Aligning Recommendation and Conversation via Dual Imitation

Jinfeng Zhou[1,2], **Bo Wang**[1,2✉], **Minlie Huang**[3], **Dongming Zhao**[4], **Kun Huang**[4],
Ruifang He[1,2], **Yuexian Hou**[1]

[1]College of Intelligence and Computing, Tianjin University, Tianjin, China

[2]State Key Laboratory of Communication Content Cognition, People's Daily Online, Beijing, China

[3]The CoAI Group, DCST, Institute for Artificial Intelligence, State Key Lab of Intelligent Technology and Systems,

[3]Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China

[4]AI Lab, China Mobile Communication Group Tianjin Co., Ltd.

{jfzhou, bo_wang}@tju.edu.cn   aihuang@tsinghua.edu.cn

Abstract

Human conversations of recommendation naturally involve the shift of interests which can align the recommendation actions and conversation process to make accurate recommendations with rich explanations. However, existing conversational recommendation systems (CRS) ignore the advantage of user interest shift in connecting recommendation and conversation, which leads to an ineffective loose coupling structure of CRS. To address this issue, by modeling the recommendation actions as recommendation paths in a knowledge graph (KG), we propose DICR (**D**ual **I**mitation for **C**onversational **R**ecommendation), which designs a dual imitation to explicitly align the recommendation paths and user interest shift paths in a recommendation module and a conversation module, respectively. By exchanging alignment signals, DICR achieves bidirectional promotion between recommendation and conversation modules and generates high-quality responses with accurate recommendations and coherent explanations. Experiments demonstrate that DICR outperforms the state-of-the-art models on recommendation and conversation performance with automatic, human, and novel explainability metrics.

1 Introduction

Conversational recommendation systems (CRS) (Liu et al., 2020; Li et al., 2022) aim to conduct recommendations during conversations with users (Gao et al., 2021). Compared with traditional recommendation systems (Wang et al., 2019; Xian et al., 2019), CRS's two main advantages are understanding user's dynamic interest during the conversation and making persuasive response with coherent explanations of the recommendation (Jannach et al., 2021). In both advantages, user interest shift plays an essential role. As the dialog in Fig. 1, the successful recommendation of "*Iron Man 3*" in the
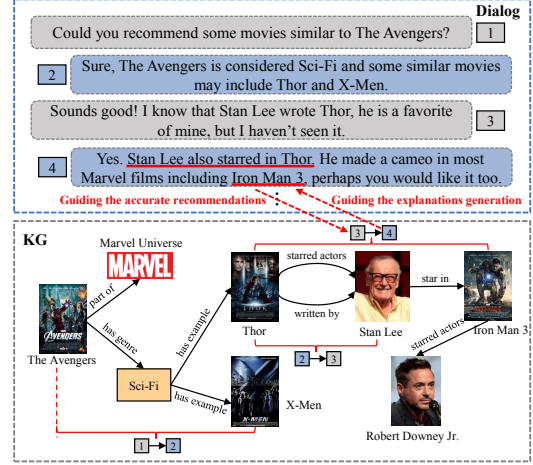

Figure 1: The interest shift process expressed in the conversation can guide the generation of explainable recommendation path. The explainable recommendation path, in turn, can guide the generation of explainable response containing accurate recommendations. Recommendation and conversation maximize mutual benefits in bidirectional guidance.

final response is achieved by tracking and reasoning the user interest shift of "*The Avengers→Sci-Fi→Thor→Stan Lee→Iron Man 3*". Furthermore, the final response is persuasive because of utilizing part of the user interest shift, i.e., "*Thor→Stan Lee→Iron Man 3*", as the explanation.

Due to the limited context (Hayati et al., 2020), a recommendation module based on knowledge graph (KG) is helpful to track the user interest shift in conversation. As shown in Figure 1, formally corresponding with the paths in KG, the user interest shift in conversation not only guides the reasoning-based prediction of recommendation, but also guides the explanation generation in response.

However, existing KG-enhanced CRS models (Chen et al., 2019; Zhou et al., 2020, 2021, 2022b; Zhang et al., 2022a) have not made full use of the user interest shift to tightly align the KG-based recommendation and conversation. Consequently, one issue is the less accurate recommendation due to

using unrelated entities in conversation to support the recommendation instead of using coherent entities in user interest shift. The other issue is the lack of explanation in response due to black-box representation of the user preference ignoring the explicit preference logic in user interest shift.

To address these issues in CRS, we propose to align the explicit behaviors of recommendation reasoning and conversation process, which are described as the recommendation paths and interest shift paths, respectively. As in Figure 1, a **recommendation path** is an explicit path in KG consisting of explicit relations of entity nodes and ending with a predicted recommended entity; an **interest shift path** is an implicit path in dialog context consisting of implicit relations of entity words. The recommendation path and interest shift path are concrete manifestations of the user interest shift in KG and dialog, respectively. The sequence of interest entities shared by the two paths facilitates the alignment of recommendation reasoning and conversation process, which can be effectively achieved by imitation learning (Ho and Ermon, 2016). Therefore, we propose a dual imitation framework named DICR (**D**ual **I**mitation for **C**onversational **R**ecommendation). DICR designs bidirectional alignment signals from dual imitation learning to improve the CRS by forcing the recommendation and conversation to behave similarly to the shared user interest shift.

Precisely, in a conversation-aware recommendation module, to align the recommendation reasoning to the conversational user interest shift, the recommendation side of the dual imitation, i.e., path imitation, adopts adversarial reinforcement learning to make the recommendation reasoning policy imitate the user interest shift in conversation. The reasoned recommendation paths are provided to the conversation module as alignment signals. In a recommendation-aware conversation module, to align the conversation process to the recommendation paths, the conversation side of the dual imitation, i.e., knowledge imitation and semantic imitation, which refine weight distribution and semantic encoding of recommendation paths by imitating the human response and the utterance statement semantic of golden explanations, respectively. These two imitations also provide the recommendation module with the rewards as the alignment signals indicating how well the predicated recommendation paths consist to the conversation context.

Our contributions are summarized as follows:

(1) To the best of our knowledge, we are the first to adopt imitation learning in CRS to integrate recommendation and conversation tightly. We design a dual imitation framework named DICR, which aligns recommendation and conversation behavior and promotes bidirectional improvement, taking recommendation paths and conversational rewards from the dual imitation as alignment signals.

(2) The dual imitation benefits the knowledge acquisition and semantic generation, promoting the accuracy of recommendation and significantly improving the explanations of recommendations in generated responses with coherent knowledge.

(3) Extensive experiments demonstrate that DICR outperforms the SOTA models on both recommendation and conversation performance with automatic, human and novel explainability metrics.

2 Related Work

Conversational recommendation systems (CRS) aim to obtain user interests through conversational interaction and make persuasive recommendations (Jannach et al., 2021; Gao et al., 2021; Ren et al., 2021). To track the user interest shift in conversation, an intuitive strategy is to ask related questions (Kostric et al., 2021; Zhang et al., 2022b) which leads to question-based CRS (Lei et al., 2020; Deng et al., 2021). Limited by predefined templates for asking and recommending, it is difficult for question-based CRS to flexibly adopt different contexts and converse in a human-like manner.

Towards more flexible conversation, generation-based CRS (Li et al., 2018; Zhang et al., 2021; Liang et al., 2021) capture user interests from context and generate responses containing persuasive explanations for recommendation. Limited by sparse context and language complexity (Lu et al., 2021; Yang et al., 2022), it is challenging for generation-based CRS to track user interest shift in conversation. As a popular solution, KG-enhanced CRS (Moon et al., 2019; Ma et al., 2021; Zhou et al., 2022a) involve knowledge graphs about explicit relations among potential interest items.

Although KG-enhanced CRS has achieved significant improvements, most approaches (Chen et al., 2019; Zhou et al., 2020, 2021, 2022b) adopt a black-box style to transfer implicit and sparse information between recommendation and conversation. The explicit interest reasoning path in KG and explicit interest shift path in conversation have
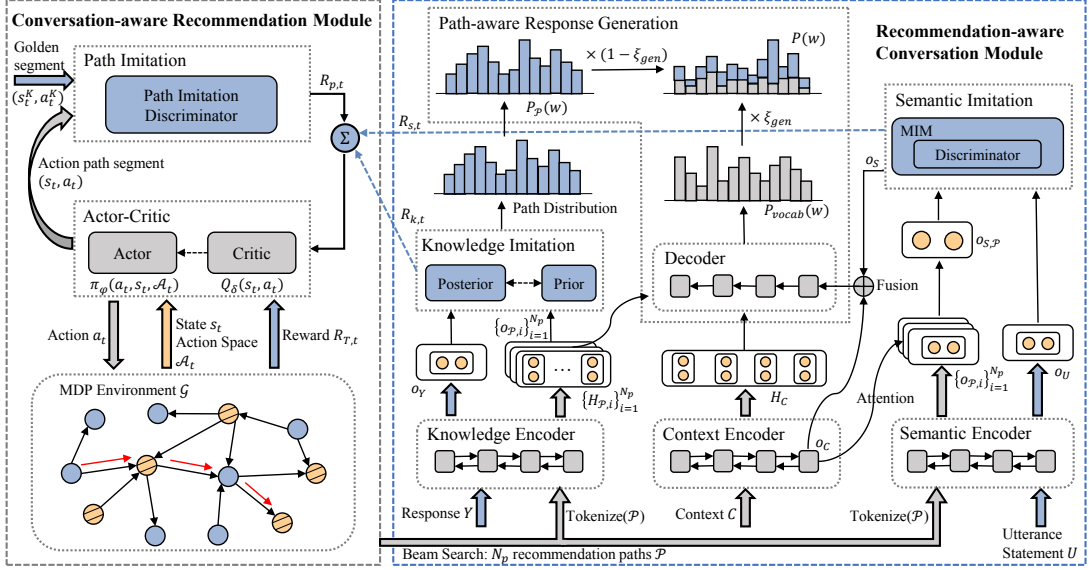
Figure 2: The dual imitation architecture of the proposed DICR model.

a good chance to align the recommendation and conversation to benefit each other.

3 Problem Formalization

A KG $\mathcal{G} = \{(e, r, e') \mid e, e' \in \mathcal{E}, r \in \mathcal{R}\}$, where $\mathcal{E}$ is entity set, and $\mathcal{R}$ is relation set. Each triplet (e, r, e') indicates that the head entity e and the tail entity e' are connected by the relation r. In this paper, a recommendation path is a multi-hop reasoning path p on $\mathcal{G}$: $p = \left\{ e_0 \xrightarrow{r_1} e_1 \xrightarrow{r_2} \ldots \xrightarrow{r_t} e_t \right\}$.

Suppose we have a conversational recommendation corpus $\mathcal{D}$ parallel to a knowledge graph $\mathcal{G}$, in which the interests (e.g., movies) mentioned in $\mathcal{D}$ are linked to the entities in $\mathcal{G}$. $C = (c_1, c_2, \ldots, c_n)$ is the conversation context, where c_i is an utterance. I is the set of recommendation items under C. $Y = (y_1, y_2, \ldots, y_m)$ is a response containing I, where y_i is a token. $K = \left\{ e_0 \xrightarrow{r_1^K} e_1 \xrightarrow{r_2^K} \ldots \xrightarrow{r_l^K} e_l \right\}$ is a golden interest shift path connecting the interest entities $e_{0,1,\ldots,l}$ in C and Y. K also matches a recommendation path p in $\mathcal{G}$. K can be extracted by identifying entities in the conversation and linking the entities to the nodes in $\mathcal{G}$. The logical utterance statement of K is U, which is the explanation of recommendation, e.g., given a one-hop reasoning path in $\mathcal{G}$ ("*Thor*", "*written_by*", "*Stan Lee*"), its tokenized U can be "*Thor is written by Stan Lee.*".

In this paper, given a conversation context C and a KG $\mathcal{G}$, we aim to generate a response Y containing the recommendation set I and explanation U. We design a novel CRS model, in which KG path reasoning obtains explicit reasoning path set $\mathcal{P}$ from $\mathcal{G}$ to help generate Y containing recommendation set I and coherent explanation U of I.

4 Approach

4.1 Architecture Overview

As shown in Fig. 2, in DICR, the **conversation-aware recommendation module** learns recommendation path reasoning policy with adversarial reinforcement learning. The path imitation discriminator aligns the recommendation paths with the golden interest shift path to reward the agent $R_{p,t}$ to optimize the reasoning policy. As a result, the top tokenized recommendation paths are provided to the conversation module as alignment signals. In the **recommendation-aware conversation module**, the knowledge imitation aligns the prior and posterior recommendation knowledge in tokenized recommendation paths and human response, respectively. The semantic imitation uses Mutual Information Maximization (MIM) to align the semantic encoding of recommendation paths with that of the utterance statement of the golden interest shift path. These imitations refine distribution of knowledge and overall words and thus benefit the path-aware response generation. They also generate rewards $R_{k,t}$ and $R_{s,t}$ as alignment signals to guide the recommendation path reasoning. Finally, DICR performs a joint training to bidirectionally promote the recommendation and conversation with alignment signals from dual imitation.

4.2 Conversation-aware Recommendation Module

In this module, we formalize the user interest shift in conversation as a Markov Decision Process (MDP) (Sutton and Barto, 2018) through KG paths to reason interest shift path with adversarial reinforcement learning. We construct KG embeddings (Bordes et al., 2013) of each entity. The entities mentioned in C are extracted by fuzzy matching, whose embeddings are averaged as the preference representation of user u in the current context.

State. We start path reasoning from the starting entity e_0 in C. The initial state $s_0 \in \mathcal{S}$ is $s_0 = \{u, e_0\}$. We encode the H-step history of entities and relations as the observed state $s_t \in \mathcal{S}$ at step t, i.e., $s_t = \{u, e_{t-H}, \ldots, r_t, e_t\}$, whose embedding $\boldsymbol{s}_t$ is obtained by concatenating the embeddings of all members of s_t, i.e., $\boldsymbol{s}_t = \boldsymbol{u} \oplus \boldsymbol{e}_{t-H} \ldots \oplus \boldsymbol{r}_t \oplus \boldsymbol{e}_t$, where $\boldsymbol{u}$ is preference representation, $\oplus$ is the concatenation operator. If the path length is smaller than H, we pad $\boldsymbol{s}_t$ with zeros.

Action. The action space $\mathcal{A}_t$ of the state s_t is defined as all outgoing edges of the entity e_t in the KG $\mathcal{G}$, excluding history entities and relations. $\mathcal{A}_t = \{(r, e) \mid (e_t, r, e) \in \mathcal{G}, e \notin \{e_0, \ldots, e_{t-1}\}\}$. As an option to terminate, $\mathcal{A}_t$ has a self-loop edge.

Transition. Given the current state s_t and the action chosen by the agent $a_t = (r_{t+1}, e_{t+1})$, the next state s_{t+1} is: $s_{t+1} = \mathcal{T}(s_t, a_t) = \{u, e_{t-H+1}, \ldots, r_{t+1}, e_{t+1}\}$, where $\mathcal{T} : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ refers to the state transition function.

Reward. We only give the agent terminal reward $R_{T,t}$, $R_{T,t}$ is 1 if the agent generates a path ends with the recommended items I_Y in the response Y, and 0 otherwise.

Policy Optimization. We adopt adversarial imitation learning (Zhao et al., 2020) based on the Actor-Critic method for policy optimization. Actor learns a path reasoning policy $\pi_\varphi(a_t, s_t, \mathcal{A}_t)$ which selects a "good" action a_t based on the current state s_t and the action space $\mathcal{A}_t$ to "fool" the discriminator in the path imitation. Critic estimates the value $Q_\delta(s_t, a_t)$ of each action a_t in the situation of the state s_t to guide the actor to choose a "good" action. We use two fully connected layers as the actor policy network:

$$\pi_\varphi(a_t, s_t, \mathcal{A}_t) = \eta\left(\boldsymbol{A}_t f\left(W_{\varphi,2}\left(f\left(W_{\varphi,1}\boldsymbol{s}_t\right)\right)\right)\right), \quad (1)$$

where $\boldsymbol{A}_t$ denotes that the action space is encoded by stacking the embedding of all actions in $\mathcal{A}_t$, and each embedding in $a_t \in \mathcal{A}_t$ is obtained by a lookup

layer. $\eta(\cdot)$ is the softmax function, $f(\cdot)$ is the ELU activation function, $W_{\varphi,1}$ and $W_{\varphi,2}$ are learnable. We design a critic network as:

$$Q_\delta(s_t, a_t) = \boldsymbol{a}_{\delta,t} \cdot f\left(W_{\delta,2}\left(f\left(W_{\delta,1}\boldsymbol{s}_t\right)\right)\right), \quad (2)$$

where $\boldsymbol{a}_{\delta,t}$ is the embedding of action a_t in the critic, $f(\cdot)$ is the ELU activation function, $W_{\delta,1}$ and $W_{\delta,2}$ are learnable.

Path Imitation   To guide the actor to generate a path in line with the user interest shift, we design the path imitation discriminator $\mathcal{I}_{p,\tau}$, which judges whether the path segment generated by the actor at each step t is similar to the golden interest shift path segment in current context. Given the current state s_t and action a_t, the probability that $\mathcal{I}_{p,\tau}$ outputs (s_t, a_t) conforms to the golden shift path segment:

$$\mathcal{I}_{p,\tau}(s_t, a_t) = \sigma(\boldsymbol{b}_{p,\tau}^T z\left(W_{p,\tau} z\left(\boldsymbol{s}_t \oplus \boldsymbol{a}_{p,t}\right)\right)), \quad (3)$$

where $\boldsymbol{a}_{p,t}$ is the embedding of a_t in $\mathcal{I}_{p,\tau}$, $z(\cdot)$ is tanh function and $\sigma(\cdot)$ is sigmoid function. $W_{p,\tau}$ and $\boldsymbol{b}_{p,\tau}$ are learnable. $\mathcal{I}_{p,\tau}$ is learned by minimizing the loss function L_τ:

$$L_\tau = -(\log\left(1 - \mathcal{I}_{p,\tau}(s_t, a_t)\right) + \log(\mathcal{I}_{p,\tau}(s_t^K, a_t^K))) \quad (4)$$

where s_t^K and a_t^K respectively denote the state and action of the golden shift process in the same step t. We further obtain the reward $R_{p,t}$ given by $\mathcal{I}_{p,\tau}$ to the actor at each step t:

$$R_{p,t} = \log\left(\mathcal{I}_{p,\tau}(s_t, a_t)\right) - \log\left(1 - \mathcal{I}_{p,\tau}(s_t, a_t)\right). \quad (5)$$

Here, the aggregation reward obtained by the agent is $R_t = \alpha R_{p,t} + (1 - \alpha)R_{T,t}$. where $\alpha \in [0, 1]$. In the final joint training with the conversation module, the agent receives two other rewards $R_{k,t}$ and $R_{s,t}$ from the conversation module. Given $Q_\delta(s_t, a_t)$, the actor and critic is updated jointly by minimizing the loss function $L_{\varphi,\delta}$:

$$L_{\varphi,\delta} = -\mathrm{E}_{a \sim \pi_\varphi}Q_\delta(s_t, a) + \left(Q_\delta(s_t, a_t) - G_t\right)^2, \quad (6)$$

where the total cumulative reward $G_t = R_t + \mathrm{E}_{a \sim \pi_\varphi}Q_\delta(s_{t+1}, a)$ is calculated by Bellman equation (Bellman, 2013). The actor, critic and path imitation discriminator is jointly optimized by minimizing $L_{REC} = L_{\varphi,\delta} + L_\tau$.

Beam Search of Recommendation Paths   After agent pre-training, we adopt beam search to generate candidate recommendation paths. Sorted by the

probability of leading to accurate recommendation, top N_p paths are tokenized into a statement containing entity and relation words, which are provided to the conversation module as alignment signals.

4.3 Recommendation-aware Conversation Module

Encoder The conversation context C, the response Y, the utterance U and the tokenized recommendation paths $\mathcal{P}$ are encoded by the context encoder, the knowledge encoder and the semantic encoder, respectively, based on Bi-RNN. Given the input sequence $X = (x_1, \ldots, x_N)$, the forward and the backward RNN respectively generate hidden states h_t^f and h_t^b for each x_t, which are concatenated to form the overall hidden state h_t:

$$h_t = [h_t^f; h_t^b] = [\overrightarrow{\text{GRU}}(x_t, h_{t-1}^f); \overleftarrow{\text{GRU}}(x_t, h_{t+1}^b)], \tag{7}$$

where $[;]$ is the concatenation operation. We denote the hidden states of all time steps as $H = (h_1, h_2, \ldots, h_N)$, $o = \left[h_N^f; h_1^b\right]$ as the final hidden state. For all input sources, we obtain H_C, H_Y, H_U, $\{H_{\mathcal{P},i}\}_{i=1}^{N_p}$ and o_C, o_Y, o_U, $\{o_{\mathcal{P},i}\}_{i=1}^{N_p}$.

Knowledge Imitation To refine the distribution of tokenized recommendation paths of leading to accurate recommendation with proper explanation, knowledge imitation makes the tokenized recommendation paths imitate the human response, which often contains correct recommendation destination without strong explanation. Given the encoded conversation context o_C and $o_{\mathcal{P}} = \{o_{\mathcal{P},i}\}_{i=1}^{N_p}$ of the encoding of tokenized recommendation paths, the o_C server as the prior information. We first obtain the prior path weight distribution by the similarity between o_C and each path p_i:

$$P(p_i \mid C) = \frac{\exp\left(o_{\mathcal{P},i} \cdot z\left(W_{K,C} o_C\right)\right)}{\sum_{j=1}^{N_p} \exp\left(o_{\mathcal{P},j} \cdot z\left(W_{K,C} o_C\right)\right)}, \tag{8}$$

where $z(\cdot)$ is the tanh function, $W_{K,C}$ is learnable.

Since the recommendation paths contain the predicated interest in response, the prior information is insufficient to calculate the recommendation paths distribution. Therefore, the imitation also involves the human response as posterior information to obtain the posterior distribution of the paths.

$$P(p_i \mid Y) = \frac{\exp\left(o_{\mathcal{P},i} \cdot z\left(W_{K,Y} o_Y\right)\right)}{\sum_{j=1}^{N_p} \exp\left(o_{\mathcal{P},j} \cdot z\left(W_{K,Y} o_Y\right)\right)}, \tag{9}$$

where $W_{K,Y}$ is learnable. We use Kullback-Leibler divergence loss L_{KL} to make $P(p_i \mid C)$ imitates $P(p_i \mid Y)$, and BOW Loss L_{BOW} to enforce the relevancy between recommendation paths distribution and response (Lian et al., 2019).

Semantic Imitation To refine the semantic encoding of tokenized recommendation paths, semantic imitation makes tokenized recommendation paths imitates the golden utterance statement of correct recommendation and coherent explanation. Given the encoded conversation context o_C and the hidden state $o_{\mathcal{P}} = \{o_{\mathcal{P},i}\}_{i=1}^{N_p}$ of the tokenized recommendation paths encoding, we apply attention (Bahdanau et al., 2015) to the $o_{\mathcal{P}}$ to obtain the context-based path aggregation representation $o_{S,\mathcal{P}} = \text{Attention}\left(o_{\mathcal{P}}, z\left(W_{S,\mathcal{P}} o_C\right)\right)$, where $z(\cdot)$ is the tanh function, $W_{S,\mathcal{P}}$ is the parameter matrix.

To make $o_{S,\mathcal{P}}$ and the semantic o_U of the encoded golden interest shift path behave similarly, we adopt the Mutual Information Maximization (Cao et al., 2021), which forces the learned context-based aggregation representation to equip with the semantic of the golden utterance statement via maximizing the mutual information between $o_{S,\mathcal{P}}$ and o_U. We use binary cross-entropy loss as the mutual information estimator. The learning objective is:

$$L_{BCE} = -\frac{1}{|\mathbb{P}| + |\mathbb{N}|} \Big(\sum_{\mathbb{P}} \log \mathcal{I}_{S,\phi}\left(o_{S,\mathcal{P}}, o_U\right) + \\ \sum_{\mathbb{N}} \log\left(1 - \mathcal{I}_{S,\phi}\left(\widetilde{o_{S,\mathcal{P}}}, o_U\right)\right) \Big), \tag{10}$$

where $\mathbb{P}$ and $\mathbb{N}$ represent the set of positive and negative samples, respectively. $\widetilde{o_{S,\mathcal{P}}}$ is the random sampled negative sample's encoding. $\mathcal{I}_{S,\phi}$ is a semantic imitation discriminator to score $o_{S,\mathcal{P}}$ and o_U via a bilinear mapping function:

$$\mathcal{I}_{S,\phi}\left(o_{S,\mathcal{P}}, o_U\right) = \sigma\left((o_{S,\mathcal{P}})^T W_{S,\phi} o_U\right), \tag{11}$$

where $\sigma(\cdot)$ is the sigmoid function, $W_{S,\phi}$ is the parameter matrix. For the response generation, an MLP layer merges the learned semantic o_S into the hidden state o_C of the conversation context as the initial hidden state of the decoder, where $o_S = o_U$ if U is available, otherwise o_S is the learned $o_{S,\mathcal{P}}$. In the inference stage, U is unknown.

Path-aware Response Generation We employ GRU to integrate context and path information to generate a response. Given the decoder state h_t, the output states H_C and $\{H_{\mathcal{P},i}\}_{i=1}^{N_p}$ of the context encoder and knowledge encoder, we apply

attention to the H_C at decoder step t: $d_t^C, v_t^C = \text{Attention}(H_C, h_t)$, where d_t^C is the attention distribution of each token in the context C, v_t^C is the aggregation vector of C. $\left\{d_t^{\mathcal{P},i}\right\}_{i=1}^{N_p}$ and $\left\{v_t^{\mathcal{P},i}\right\}_{i=1}^{N_p}$ are obtained for $H_{\mathcal{P},i}$ of each path p_i. We obtain the overall path representation $v_t^{\mathcal{P}} = \sum_{i=1}^{N_p} \mu_{\mathcal{P},i} \cdot v_t^{\mathcal{P},i}$, where $\mu_{\mathcal{P},i} = P(p_i \mid Y)$ in training, and $\mu_{\mathcal{P},i} = P(p_i \mid C)$ in inference. To reduce the impact of inaccurate recommended paths, we design a fusion gate g_t to determine the contribution of $v_t^{\mathcal{P}}$ to the fusion information v_t:

$$g_t = \sigma\left(W_g\left[d_t^C; v_t^{\mathcal{P}}\right]\right), v_t = g_t d_t^C + (1 - g_t) v_t^{\mathcal{P}}, \quad (12)$$

where W_g is learnable. Hence. the decoder updates its state as: $h_{t+1} = \text{GRU}(h_t, [y_t; v_t])$, where y_t is the embedding of predicted word at time step t. h_t and v_t are also used to obtain the generation probability $P_{vocab}(w_t)$ over the vocabulary at time step t, formalized as $P_{vocab}(w_t) = \rho([h_t; v_t])$, where $\rho(\cdot)$ is a two-layer MLP with a softmax function. Furthermore, we adopt a pointer copy mechanism to copy tokens from the tokenized recommendation paths $\mathcal{P}$, which ensures that the logical knowledge in the path can be copied to enrich the explanation in the response. At time step t, the probability of copying tokens from $\mathcal{P}$ is a weighted sum of copying tokens from all paths over the path distribution:

$$P_{\mathcal{P}}(w_t) = \sum_{i=1}^{N_p} \mu_{\mathcal{P},i} \cdot \sum_{\{j : p_i^j = w_t\}} d_{t,j}^{\mathcal{P},i}, \quad (13)$$

where p_i^j is the token in the path p_i, $d_{t,j}^{\mathcal{P},i}$ is the attention weights of the jth token in p_i. We use a pointer generation probability ξ_t^{gen} (See et al., 2017) to obtain the overall probability distribution:

$$P(w_t) = \xi_t^{gen} P_{vocab}(w_t) + (1 - \xi_t^{gen}) P_{\mathcal{P}}(w_t), \quad (14)$$

where $\xi_t^{gen} = \sigma(W_{gen}[y_{t-1}; h_t; v_t])$, W_{gen} is learnable. When training the conversation module, we use additional NLL Loss to quantify the difference between the golden and generated response:

$$L_{NLL} = -\frac{1}{|Y|} \sum_{t=1}^{|Y|} \log(P(y_t \mid y_{<t}, C, \mathcal{P})). \quad (15)$$

In summary, the conversation module is jointly optimized by minimizing the joint loss $L_{GEN} = L_{KL} + L_{BOW} + L_{BCE} + L_{NLL}$.

4.4 Bidirectional Improvement of Two Modules

After training recommendation and conversation modules, we conduct a bidirectional joint training. The conversation module provides the recommendation module with the rewards $R_{k,t}$ and $R_{s,t}$ from the knowledge imitation and semantic imitation, respectively. $R_{k,t}$ is the knowledge consistency between recommendation paths and human response. If path $p \in \mathcal{P}$, $R_{k,t} = \log(\mu_{\mathcal{P},i}) + \log(1 - \mu_{\mathcal{P},i})$, where i is the index of p in $\mathcal{P}$, otherwise $R_{k,t} = 0$. $R_{s,t}$ is the semantic similarity between the path segment generated at step t and the golden utterance. $R_{s,t} = \log(\mathcal{I}_{S,\phi}(o_{p,t}, o_U)) + \log(1 - \mathcal{I}_{S,\phi}(o_{p,t}, o_U))$, where $o_{p,t}$ is the hidden state of the tokenized path segment encoded by the semantic encoder. The aggregation reward is $R_t = \alpha R_{p,t} + \beta R_{k,t} + \gamma R_{s,t} + (1 - \alpha - \beta - \gamma) R_{T,t}$, where $\alpha + \beta + \gamma \in [0, 1]$. If the path is shorter than the maximum reasoning length, $\beta = 0$.

The recommendation module provides the conversation module with optimized recommendation paths to guide the response generation. In this way, the bidirectional joint training optimizes the alignment between recommendation and conversation and promotes the overall performance of CRS.

5 Experiments

5.1 Experiment Setup

Dataset We did experiments on OpenDialKG (Moon et al., 2019), a dialog↔KG parallel corpus for CRS, where the mentions of KG entities and their factual connections in a dialog are annotated. The user interest shift path is extracted from context-response pairs, where its start entity is in the context, and destination entity is in the response. Each path is tokenized into an utterance statement that weaves together the entities and relations mentioned in the conversation. More details on data and experiments are in Appendix A and B.

Models for Comparison (1) *TextCNN* (Kim, 2014) is CNN-based recommendation model. (2) *Trans* (Vaswani et al., 2017) is a Transformer-based response generation model. (3) *KBRD* (Chen et al., 2019) is a knowledge-based CRS that enhances user preference with a KG. (4) *KGSF* (Zhou et al., 2020) is a KG-based CRS aligning the semantic space of two KGs. (5) *RevCore* (Lu et al., 2021) is a review-enhanced CRS. (6) *CRFR* (Zhou et al., 2021) is a fragments reasoning-based CRS. (7) C^2-

Models	Recall@1	Recall@10	Recall@25
TextCNN	0.059	0.177	0.235
KBRD	0.104	0.407	0.490
KGSF	0.119	0.436	0.523
RevCore	0.124	0.432	0.516
CRFR	0.130	0.458	0.543
C^2-CRS	0.112	0.465	0.541
ACRG	0.185	0.490	0.629
DICR	**0.211***	**0.511***	**0.643***
w/o PI	0.203	0.497	0.635
w/o KI	0.201	0.494	0.632
w/o SI	0.200	0.500	0.631

Table 1: Overall evaluation on recommendation. w/o refers to removing the component from DICR. "$*$" indicates the statistical significance for $p < 0.001$ compared with the best baseline (t-test with p-value < 0.001).

Models	Bleu-1	Bleu-2	Dist-1	Dist-2	F1
Trans	0.388	0.309	0.027	0.103	0.050
KBRD	0.408	0.324	0.055	0.162	0.108
KGSF	0.416	0.330	0.062	0.203	0.123
RevCore	0.409	0.323	0.057	0.195	0.112
CRFR	0.421	0.334	0.064	0.208	0.135
C^2-CRS	0.417	0.331	**0.065**	**0.209**	0.145
ACRG	0.422	0.326	0.054	0.161	0.270
DICR	**0.478***	**0.366***	0.059	0.183	**0.319***
w/o PI	0.467	0.357	0.057	0.177	0.308
w/o KI	0.464	0.357	0.055	0.171	0.293
w/o SI	0.466	0.358	0.057	0.175	0.302
w/o FG	0.458	0.352	0.056	0.168	0.305
w/o BI	0.439	0.338	0.058	0.174	0.288

Table 2: Overall evaluation on conversation. w/o and "$*$" have the same meaning as those in Table 1.

CRS (Zhou et al., 2022b) is a contrastive learning-based CRS. (8) We design *ACRG* as a variant of DICR and removes all imitation components and rich interaction between two modules from DICR.

Implementation Details We implemented our model with Pytorch. In the recommendation module, the history length $H = 1$ and the maximum length of the reasoning path is 3. The maximum action space is 250. We trained the KG with the embedding size 128. The rewards weights are $\alpha = \gamma = 0.006$, $\beta = 0.001$. The conversation module receives $N_p = 10$ recommendation paths. All encoders and decoders have 2-layers with 800 hidden units for each layer. The word embedding is initialized with word2vec and size 300. We used the Adam optimizer (Kingma and Ba, 2015), the batch size is 32 and the learning rate is 0.0001. We trained our model with four steps. We first trained the model to minimize the L_{REC} loss, then minimized the BOW loss and BCE loss for pre-training knowledge imitation and semantic imitation components, and then minimized the L_{GEN} loss. Finally, we jointly trained the whole model.

5.2 Evaluation on Recommendation

In recommendation evaluation, we use Recall@K (K=1,10,25) indicating whether the top-k predicted items include the golden recommendation item.

Overall Evaluation As in Table 1, DICR outperforms all the baselines significantly, benefiting from using rich dual imitation signals as the rewards for the recommendation agent. Compared with the best results of CRFR and C^2-CRS, DICR achieves 62.3%, 9.9%, and 18.4% improvements on three metrics. Despite substantial progress

achieved by extra knowledge (i.e., KBRD, KGSF, RevCore, C^2-CRS) and fragments reasoning (i.e., CRFR), their performance is still inferior to ACRG indicating that black-box preference representation is a sub-optimal interest expression scheme.

Ablation Study of Dual Imitation We separately remove path imitation, knowledge imitation and semantic imitation from DICR to examine their contribution, called w/o PI, w/o KI and w/o SI. In Table 1, all imitation components contribute to the recommendation performance, with the rewards within module (PI) or across modules (KI and SI) for the reasoning policy learning. On the one hand, the conversational rewards (i.e., from KI and SI) are used as the alignment signals and guide the recommender to learn user interest shift policies. On the other hand, the dual-reward-reinforced (i.e., from PI, KI and SI) recommendation paths serve as the alignment signals in turn improve the conversation by promoting the positive cyclic learning of bidirectional interaction with dual imitation.

5.3 Evaluation on Conversation

Overall Evaluation To evaluate the overall performance of the conversation, we use BLEU-1/2 and Distinct-1/2 (Dist-1/2) to evaluate the quality and diversity of the generated responses. F1 is the F1-score measuring how well the responses contain golden knowledge. In Table 2, DICR outperforms all baselines significantly in most metrics. DICR achieves 13.3%, 9.6% improvements on Bleu-1/2 compared to the best baselines, which supports the effectiveness of our method, i.e., aligning the recommendation reasoning and conversation process. DICR achieves the best result on F1, demonstrating that explicit recommendation paths (i.e., DICR, ACRG) are superior to implicit embedding seman-

Models	Hit	G-Inter	G-Inner	P-Inter	P-Inner
KBRD	0.250	0.639	0.194	-	-
KGSF	0.261	0.662	0.201	-	-
RevCore	0.245	0.670	0.222	-	-
CRFR	0.293	0.691	0.247	0.485	0.134
C^2-CRS	0.310	0.695	0.319	-	-
ACRG	0.318	0.548	0.288	0.527	0.279
DICR	**0.426***	**0.720***	**0.348***	**0.689***	**0.330***
w/o PI	0.405	0.685	0.320	0.657	0.309
w/o KI	0.386	0.680	0.342	0.651	0.327
w/o SI	0.403	0.687	0.327	0.663	0.319
w/o FG	0.385	0.667	0.263	0.641	0.256
w/o BI	0.354	0.592	0.298	0.570	0.287

Table 3: Evaluation on hit and explainability. w/o and "∗" have the same meaning as those in Table 1.

Models	Flu.	Coher.	Inform.	Explain.
Trans	1.644	1.302	1.002	0.750
KBRD	1.688	1.328	1.235	0.875
KGSF	1.665	1.351	1.241	0.895
RevCore	1.692	1.383	1.248	0.912
CRFR	1.696	1.386	1.262	0.940
C^2-CRS	1.673	1.342	1.234	0.900
ACRG	1.790	1.608	1.045	0.950
DICR	**1.815***	**1.633***	**1.300***	**1.058***
kappa	0.566	0.518	0.521	0.490

Table 4: Human evaluation on conversation. "Flu.", "Coher.", "Inform.", "Explain." denote fluency, coherence, informativeness and explainability. The agreement ratio $kappa \in [0.41, 0.6]$ denotes the moderate agreement. "∗" indicates t-test with p-value < 0.05.

tics (i.e., baselines except for ACRG) in guiding the generation of knowledge-rich responses.

Hit and Explainability   Accurate recommendations with coherent explanations is one of our main contributions. We propose **"Hit"** to measure the recommendation success rate in conversation. Hit is the hit rate at which recommended items in the golden response are included in the generated response. The explainability of the response is evaluated by logically linked entity pairs which are necessary for coherent explanation. Specifically, **"Inter"** counts the entity links between context and response, which evaluates contextually coherent explanation across context and response. **"Inner"** counts the entity links within the response, which evaluates self-consistent explanation in response. **"G"** counts the entity links that can be matched in KG and thus evaluates the explanations with global KG knowledge. **"P"** counts the entity links that can be matched in the recommendation paths in KG and thus evaluates how well the recommendation paths support the explanation generation. Finally, we have four combined indicators, **"G-Inter, G-Inner, P-Inter, P-Inner"**, e.g., **"G-Inter"** evaluates the coherent explanation according to KG knowledge.

In Table 3, DICR outperforms all baselines and obtains a significant improvement on Hit and explainability. First, DICR achieves 34% improvements on Hit compared to the best ACRG, which verifies the effectiveness of the dual imitation mechanism for aligning the consistent behavior of the recommendation reasoning and conversation process. Second, DICR obtains 3.6% and 9.1% gains on G-Inter and G-Inner compared to the best C^2-CRS, which shows that DICR prefers to generate logically coherent explanations within responses and across context and response. Third, DICR improves 30.7% and 18.3% on P-Inter and P-Inner

compared to the best ACRG, which indicates that the conversation side of the dual imitation (i.e., KI and SI) can effectively identify and integrate recommendation paths for response generation.

Human Evaluation   In human evaluation, we randomly sampled 200 contexts. Each context is associated with eight responses from eight comparison models, respectively. Six well-educated annotators evaluate each response with four scores: fluency, coherence, informativeness, and explainability. Fluency and coherence evaluate the language quality of responses. Informativeness evaluates whether the response incorporates rich knowledge. Explainability evaluates whether the response explains the reason for the recommended item. The scores are set to $\{0, 1, 2\}$. The model name is masked during the evaluation for a fair comparison. The Fleiss' kappa (Fleiss, 1971) measures the agreement among the annotators. As the results in Table 4, the superior of DICR on all indicators support the observations from automatic evaluations.

Ablation Study   In Tables 2 and 3: (1) We separately remove the path imitation, knowledge imitation, and semantic imitation to examine their contribution, namely w/o PI, w/o KI, and w/o SI, respectively. In the results, the path imitation mainly benefits the inner coherent of explainability ("G/P-Inner"), which verifies its designed advantage to indirectly guide the explanation logic by accurate recommendation paths. The knowledge imitation mainly benefits recommendation hit ("Hit"), inter coherent of recommendation ("G/P-Inter") and distinct of responses ("Dist-1/2"), which verifies its designed advantage to refine the distribution of recommendation paths for accurate recommendation and to encourage diverse explanations in response. The semantic imitation also mainly benefits inner

Cases		Dialog-1	Dialog-2
Context		• Could you recommend movies by the director **Martin Campbell**?	• I love adventure films. Do you know any? • Sure, Tower Heist, Touching the void, Total Recall and Tomorrow Never Dies. Which one do you like? • **Tower Heist** is great who is that starring again?
Response	**DICR**	• **Sure, Martin Campbell directed the movie GoldenEye, which is a thriller. Do you like that genre?**	**Nina Arianda is in it. She also starred in Higher Ground.**
	Trans	• Yes, he directed The Matrix. Have you seen it ?	• Yes, I like the actors. Can you tell me more about the movie?
	KBRD	• Sure, he starred in the GoldenEye and the World. Are you interested in films?	• I don't know that. But I know that she starred in The Golden Compass.
	KGSF	• Sure, he directed the thriller and stars Daniel Craig Monahan as well. Have you seen either?	• I like her in Your Highness. She starred in Titan A.E. Have you seen them?
	RevCore	• Sure, he directed The Green Mile.	• Yes, he did. He also directed The Croods. Have you seen that one?
	CRFR	• Sure, he has directed a lot of movies like GoldenEye, and also directed Casino Royale.	• I love her acting in The Imaginarium of Doctor Parnassus! Are you a fan of Lily Cole?
	C²-CRS	• Martin Campbell starred in GoldenEye, which also Martin Campbell. Have you seen that one?	• Yes, she starred in The Tree of Life and An Bits of the Year.
	ACRG	• Sure! He starred in GoldenEye. Have you seen it?	• Yes, it is a great movie. It also stars Nina Arianda.

Zero-hop interest · One-hop interest · Two-hop interest · → Correct relation · ⊗ Incorrect relation

Figure 3: Cases generated by different models, indicting multi-hop entities and correct/incorrect relations.

coherent of explainability and is more important to inter coherent of recommendation than path imitation, which verifies its designed advantages to improve the semantic of response by promoting inner and inter coherent. (2) We remove the fusion gate, namely w/o FG. The results show that the dynamic information fusion mechanism achieves an impressive enhancement. (3) We remove the bidirectional improvement in the training, namely w/o BI. The results indicate that the tightly information interaction between recommendation and conversation with the alignment signals of dual imitation as the bridge is crucial to the overall performance.

5.4 Case Study

In Figure 3, two cases from eight models are selected, among which DICR has two advantages: (1) The items recommended by DICR is more accurate and likely to have explicit multi-hop relation with the items mentioned by the user, being consistent to "Hit" and "G/P-Inter" in Table 3, e.g., in Dialog-2, "Higher Ground" in response and "Tower Heist" in context share the actor "Nina Arianda." This is evidence of improving recommendation by tracking user interest shift in conversation, which mainly benefits from the path imitation and the knowledge imitation, as verified by ablation study in Table 3; (2) DICR naturally tells the items' relation as an explanation, being consistent to "G/P-Inner" in Table 3 and "Explain." in Table 4, e.g., in Dialog-1, the director "Martin Campbell", the movie "GoldenEye" and the genre "thriller" derive from a recommendation path with coherent relation. This is evidence of

improving conversation by involving recommendation path as an explanation, which mainly benefits from the semantic imitation, as verified in Table 3.

6 Conclusions

We propose DICR, which adopts the dual imitation to align CRS's recommendation and conversation behavior explicitly. Using recommendation paths and conversational rewards as alignment signals for tight interaction between recommendation and conversation, DICR achieves accurate recommendations and coherent explanations in generated responses. The effectiveness of DICR is verified by designed novel explainability evaluations together with human and existing automatic metrics.

Limitations

We discuss two main limitations of this work which can be further studied in future work. The first one is the reliance of explicit knowledge in knowledge graph. Although using knowledge graph is a common advantage of most current CRS studies, and explicit relations between entities leads to effective and reliable reasoning for recommendation, there are still a large amount of implicit knowledge in unstructured resources which cannot be extracted as explicit triplet, e.g., the multidimensional similarity between entities, but can be further extra supplement to dialog context.

The second one is the task of next-turn recommendation. As the main contribution of this work, although the modeling of user interest shift signifi-

cantly improve the performance of making recommendation in next-turn response, the user interest shift modeling can also naturally help us to guide the user interests towards proper recommendation through smooth and persuasive multi-turn conversation with users. To address this limitation, in the future, we will extend the idea to align the KG-based reasoning and conversation process towards long-term global goal instead of local target.

Ethics Consideration

All models in this paper are trained on public corpus. The OpenDialKG (Moon et al., 2019) dataset do not contain personal information and unethical language. We also ensure the anonymization of the human evaluation. We belieive that this work honors the ethical code of EMNLP.

Acknowledgements

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

R Bellman. 2013. Dynamic programming, courier corporation. *New York, NY*, 707.

Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 2787–2795.

Jiangxia Cao, Xixun Lin, Shu Guo, Luchen Liu, Tingwen Liu, and Bin Wang. 2021. Bipartite graph embedding via mutual information maximization. In *WSDM '21, The Fourteenth ACM International Conference on Web Search and Data Mining, Virtual Event, Israel, March 8-12, 2021*, pages 635–643. ACM.

Qibin Chen, Junyang Lin, Yichang Zhang, Ming Ding, Yukuo Cen, Hongxia Yang, and Jie Tang. 2019. To-wards 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.

Yang Deng, Yaliang Li, Fei Sun, Bolin Ding, and Wai Lam. 2021. Unified conversational recommendation policy learning via graph-based reinforcement learning. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, pages 1431–1441. ACM.

Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.

Chongming Gao, Wenqiang Lei, Xiangnan He, Maarten de Rijke, and Tat-Seng Chua. 2021. Advances and challenges in conversational recommender systems: A survey. *AI Open*, 2:100–126.

Shirley Anugrah Hayati, Dongyeop Kang, Qingxiaoyang 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.

Jonathan Ho and Stefano Ermon. 2016. Generative adversarial imitation learning. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 4565–4573.

Dietmar Jannach, Ahtsham Manzoor, Wanling Cai, and Li Chen. 2021. A survey on conversational recommender systems. *ACM Comput. Surv.*, 54(5):105:1–105:36.

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.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Ivica Kostric, Krisztian Balog, and Filip Radlinski. 2021. Soliciting user preferences in conversational recommender systems via usage-related questions. In *RecSys '21: Fifteenth ACM Conference on Recommender Systems, Amsterdam, The Netherlands,*

27 September 2021 - 1 October 2021, pages 724–729. ACM.

Wenqiang Lei, Gangyi Zhang, Xiangnan He, Yisong Miao, Xiang Wang, Liang Chen, and Tat-Seng Chua. 2020. 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.

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.

Shuokai Li, Ruobing Xie, Yongchun Zhu, Xiang Ao, Fuzhen Zhuang, and Qing He. 2022. User-centric conversational recommendation with multi-aspect user modeling. In *Proceedings of the 45nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. Association for Computing Machinery.

Rongzhong Lian, Min Xie, Fan Wang, Jinhua Peng, and Hua Wu. 2019. Learning to select knowledge for response generation in dialog systems. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 5081–5087. ijcai.org.

Zujie Liang, Huang Hu, Can Xu, Jian Miao, Yingying He, Yining Chen, Xiubo Geng, Fan Liang, and Daxin Jiang. 2021. Learning neural templates for recommender dialogue system. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 7821–7833. Association for Computational Linguistics.

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.

Yu Lu, Junwei Bao, Yan Song, Zichen Ma, Shuguang Cui, Youzheng Wu, and Xiaodong He. 2021. Revcore: Review-augmented conversational recommendation. In *Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021*, volume ACL/IJCNLP 2021 of *Findings of ACL*, pages 1161–1173. Association for Computational Linguistics.

Wenchang Ma, Ryuichi Takanobu, and Minlie Huang. 2021. Cr-walker: Tree-structured graph reasoning and dialog acts for conversational recommendation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 1839–1851. Association for Computational Linguistics.

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.

Xuhui Ren, Hongzhi Yin, Tong Chen, Hao Wang, Zi Huang, and Kai Zheng. 2021. Learning to ask appropriate questions in conversational recommendation. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, pages 808–817. ACM.

Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 1073–1083. Association for Computational Linguistics.

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.

Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, and Tat-Seng Chua. 2019. Explainable reasoning over knowledge graphs for recommendation. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 5329–5336. AAAI Press.

Yikun Xian, Zuohui Fu, S. Muthukrishnan, Gerard de Melo, and Yongfeng Zhang. 2019. Reinforcement knowledge graph reasoning for explainable recommendation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019*, pages 285–294. ACM.

Zhitong Yang, Bo Wang, Jinfeng Zhou, Yue Tan, Dongming Zhao, Kun Huang, Ruifang He, and Yuexian Hou. 2022. TopKG: Target-oriented dialog via global planning on knowledge graph. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 745–755, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Jun Zhang, Yan Yang, Chencai Chen, Liang He, and Zhou Yu. 2021. KERS: A knowledge-enhanced framework for recommendation dialog systems with multiple subgoals. In *Findings of the Association for Computational Linguistics: EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 16-20 November, 2021*, pages 1092–1101. Association for Computational Linguistics.

Tong Zhang, Yong Liu, Boyang Li, Peixiang Zhong, Chen Zhang, Hao Wang, and Chunyan Miao. 2022a. Toward knowledge-enriched conversational recommendation systems. In *Proceedings of the 4th Workshop on NLP for Conversational AI, ConvAI@ACL 2022, Dublin, Ireland, May 27, 2022*, pages 212–217. Association for Computational Linguistics.

Yiming Zhang, Lingfei Wu, Qi Shen, Yitong Pang, Zhihua Wei, Fangli Xu, Bo Long, and Jian Pei. 2022b. Multiple choice questions based multi-interest policy learning for conversational recommendation. In *WWW '22: The ACM Web Conference 2022, Virtual Event, Lyon, France, April 25 - 29, 2022*, pages 2153–2162. ACM.

Kangzhi Zhao, Xiting Wang, Yuren Zhang, Li Zhao, Zheng Liu, Chunxiao Xing, and Xing Xie. 2020. Leveraging demonstrations for reinforcement recommendation reasoning over knowledge graphs. 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 239–248. ACM.

Jinfeng Zhou, Bo Wang, Ruifang He, and Yuexian Hou. 2021. CRFR: improving conversational recommender systems via flexible fragments reasoning on knowledge graphs. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, pages 4324–4334. Association for Computational Linguistics.

Jinfeng Zhou, Bo Wang, Zhitong Yang, Dongming Zhao, Kun Huang, Ruifang He, and Yuexian Hou. 2022a. CR-GIS: Improving conversational recommendation via goal-aware interest sequence modeling. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 400–411, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Kun Zhou, Wayne Xin Zhao, Shuqing Bian, Yuanhang Zhou, Ji-Rong Wen, and Jingsong Yu. 2020. 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.

Yuanhang Zhou, Kun Zhou, Wayne Xin Zhao, Cheng Wang, Peng Jiang, and He Hu. 2022b. C2-CRS: coarse-to-fine contrastive learning for conversational recommender system. *CoRR*, abs/2201.02732.

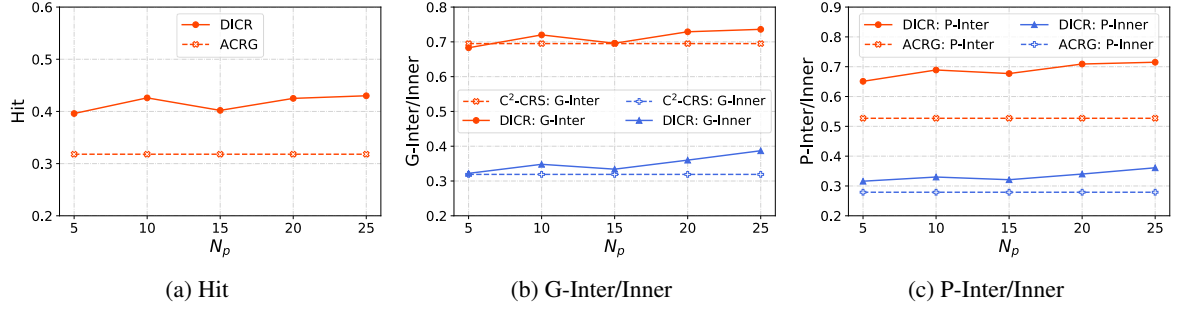| | (a) Hit | (b) G-Inter/Inner | (c) P-Inter/Inner |

Figure 4: The influence of the number of recommendation paths on Hit and Explainability. As the number of recommendation paths increases, DICR improves on Hit and Explainability metrics and outperforms the best baseline on most cases.

Corpus Info.	#Domain	Movie,Book
	#Dialogues	15,673
	#Turns	91,209
	#Split Ratio	7:1.5:1.5
KG Info.	#Entities	100,813
	#Relations	1,358
	#Triplets	1,190,658

Table 5: Statistics of our datasets after preprocessing.

A Dataset

The statistics of OpenDialKG after preprocessing are in Table 5.

We did not employ other CRS datasets. The reason is that compared with OpenDialKG, in other datasets like REDIAL (Li et al., 2018), dialogs mention the recommended items without rich related information and tend to mention only movie names rather than an in-depth discussion on the movie preference, which is considered as the recommended explanation in this paper. As reported in the *CRFR* (Zhou et al., 2021), OpenDialKG's advantages improve the performance of *CRFR* and the compared CRS baselines in our experiments.

B Analysis of the Number of Recommendation Paths

We analyze the influence of the number of recommendation paths on Hit and explainability, and Figure 4 presents the results.

First, as shown in Figure 4(a), with the increase of the number of recommendation paths which are used as the alignment signals for the recommendation side of the dual imitation, the Hit scores slightly improve in fluctuations. This indicates that the conversation side of the dual imitation can effectively identify the golden recommendation paths and prompt the conversation process to align the

recommendation reasoning.

Second, in Figure 4(b) and 4(c), G-Inter/Inner and P-Inter/Inner both improve distinctly as the number of paths increases. This improvement is attributed to the advantage that knowledge imitation and semantic imitation endow the DICR with discerning and integrating the coherent knowledge in the recommendation paths as the recommended explanation in response. This advantage aligns the recommendation reasoning to explanation generation, which helps the model refine the discerned knowledge and display them in the generated response.