p_{\text{agg}} \parallel p_{u_0} \right) \right), \quad (6)$$

$$p_{\text{agg}} = \eta \cdot p_{\text{agg}} + (1 - \eta) \cdot p_{u_0},$$

where  $\sigma(\cdot)$  is the sigmoid function, and  $\parallel$  is the concatenation operation.  $W_g$  is the parameter learned separately by the two gate networks. The final fusion representation  $p_{\text{agg}}$  is the user's preference representation  $p_u$ .

Finally, we conduct inner product of user and item representations to predict their matching score:  $\hat{y}(u, i) = \text{softmax} \left( p_u^T e_i \right)$ , where  $e_i$  is the embedding for item entity  $i$ .

### 4.3 Reasoning information enhanced dialog

To enhance dialog, we use the Transformer's decoder (Vaswani et al., 2017) to merge the user's preference with the information of the path fragments from  $\mathcal{G}_D$  and  $\mathcal{G}_C$  respectively at token level. The decoder can flexibly select semantics in two kinds of reasoning fragments to enhance the informativeness and interpretability of the generated response. Intuitively, human interest shift is a continuous process, and the reasoning paths provide explanations of recommendations. Therefore, we encode each path fragment  $p_{D_i} \in \mathcal{P}_D$  of  $\mathcal{G}_D$  inference into a path vector  $m^{(\mathcal{H})}$ , and concatenate  $m^{(\mathcal{H})}$  to the matrix  $M_D^{(\mathcal{H})}$ . Specifically, we concatenate the embeddings of all entities on the fragment, and use MLP to obtain the representation of the path fragment information:

$$m^{(\mathcal{H})} = \text{MLP} \left( (e_0 \parallel e_1 \parallel \dots \parallel e_T) \right), \quad (7)$$

where  $\parallel$  is the concatenation operation. For the concept information inferred from  $\mathcal{G}_C$ , we merge the embeddings of the concept words of each order on

the path fragments into the matrix  $N_C^{(\mathcal{H})}$ . In the decoder's  $i$ -th layer, we first fuse the output of the self-attention sub-layer with the word bias of the user preference representation:  $V^{i-1} = V^{i-1} + \mathcal{Z}(p_u)$ , Where  $\mathcal{Z} : \mathbb{R}^{d_u} \rightarrow \mathbb{R}^{d_w}$ . Next, we add two multi-head attention modules, so that the output  $V^{i-1}$  after fusing user preference can conduct attention operation with path information  $M_D^{(\mathcal{H})}$  and concept information  $N_C^{(\mathcal{H})}$  in turn. Finally, the output embedding matrix  $V^i$  of the  $i$ -th layer is obtained through a fully connected feed-forward network.

Merged path fragments information is supposed to improve the possibility of reasoning entities or words to appear in response. In this way, we can significantly improve the coherence and interpretability of response. We will demonstrate it in experimental section.

### 4.4 Optimization

We train the parameters in four steps. First, we pre-train the entity representation of KGs as Sec.4.1. Then, use Eq.3 to optimize the policy parameters  $\theta_{RL}^{\mathcal{G}_D}$  and  $\theta_{RL}^{\mathcal{G}_C}$  of the two agents. Next, to optimize the recommendation, we adopt cross-entropy loss:

$$L_{Rec} = - \sum_{j=1}^O \sum_{i=1}^I \left[ y_{ij} \log \hat{y}^{(j)}(u, i) - (1 - y_{ij}) \log \left( 1 - \hat{y}^{(j)}(u, i) \right) \right], \quad (8)$$

where  $i$  is the index of items,  $j$  is the conversation index. After completing the optimization of recommendation, we get the parameter  $\theta_{Rec}$ . On this basis, we use the loss function of the generative model to learn the dialogue parameter  $\theta_{Gen}$ :

$$L_{Gen} = - \frac{1}{K} \sum_{k=1}^K \log \left( p_{\theta_{Gen}} \left( h_k \mid h_1, \dots, h_{k-1} \right) \right), \quad (9)$$

where  $K$  is the count of turns in a conversation  $\mathcal{H}$ . We calculate  $L_{Gen}$  for each utterance  $h_k$  from  $\mathcal{H}$  and perform gradient descent to update parameters.

## 5 Experiments

### 5.1 Experimental Setup

**Dataset.** Experiments are conducted on the ReDial dataset (Li et al., 2018). ReDial has 10,006 multi-turn conversations and a total of 182,150 utterances of movie recommendation seekers and recommenders. Movies mentioned in utterances are manually annotated. We identified unlabeled

Model	R@1	R@10	R@50
Popularity	0.012	0.061	0.179
TextCNN	0.013	0.068	0.191
REDIAL	0.024	0.140	0.320
KBRD	0.032	0.166	0.345
KGSF	0.037	0.183	0.371
<b>CRFR</b>	<b>0.040</b>	<b>0.202</b>	<b>0.399</b>
+RandomWalk	0.033	0.170	0.361

Table 1: **Evaluation of recommendation.** +RandomWalk means using random walk to replace the policy reasoning part. (t-test with p-value < 0.05)

entities in utterances by NER and linked them to DBpedia nodes. We divided the training, validation and test set according to the ratio of 8:1:1.

**Baselines.** Used baselines are as follows. *Popularity* only sorts the items according to historical recommendation frequency. *TextCNN* (Kim, 2014) is a recommendation model learning user preference representations by CNN-based encoding of contextual utterances. *Transformer* (Vaswani et al., 2017) is a dialog generation model of classical Transformer framework. *REDIA* (Li et al., 2018) is the benchmark CRS model on the ReDial corpus, which is mainly based on HRED (Sordoni et al., 2015; Serban et al., 2016), a recommendation system based on autoencoder and a sentiment analysis module. *KBRD* (Chen et al., 2019) is a Knowledge-Based CRS model that only uses DBpedia to enhance the user’s representation. *KGSF* (Zhou et al., 2020a) is a state-of-the-art CRS model that aligns two KGs with mutual information maximization for conversation recommendation.

**Implementation Details.** The default parameter settings can be found in appendix.

## 5.2 Evaluation on Recommendation

To evaluate the recommendation performance of our model, we adopt widely-used Recall metrics including Recall@1, Recall@10 and Recall@50. They evaluate whether the top-k auto-recommended items contain the ground-truth item recommended by human recommended.

**Overall Evaluation.** As shown in Table 1, REDIAL exceeds the classical recommendation models, i.e., Popularity and TextCNN, by using items in the dialogue contexts. KBRD and KGSF get further improvement by fusing the entities and items information in the knowledge graph. Our model achieves the best results using explicit multi-hop

Model	Dist-2	Dist-3	Dist-4	Ratio
Transformer	0.148	0.151	0.137	0.194
REDIAL	0.225	0.236	0.228	0.158
KBRD	0.281	0.379	0.439	0.298
KGSF	0.302	0.433	0.521	0.324
<b>CRFR</b>	0.345	<b>0.516</b>	<b>0.639</b>	<b>0.383</b>
- Concept	0.298	0.407	0.478	0.335
- DBpedia	<b>0.355</b>	0.503	0.594	0.267
- Preference	0.287	0.374	0.422	0.240

Table 2: **Evaluation of diversity and informativeness.** Dist refers to Distinct. Ratio refers to Item Ratio. The last three rows are ablation results. (t-test with p-value < 0.05)

policy reasoning to obtain the optimal set of path fragments learning user’s interest shift, which outperforms the best baselines with a large margin.

**The Effect of Policy Reasoning.** Here, we specially examine the contribution of policy reasoning in our model. To this end, we replace the multi-hop policy reasoning by random walk (Spitzer, 2013) in the knowledge graph to obtain multi-hop path fragments. As shown in Table 1, the use of random walk significantly reduce the recommendation performance. This happens because random walk is undirected, while a significant advantage of multi-hop policy reasoning is being guided by real user interest shift.

**The Effect of Explicit Reasoning.** We also specially examine the contribution of explicit reasoning in our model. As shown in Figure 3, compared with the baselines, the advantage of our model becomes more significant with the increase of the average number of neighbors of KG nodes. This happens because more neighbors mean more noise for reasoning. Implicit reasoning through aggregating all direct neighborhood is very sensitive to such noise. However, our explicit high-order reasoning is more effective to find correct reasoning directions from noisy neighbors.

## 5.3 Evaluation on Conversation

**Diversity and Informativeness.** To evaluate the conversation, we firstly calculate "Distinct-n" to measure the diversity of responses and calculate "Item Ratio" which is the proportion of responses containing items to evaluate the informativeness of responses.

As shown in Table 2, CRFR outperforms all baselines on corpus-level language diversity and greatly improves the item ratio. Transformer has

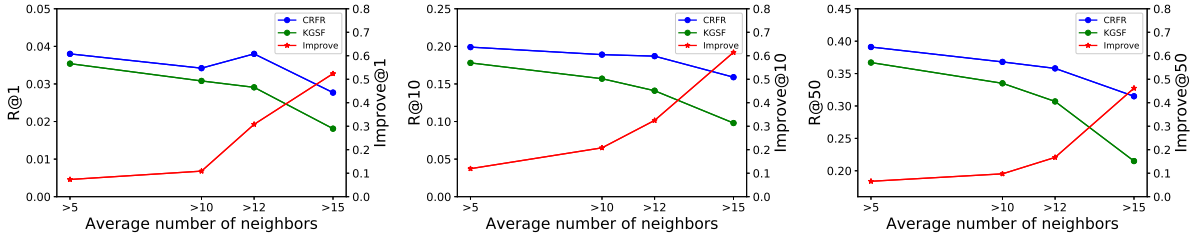


Figure 3: **The effect of explicit reasoning.** As the average number of neighbors increases, the performance of our model is always higher than the SOTA baseline, and the improvement is more and more significant.

significant advantages in modeling the relationship between words and items using self-attention mechanism. On this basis, KBRD and KGSF utilize external knowledge to increase the occurrence of the item, but they only use low-order KG information of entities and concepts mentioned in the context. Our method further enhances the occurrence of valuable items and entities consistent with user interest by integrating the high-order reasoning path fragments into response generation. In this way, our model enhances the informativeness of responses.

**Interpretability.** Secondly, we evaluate the interpretability of response by examining whether there are two logically linked entities in the response or across the response and the context. "logically linked" means being linked in DBpedia. The idea is that containing two logically linked entities are often a necessary condition to have a interpretative expression in a response. In Table 3, "2ER" (2 Entities Ratio) indicates the proportion of responses containing at least two entities in all responses containing entities. "Inner-Con." counts the logically linked entity pairs in responses. "Inter-Con." counts the logically linked entity pairs across the response and context. As shown in Table 3, our model is better than baselines on interpretability with a large margin. This happens because the entities in our explicit reasoning path are exactly logically linked, and this naturally increase the occurrence of logically linked entities pairs in response or across the response and the context.

**Human Evaluation.** In human evaluation, we examine "Flu." (fluency), "Coher." (coherence), "Info." (informativeness) and "Inter." (interpretability). Fluency and coherence are used to evaluate the language quality of generated responses. Informativeness evaluates whether the response has incorporated rich entity knowledge. Interpretability evaluates whether the response explain the reason of recommended item. 100 multi-turn conversa-

Model	2ER	Inner-Con.	Inter-Con.
KBRD	0.141	0.066	0.038
KGSF	0.161	0.078	0.039
<b>CRFR</b>	<b>0.331</b>	<b>0.130</b>	<b>0.076</b>
-Concept	0.261	0.083	0.052
-DBpedia	0.247	0.078	0.045
-Preference	0.171	0.106	0.043

Table 3: **Evaluation of interpretability.** We abbreviate 2 Entities Ratio as 2ER. The last three rows are ablation results. (t-test with p-value < 0.05)

Model	Flu.	Coher.	Info.	Inter.
Transformer	1.92	1.91	1.81	1.70
REDIAL	2.01	1.97	1.69	1.62
KBRD	2.18	2.09	1.95	1.88
KGSF	2.32	2.21	2.12	2.08
<b>CRFR</b>	<b>2.53</b>	<b>2.39</b>	<b>2.36</b>	<b>2.27</b>

Table 4: **Human evaluation of conversation.** (t-test with p-value < 0.05)

tions are randomly sampled from the test set for human evaluation. The responses are scored from 1 to 3 for each indicator by 5 workers. The average score is finally calculated. Results are shown in Table 4. Among all the baselines, KGSF has the best performance on all indicators. However, KGSF is more inclined to generate words that are repeated in the context and have no practical meaning, and the recommendation response is less explanatory. Our model achieves the best performance on all human evaluation indicators. This further verifies the advantages of our model in automatic evaluation.

**Ablation Study.** In ablation study, we examine three variants of CRFR: (1) "-Concept", which removes the reasoning path fragments information from  $\mathcal{G}_C$ ; (2) "-DBpedia", which removes the reasoning path fragments information from  $\mathcal{G}_D$ ; (3) "-Preference", which removes token level user preference information. As shown in Table 2 and Ta-

Context-1	Have you seen <a href="#">Eight Crazy Nights</a> . That seems like a fun movie Not seen it though.
KBRD	I haven't seen that one.
KGSF	I haven't seen that one of any of the first?
CRFR	I also recommend watching <a href="#">Jack and Jill (2011 film)</a> , it was a few of the <a href="#">Adam Sandler</a> movie.
Ground-truth	It's a funny movie I watch it every year.
Context-2	I'm fantastic! thanks! What kind of movies do you enjoy? I'm looking for some <a href="#">scary</a> movies like <a href="#">Annabelle (film)</a> or <a href="#">The Conjuring</a> , I also really liked <a href="#">Sinister (film)</a> .
KBRD	I haven't seen that one.
KGSF	Have you seen <a href="#">The Omen</a> ?
CRFR	I also like <a href="#">Sinister 2</a> if you like <a href="#">ghost</a> movies.
Ground-truth	Oh nice! Have you seen <a href="#">The Sixth Sense</a> or <a href="#">The Ring (2002 film)</a> .
Context-3	Hello, what are you looking for? \\ Hey! How are you? \\ Not too bad. I'm looking for a <a href="#">kid</a> movie! Like <a href="#">The Incredibles (film score)</a> .
KBRD	How about <a href="#">Toy Story</a> ?
KGSF	I would recommend <a href="#">WALL-E</a> .
CRFR	I would recommend <a href="#">The Croods</a> . It is a really <a href="#">cute</a> movie with <a href="#">Nicholas Cage</a> .
Ground-truth	Oh, cool, Have you seen <a href="#">Isle Of Dogs (2018)</a> ? That's a great new one I have n't seen <a href="#">The Incredibles (film score)</a> .
Context-4	I'm open to any movie what would you recommend? I like action <a href="#">comedy</a> like <a href="#">Rush Hour (film series)</a> have you seen it? \\ No, i have not seen it, is it <a href="#">interesting</a> ?
KBRD	Have you seen <a href="#">Who Am I (2014 film)</a> ?
KGSF	Have you seen <a href="#">21 Jump Street (film)</a> ?
CRFR	I also recommend <a href="#">Rush Hour 2</a> full of <a href="#">laughs</a> .
Ground-truth	Yes, very filled with action and <a href="#">Christ tucker</a> and <a href="#">Jackie chan</a> are great! Do you like action or something else?

Table 5: Cases generated by different models. Crossing context and response, [Entities connectivity](#) in DBpedia and [Concepts shift](#) in ConceptNet are highlighted. \\ denotes to switch conversation roles.

ble 3, all these three features are indispensable. Especially, user preference information at the token level is most essential, without which the informativeness and interpretability drop significantly. Its function is to flexibly select the information of reasoning path fragments from  $\mathcal{G}_C$  and  $\mathcal{G}_D$  to improve the quality of the generated response.

#### 5.4 Case Study

Four cases from three CRS models and ground truth are selected in Table 5. Compared with the other two models, CRFR has four main advantages: (1) CRFR is more likely to recommend a specific film instead of chatting without recommendation, being consistent to "Ratio" in Table 2; (2) The items recommended by CRFR is more likely to have explicit relation with the items mentioned by the user, being consistent to "Inter-Con." in Table 3, e.g., in the 1st case, "Jack and Jill" in response and "Eight Crazy Nights" in context share the actor "Adam Sandler". This benefits from our explicit multi-hop reasoning in item-oriented KG; (3) Especially, one main advantage of CRFR is to naturally tell the items' relation as an explanation, being consistent to "Inner-Con." in Table 3 and "Inter." in Table 4. It is noted that this ability is even often absent in human ground-truth in selected cases. This benefits from using personalized embedding of reasoning fragments in response decoder; (4) Another important advantage of CRFR is to make more friendly

and persuasive explanation with descriptive words related to the feature words of user intention in the context, e.g., *ghost* vs. *scary*, *cute* vs. *kid* and *laughs* vs. *comedy*. This benefits from our explicit multi-hop reasoning in concept-oriented KG.

## 6 Conclusion

We propose CRFR, which significantly improves the agent response in conversational recommendation by exhibiting items having more clear higher-order relations with users' contextual intention and containing more persuasive explanation. As the essential advantage of this approach, explicit reasoning of high-order fragments in two heterogeneous knowledge graphs is performed by a reinforcement learning model. High-order path fragments obtained by explicit reasoning on item-oriented KG help the model to better track the user preference shift in conversation. The same reasoning on concept-oriented KG further improves the interpretability of response with informative concept words. Heterogeneous fragments are personalized encoded to finally enhance the response generation. Extensive experiments demonstrated that CRFR is superior to the SOTA baselines on recommendation, explanation and language quality.

In future, we will explore to make better use of path information to further improve the interpretability of responses.



## Acknowledgements

This work was supported by a grant from the National Key Research and Development Program of China (2018YFC0809804), State Key Laboratory of Communication Content Cognition (Grant No.A32003), the Artificial Intelligence for Sustainable Development Goals (AI4SDGs) Research Program, National Natural Science Foundation of China (U1736103, 61976154, 6140232361876128), the National Key Research and Development Program (2017YFE0111900).

## References

- Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. 2009. [Dbpedia - A crystallization point for the web of data](#). *J. Web Semant.*, 7(3):154–165.
- Qibin Chen, Junyang Lin, Yichang Zhang, Ming Ding, Yukuo Cen, Hongxia Yang, and Jie Tang. 2019. [Towards knowledge-based recommender dialog system](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 1803–1813. Association for Computational Linguistics.
- Zhongxia Chen, Xiting Wang, Xing Xie, Mehul Parsana, Akshay Soni, Xiang Ao, and Enhong Chen. 2020. [Towards explainable conversational recommendation](#). In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 2994–3000. ijcai.org.
- Michael Edwards and Xianghua Xie. 2016. [Graph convolutional neural network](#). In *Proceedings of the British Machine Vision Conference 2016, BMVC 2016, York, UK, September 19-22, 2016*. BMVA Press.
- Chongming Gao, Wenqiang Lei, Xiangnan He, Maarten de Rijke, and Tat-Seng Chua. 2021. [Advances and challenges in conversational recommender systems: A survey](#). *CoRR*, abs/2101.09459.
- Shirley Anugrah Hayati, Dongyeop Kang, Qingxi-aoyang Zhu, Weiyan Shi, and Zhou Yu. 2020. [INSPIRED: toward sociable recommendation dialog systems](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 8142–8152. Association for Computational Linguistics.
- Dietmar Jannach, Ahtsham Manzoor, Wanling Cai, and Li Chen. 2020. [A survey on conversational recommender systems](#). *CoRR*, abs/2004.00646.
- Dongyeop Kang, Anusha Balakrishnan, Pararth Shah, Paul A. Crook, Y-Lan Boureau, and Jason Weston. 2019. [Recommendation as a communication game: Self-supervised bot-play for goal-oriented dialogue](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 1951–1961. Association for Computational Linguistics.
- Yoon Kim. 2014. [Convolutional neural networks for sentence classification](#). In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar; A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1746–1751. ACL.
- Thomas N. Kipf and Max Welling. 2017. [Semi-supervised classification with graph convolutional networks](#). In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.
- Wenqiang Lei, Xiangnan He, Yisong Miao, Qingyun Wu, Richang Hong, Min-Yen Kan, and Tat-Seng Chua. 2020a. [Estimation-action-reflection: Towards deep interaction between conversational and recommender systems](#). In *WSDM '20: The Thirteenth ACM International Conference on Web Search and Data Mining, Houston, TX, USA, February 3-7, 2020*, pages 304–312. ACM.
- Wenqiang Lei, Gangyi Zhang, Xiangnan He, Yisong Miao, Xiang Wang, Liang Chen, and Tat-Seng Chua. 2020b. [Interactive path reasoning on graph for conversational recommendation](#). In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, pages 2073–2083. ACM.
- Hanze Li, Scott Sanner, Kai Luo, and Ga Wu. 2020. [A ranking optimization approach to latent linear critiquing for conversational recommender systems](#). In *RecSys 2020: Fourteenth ACM Conference on Recommender Systems, Virtual Event, Brazil, September 22-26, 2020*, pages 13–22. ACM.
- Raymond Li, Samira Ebrahimi Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Chris Pal. 2018. [Towards deep conversational recommendations](#). In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 9748–9758.
- Lizi Liao, Ryuichi Takanobu, Yunshan Ma, Xun Yang, Minlie Huang, and Tat-Seng Chua. 2019. [Deep conversational recommender in travel](#). *CoRR*, abs/1907.00710.
- Zeming Liu, Haifeng Wang, Zheng-Yu Niu, Hua Wu, Wanxiang Che, and Ting Liu. 2020. [Towards conversational recommendation over multi-type dialogs](#).

- In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 1036–1049. Association for Computational Linguistics.
- Kai Luo, Scott Sanner, Ga Wu, Hanze Li, and Hojin Yang. 2020. [Latent linear critiquing for conversational recommender systems](#). In *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, pages 2535–2541. ACM / IW3C2.
- Xin Lv, Xu Han, Lei Hou, Juanzi Li, Zhiyuan Liu, Wei Zhang, Yichi Zhang, Hao Kong, and Suhui Wu. 2020. [Dynamic anticipation and completion for multi-hop reasoning over sparse knowledge graph](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 5694–5703. Association for Computational Linguistics.
- Wenchang Ma, Ryuichi Takanobu, Minghao Tu, and Minlie Huang. 2020. [Bridging the gap between conversational reasoning and interactive recommendation](#). *CoRR*, abs/2010.10333.
- Seungwhan Moon, Pararth Shah, Anuj Kumar, and Rajen Subba. 2019. [Opendialkg: Explainable conversational reasoning with attention-based walks over knowledge graphs](#). In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 845–854. Association for Computational Linguistics.
- Rajdeep Sarkar, Koustava Goswami, Mihael Arcan, and John P. McCrae. 2020. [Suggest me a movie for tonight: Leveraging knowledge graphs for conversational recommendation](#). In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, pages 4179–4189. International Committee on Computational Linguistics.
- Michael Sejr Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. [Modeling relational data with graph convolutional networks](#). In *The Semantic Web - 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3-7, 2018, Proceedings*, volume 10843 of *Lecture Notes in Computer Science*, pages 593–607. Springer.
- Iulian Vlad Serban, Alessandro Sordoni, Yoshua Bengio, Aaron C. Courville, and Joelle Pineau. 2016. [Building end-to-end dialogue systems using generative hierarchical neural network models](#). In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, pages 3776–3784. AAAI Press.
- Alessandro Sordoni, Yoshua Bengio, Hossein Vahabi, Christina Lioma, Jakob Grue Simonsen, and Jian-Yun Nie. 2015. [A hierarchical recurrent encoder-decoder for generative context-aware query suggestion](#). In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015*, pages 553–562. ACM.
- Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. [Conceptnet 5.5: An open multilingual graph of general knowledge](#). In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, pages 4444–4451. AAAI Press.
- Frank Spitzer. 2013. *Principles of random walk*, volume 34. Springer Science & Business Media.
- Yueming Sun and Yi Zhang. 2018. [Conversational recommender system](#). In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018*, pages 235–244. ACM.
- Richard S Sutton and Andrew G Barto. 2018. *Reinforcement learning: An introduction*. MIT press.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008.
- Jun Xu, Haifeng Wang, Zheng-Yu Niu, Hua Wu, Wanxiang Che, and Ting Liu. 2020. [Conversational graph grounded policy learning for open-domain conversation generation](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 1835–1845. Association for Computational Linguistics.
- Kerui Xu, Jingxuan Yang, Jun Xu, Sheng Gao, Jun Guo, and Ji-Rong Wen. 2021. [Adapting user preference to online feedback in multi-round conversational recommendation](#). In *WSDM '21, The Fourteenth ACM International Conference on Web Search and Data Mining, Virtual Event, Israel, March 8-12, 2021*, pages 364–372. ACM.
- Yongfeng Zhang, Xu Chen, Qingyao Ai, Liu Yang, and W. Bruce Croft. 2018. [Towards conversational search and recommendation: System ask, user respond](#). In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018*, pages 177–186. ACM.
- Kun Zhou, Wayne Xin Zhao, Shuqing Bian, Yuanhang Zhou, Ji-Rong Wen, and Jingsong Yu. 2020a. [Improving conversational recommender systems via knowledge graph based semantic fusion](#). In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, pages 1006–1014. ACM.

Kun Zhou, Yuanhang Zhou, Wayne Xin Zhao, Xiaoke Wang, and Ji-Rong Wen. 2020b. [Towards topic-guided conversational recommender system](#). In *Proceedings of the 28th International Conference on Computational Linguistics, COLING 2020, Barcelona, Spain (Online), December 8-13, 2020*, pages 4128–4139. International Committee on Computational Linguistics.

Jie Zou, Yifan Chen, and Evangelos Kanoulas. 2020. [Towards question-based recommender systems](#). In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020*, pages 881–890. ACM.

## A Appendix

**Implementation Details.** The default parameter settings across all experiments are as follows. We set the embedding dimension of entities in the knowledge graph to 128. In the policy reasoning module, we set the maximum reasoning path length to  $T=3$ , so the state vector  $s_t$  is of size 512. We prune the action space with the maximum size 50 and 128 for graph  $\mathcal{G}_D$  and  $\mathcal{G}_C$ , respectively. The discount factor  $\gamma$  is 0.99. In graph  $\mathcal{G}_D$ , the rewards  $\alpha$  and  $\beta$  are 2.0 and 0.5. The thresholds  $\varepsilon_1$  and  $\varepsilon_2$  are 0.8 and 0.5. In graph  $\mathcal{G}_C$ , the rewards  $\alpha$  and  $\beta$  are 2.0 and 1.0. The thresholds  $\varepsilon_1$  and  $\varepsilon_2$  are 0.9 and 0.8. For the policy network, The parameter matrix of the two fully connected layers  $W_1 \in \mathbb{R}^{512 \times 256}$  and  $W_2 \in \mathbb{R}^{256 \times 128}$ . The action dropout rate is 0.5. Our policy reasoning model is trained for 20 epochs using Adam optimization with the learning rate of 1e-4 and the batch size of 32. When using beam search for path reasoning, we set the sampling sizes at each step to [5,4,1] for graph  $\mathcal{G}_D$  and [25,5,1] for graph  $\mathcal{G}_C$ . In the recommendation module, we set hidden representation dimensions to 128. In the dialog module, all embedding dimensions are set to 300, and word2vec is used to initialize word embedding. In particular,  $d_u$  and  $d_w$  are 128 and 300 respectively. In the training phase, we use Adam optimizer with the default parameter setting. We set the batch size of both modules to 32. The learning rate is 4e-4 for the recommendation and 1e-3 for the conversation. Gradient clipping restricts the norm of the gradients within [0, 0.1]. We implement the models in PyTorch and train on an NVIDIA P100. The total training time takes approximately 24 hours.