

CR-GIS: Improving Conversational Recommendation via Goal-aware Interest Sequence Modeling

Jinfeng Zhou^{1,2}, Bo Wang^{1,2*}, Zhitong Yang^{1,2}, Dongming Zhao³, Kun Huang³,
Ruifang He^{1,2}, Yuexian Hou¹

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

²State Key Laboratory of Communication Content Cognition,

²People's Daily Online, Beijing, China

³AI Lab, China Mobile Communication Group Tianjin Co., Ltd.

{jfzhou, bo_wang, yyyyyyzt}@tju.edu.cn

Abstract

Conversational recommendation systems (CRS) aim to determine a goal item by sequentially tracking users' interests through multi-turn conversation. In CRS, implicit patterns of user interest sequence guide the smooth transition of dialog utterances to the goal item. However, with the convenient explicit knowledge of knowledge graph (KG), existing KG-based CRS methods over-rely on the explicit separate KG links to model the user interests but ignore the rich goal-aware implicit interest sequence patterns in a dialog. In addition, interest sequence is also not fully used to generate smooth transited utterances. We propose CR-GIS with a parallel star framework. First, an interest-level star graph is designed to model the goal-aware implicit user interest sequence. Second, a hierarchical Star Transformer is designed to guide the multi-turn utterances generation with the interest-level star graph. Extensive experiments verify the effectiveness of CR-GIS in achieving more accurate recommended items with more fluent and coherent dialog utterances.

1 Introduction

Traditional recommendation systems often interact with users in a one-shot, one-directional manner (Jannach et al., 2021), that is, users passively receive the static recommendation list and the recommendation system lacks the ability to understand and proactively guide the dynamic shift of users' interests. Conversational Recommendation Systems (CRS) (Sun and Zhang, 2018; Li et al., 2018) solve these problems by supporting multi-turn goal-oriented (Kang et al., 2019; Zhou et al., 2020b) dialog to proactively track and guide real-time user interests shift (Gao et al., 2021).

In current studies, one most popular way of CRS is to determine an item meeting the user preference through multi-turn conversation. Regarding the

item to be determined as the goal, we regard this kind of CRS as "goal-aware" CRS. In this paper, we suppose that the performance of goal-aware CRS is highly dependent on a goal-aware sequence of user interests which is expressed in the form of logically transited utterance sequence in dialog. From the view of recommendation, in a goal-aware CRS dialog, a proper dynamic user preference at any time should be coherent to the preorder interests and towards the final goal in the interest sequence. From the view of conversation, a goal-aware CRS dialog is a smoothly transited sequence of utterances guided by the sequence of user interest. Therefore, modeling the goal-aware sequence of user interests is essential for goal-aware CRS. For example, in the "Dialog" of Figure 1, there are similarities in storyline or other aspects between neighbor entities in the interest sequence, i.e., "Shutter Island" → "Inception" → "Leonardo" → "Source Code", and between each interest entity and the recommendation goal "The Butterfly Effect".

For modeling the goal-aware interest sequence, KG-based CRS methods are widely studied using knowledge graph (KG) (Zhou et al., 2020a; Lu et al., 2021; Zhou et al., 2022) to track the interest sequence in the dialog (Zhou et al., 2021; Ma et al., 2021). Although KG's explicit logical links between interest entities greatly facilitate the modeling of interest sequence, it also leads to over-rely on KG knowledge and weakens the key role of dialogue behavior in CRS. Specifically, we note two major consequent issues: (1) From the view of recommendation, most KG-based CRS model the user preference only on the explicit separate interest links in KG and ignore the rich implicit interests sequence in dialogue which is absent in KG knowledge. For example, in the "Dialog" of Figure 1, each interest entity may have implicit semantic relations with the goal "The Butterfly Effect" beyond KG links. Furthermore, the goal of the sequence is also absent in the user preference modeling. (2)

*Corresponding author.

From the view of conversation, current KG-based CRS mainly use entities in the interest sequence to enhance the semantic generation of every single response instead of using the entire sequence to enhance the smooth transition of multiple utterances towards the recommendation goal.

To address these two issues, we propose **CR-GIS**, which jointly improves the **C**onversation and **R**ecommendation by modeling **G**oal-aware **I**nterest **S**equence in CRS. To this end, we design a novel parallel star structure with two advantages: (1) For the first issue of implicit relations modeling in goal-aware interests sequence, on the base of pre-encoded explicit knowledge from KG, we propose an interest-level star graph to encode the implicit relations between interest entities in dialog. The satellite nodes of the star graph are interest entities in the current ongoing dialog, which are sequenced by their adjacency relationship in the dialog. The key advantage of interest-level star graph is that the central star node, which connects all satellite nodes, can be fused to the recommendation goal with the Mutual Information Maximization (MIM) method. This alignment makes the interest sequence modeling goal-oriented. (2) For the second issue of smooth transition in goal-aware conversation, to be paralleled to the interest-level star graph, we design a hierarchical Star Transformer encoder (Guo et al., 2019) for words and utterances. By aligning the utterance state representation of the "utterance-level star" with the "interest-level star", we enhance the CRS to generate smoothly transitioned utterances towards the recommendation goal.

Our contributions are summarized as follows:

(1) To sufficiently model the goal-aware user interest sequence in CRS, we propose an interest-level star graph to model the implicit sequence of interest entities in dialog and make the interest sequence modeling be aware of the recommendation goal with a goal-oriented fusion mechanism.

(2) To effectively generate the smoothly transitioned responses, we align an utterance-level star transformer to the interest-level star so as to make the responses generation follow the interest sequence and be aware of the recommendation goal.

(3) Extensive experiments conducted on OpenDialKG (Moon et al., 2019) and TG-ReDial (Zhou et al., 2020b) datasets demonstrate that our model outperforms the SOTA baseline models in successfully reaching recommendation goals through smooth transitioned utterances.

2 Related Work

Current CRS can be divided into two types. One is the recommendation-based CRS which aims to ask users questions about interests over pre-define slots (Sun and Zhang, 2018; Lei et al., 2020; Zou et al., 2020; Deng et al., 2021; Zhang et al., 2021; Kostric et al., 2021) and make responses considering users' feedback (Luo et al., 2020; Xu et al., 2021). Recommendation-based CRS mainly suffers from the inflexibility of pre-defined templates.

The other is the generation-based CRS (Li et al., 2018; Hayati et al., 2020) which understands user preferences (Moon et al., 2019; Zhou et al., 2020b; Lu et al., 2021; Zhou et al., 2022) and generates human-like responses (Liao et al., 2019; Liang et al., 2021) in line with user interests. Closely related to our work, Chen et al. (2019) and Zhou et al. (2020a) integrate KG knowledge to understand users' interest. However, they simply aggregate entities of KG in the utterance, instead of exploiting the implicit interest sequence in the conversation. Zhou et al. (2021) and Ma et al. (2021) adopt reasoning-based methods to predict the shift direction of user interest, but also limits to the explicit interest sequence in KG.

3 Problem Formalization

A KG \mathcal{G} with entity set \mathcal{E} and relation set \mathcal{R} is $\mathcal{G} = \{(e, r, e') \mid e, e' \in \mathcal{E}, r \in \mathcal{R}\}$ where (e, r, e') is a relation r from the entity e to the entity e' . Suppose we have a CRS corpus D and a KG \mathcal{G} parallel to D , in which the interest entities mentioned in D are linked to the entities in \mathcal{G} . $U = \{u_1, u_2, \dots, u_n\}$ is the conversation history, where $u_i = \{w_{i,1}, w_{i,2}, \dots, w_{i,m}\}$ is the word token sequence in the i -th utterance. $S = \{s_1, s_2, \dots, s_k\}$ is the interest entity in each utterance of U , and $s_i \in S$ is linked to \mathcal{G} , i.e., $s_i \in \mathcal{E}$. In a response $Y = \{y_1, y_2, \dots, y_m\}$, a recommendation goal entity set $G = \{g_1, g_2, \dots, g_t\}$ is identified in advance, where $g_i \in G$ is an entity linking to \mathcal{G} , i.e., $g_i \in \mathcal{E}$. n, m, k, t represent the length of the historical utterance sequence, the length of the token sequence, the length of the interest entity sequence and the number of recommendation goals, respectively. Our task is to learn a recommendation model and a response generation model $P(Y|U, S, G)$ with the D and \mathcal{G} . The former model the implicit user interests sequence through S and G and help the latter to generate responses that smoothly progress to the recommendation goal G .

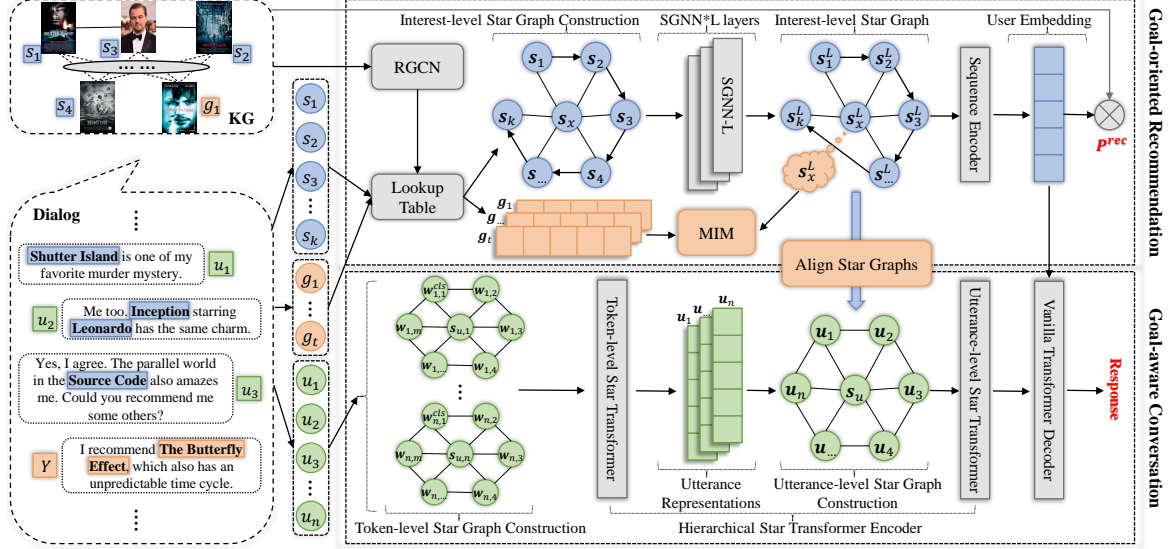


Figure 1: The architecture of the proposed CR-GIS model.

4 Approach

4.1 Architecture Overview

As shown in Figure 1, proposed CR-GIS contains five parts: (1) Explicit-Implicit relations encoder. The explicit KG relations encoder adopts R-GCN to learn the KG-based representation of entities. The interest entities sequence in the conversation is constructed into an interest-level star graph, in which the implicit user interest sequence is learned by the implicit dialog relations encoder employing a star graph neural network (SGNN). (2) The goal-oriented fusion module adopts mutual information maximization (MIM) to bridge the goal G and the star node of SGNN, which makes the user interest sequence modeling goal-aware. (3) The goal-aware recommendation module obtains the user’s interests representation with a sequence encoder mining the interest sequence. (4) The goal-aware response generation module uses hierarchical Star Transformer to encode the token-level and utterance-level star graph from conversation history. (5) We align the interest-level star graph to the utterance-level star graph to improve the goal-aware transition of multi-turn response generation.

4.2 Explicit-Implicit Relations Encoder

Explicit KG Relations Encoder We use R-GCN (Schlichtkrull et al., 2018) to encode explicit KG relations and get the entity embedding matrix E .

Implicit Dialog Relations Encoder With embedding matrix E of KG entities, we adopt star graph neural network (SGNN) (Pan et al., 2020) to encode

the implicit semantic links between the entities in the interest sequence in dialog. Given the interest sequence $S = \{s_1, s_2, \dots, s_k\}$ in dialog context, we construct an interest-level star graph with one star node s_x and each s_i is a satellite node. The adjacency relationship in the dialog between the satellite nodes $\{(s_i, s_{i+1}) \mid s_i, s_{i+1} \in S\}$ is maintained. Star node s_x is linked to each satellite s_i .

SGNN adopts a cyclic updating between s_x and each s_i . s_i is initialized by embedding matrix E , and s_x is initialized as the average value of all s_i . The updated representation $\hat{s}_i^{(l+1)}$ of s_i at the $(l+1)$ -th layer of the SGNN is calculated as:

$$\begin{aligned}
 z_i^{(l+1)} &= \sigma \left(\mathbf{W}_{z,1} \mathbf{a}_i^{(l+1)} + \mathbf{W}_{z,2} \mathbf{s}_i^{(l)} \right), \\
 v_i^{(l+1)} &= \sigma \left(\mathbf{W}_{v,1} \mathbf{a}_i^{(l+1)} + \mathbf{W}_{v,2} \mathbf{s}_i^{(l)} \right), \\
 \tilde{s}_i^{(l+1)} &= \rho \left(\mathbf{W}_{s,1} \mathbf{a}_i^{(l+1)} + \mathbf{W}_{s,2} \left(v_i^{(l+1)} \odot \mathbf{s}_i^{(l)} \right) \right), \\
 \hat{s}_i^{(l+1)} &= \left(1 - z_i^{(l+1)} \right) \odot \mathbf{s}_i^{(l)} + z_i^{(l+1)} \odot \tilde{s}_i^{(l+1)}, \\
 \mathbf{a}_i^{(l+1)} &= \left[\mathbf{A}_i^I \left(\left(\left[\mathbf{s}_1^{(l)}; \dots; \mathbf{s}_k^{(l)} \right] \right)^T \mathbf{W}_a^I + \mathbf{b}_a^I \right); \right. \\
 &\quad \left. \mathbf{A}_i^O \left(\left(\left[\mathbf{s}_1^{(l)}; \dots; \mathbf{s}_k^{(l)} \right] \right)^T \mathbf{W}_a^O + \mathbf{b}_a^O \right) \right], \quad (1)
 \end{aligned}$$

where $\mathbf{W}_{z,1}, \mathbf{W}_{v,1}, \mathbf{W}_{s,1} \in \mathbb{R}^{d_e \times 2d_e}$ and $\mathbf{W}_{z,2}, \mathbf{W}_{v,2}, \mathbf{W}_{s,2} \in \mathbb{R}^{d_e \times d_e}$ are the learnable matrix. $\rho(\cdot)$ is the tanh function. $\sigma(\cdot)$ is the sigmoid function. \odot is the Hadamard product. $\mathbf{s}_i^{(l)}$ is the representation of s_i at the l -th layer of the SGNN. $\mathbf{a}_i^{(l+1)}$ is the information propagated by the adjacent nodes of s_i on the interest-level star graph

through the incoming matrix \mathbf{A}^I and the outgoing matrix \mathbf{A}^O proposed by GGNN (Li et al., 2016). $[\cdot; \cdot]$ is the concatenation operation. $\mathbf{A}_i^I, \mathbf{A}_i^O \in \mathbb{R}^{1 \times k}$ are the weights of the i -th row in the \mathbf{A}^I and \mathbf{A}^O , respectively. $\mathbf{W}_a^I, \mathbf{W}_a^O \in \mathbb{R}^{d_e \times d_e}$ are the learnable matrix. $\mathbf{b}_a^I, \mathbf{b}_a^O \in \mathbb{R}^{d_e}$ are the bias vector. The information injected from the s_x is controlled by the self-attention mechanism to calculate the similarity $\gamma_i^{(l+1)}$ between each s_i and s_x . Furthermore, the representation of s_i at the $(l+1)$ -th layer is $\mathbf{s}_i^{(l+1)} = (1 - \gamma_i^{(l+1)}) \hat{\mathbf{s}}_i^{(l+1)} + \gamma_i^{(l+1)} \mathbf{s}_x^{(l)}$. The representation $\mathbf{s}_x^{(l+1)}$ of s_x at the $(l+1)$ -th layer is obtained by aggregating all s_i at the $(l+1)$ -th layer with attention mechanism. In this way, s_x represents the information of entire interest sequence.

4.3 Goal-oriented Fusion Mechanism

The goal-oriented fusion mechanism is motivated by the intuition that the user interest sequence is aware of the recommendation goals in the conversation. Therefore, to connect the interest-level star graph with the the goals, we bridge the star node of SGNN and the goals with the Mutual Information Maximization (Hjelm et al., 2019). Specifically, given the set of goal entities $G = \{g_1, g_2, \dots, g_t\}$ and the representation $\mathbf{s}_x^{(l+1)}$ of the star node at the $(l+1)$ -th layer, we design a loss function by the contrastive learning that maximizes the mutual information between the goal and the star node:

$$L_{MIM}(s_x, g_i) = \mathbb{E}_{g_i \in G} [f(s_x, g_i) - \log \sum_{\tilde{g}_j \in \tilde{G}} \exp(f(s_x, \tilde{g}_j))], \quad (2)$$

where $\tilde{g}_j \in \tilde{G}$ is the negative node we randomly sampled from KG. $f(\cdot, \cdot)$ is a scoring function implemented with a bilinear mapping network:

$$f(s_x, g_i) = \sigma \left(\left(\mathbf{s}_x^{(l+1)} \right)^T \cdot \mathbf{W}_{MIM} \cdot \mathbf{g}_i \right), \quad (3)$$

where $\mathbf{W}_{MIM} \in \mathbb{R}^{d_e \times d_e}$ is the learnable matrix and $\sigma(\cdot)$ is the sigmoid function. \mathbf{g}_i is the embedding of g_i by looking up from \mathbf{E} . Here, we adopt a pre-training way to enhance the goal aware ability of the sequence by minimizing the loss L_{MIM} .

4.4 Goal-oriented Recommendation

After the multi-dimensional information is fused into the entity representation, in the output of the L -layer SGNN, we obtain the star node embedding vector \mathbf{s}_x^L and the embedding matrix $\mathbf{S}^L =$

$[\mathbf{s}_1^L; \mathbf{s}_2^L; \dots; \mathbf{s}_k^L]$ composed of entities in the interest sequence, i.e., satellite nodes. To model sequence information, we integrate a learnable position embedding matrix $\mathbf{P} \in \mathbb{R}^{k \times d_e}$ to enhance the sequence representation of interest entities. The sequence representation \mathbf{S}_P^L is obtained by summing two embedding metrics: $\mathbf{S}_P^L = \mathbf{S}^L + \mathbf{P}$. In addition, the star node, as an intermediary between the sequence and the goal, contains personalized information of the goal. Therefore, we combine the position-enhanced interest sequence \mathbf{S}_P^L and the star node \mathbf{s}_x^L as the goal-enhanced sequence representation $\mathbf{S}_{P,x}^L = [\mathbf{S}_P^L; \mathbf{s}_x^L]$.

To better capture the preference expressed in the interest sequence, we propose a sequence encoder, which is composed of a multi-head self-attention layer (*MHA*) and a point-wise feed-forward network. The *MHA* extracts the information of different representation subspaces, which is defined as:

$$\begin{aligned} \mathbf{S}_F &= \text{MHA}(\mathbf{F}) = [\text{head}_1; \dots; \text{head}_h] \mathbf{W}^O, \\ \text{head}_i &= \text{Attention} \left(\mathbf{F} \mathbf{W}_i^Q, \mathbf{F} \mathbf{W}_i^K, \mathbf{F} \mathbf{W}_i^V \right), \\ \text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) &= \text{softmax} \left(\frac{\mathbf{Q} \mathbf{K}^T}{\sqrt{d_e/h}} \right) \mathbf{V}, \end{aligned} \quad (4)$$

where $\mathbf{F} = \mathbf{S}_{P,x}^L$ is the input of MHA. The projection matrix $\mathbf{W}_i^Q \in \mathbb{R}^{d_e \times d_h}$, $\mathbf{W}_i^K \in \mathbb{R}^{d_e \times d_h}$, $\mathbf{W}_i^V \in \mathbb{R}^{d_e \times d_h}$ and $\mathbf{W}^O \in \mathbb{R}^{d_e \times d_e}$ are the learnable parameters for each attention head. $d_h = d_e/h$ is the dimension of attention heads. $\mathbf{Q} = \mathbf{F} \mathbf{W}_i^Q$, $\mathbf{K} = \mathbf{F} \mathbf{W}_i^K$ and $\mathbf{V} = \mathbf{F} \mathbf{W}_i^V$ are query, key and value, respectively. $\sqrt{d_e/h}$ is the scale factor to avoid large inner product values.

Towards a nonlinear sequence encoder, we use a point-wise feed-forward network (FFN):

$$\begin{aligned} \mathbf{F} &= \left[\text{FFN}(\mathbf{S}_{F,1})^T; \dots; \text{FFN}(\mathbf{S}_{F,k})^T \right], \quad (5) \\ \text{FFN}(\mathbf{x}) &= \max(0, \mathbf{x} \mathbf{W}_{F,1} + \mathbf{b}_{F,1}) \mathbf{W}_{F,2} + \mathbf{b}_{F,2}, \end{aligned}$$

where $\mathbf{W}_{F,1}, \mathbf{W}_{F,2}, \mathbf{b}_{F,1}, \mathbf{b}_{F,2}$ are trainable parameters. Note that the sequence encoder, i.e., MHA layer and FFN layer, can be multiply stacked. We take the first embedding vector of the matrix output by the sequence encoder as the user's preference representation \mathbf{p}_u in current context, i.e., $\mathbf{p}_u = \mathbf{F}_1$.

Given the learned user preference, we calculate the probability of recommending an item: $P_i^{rec} = \text{softmax}(\mathbf{p}_u^T \cdot \mathbf{e}_i)$, where \mathbf{e}_i is the learned item embedding looking up from \mathbf{E} . To train the recommendation module, we use cross-entropy as the optimization objective:

$$L_{REC} = - \sum_{i=1}^M [P_i \cdot \log(P_i^{rec}) - (1 - P_i) \cdot \log(1 - P_i^{rec})] + \alpha * L_{MIM}, \quad (6)$$

where i is the item index. α is a hyperparameter representing the weight of MIM loss.

4.5 Goal-aware Response Generation

Parallel to the interest-level star graph, we design a hierarchical Star Transformer to encode the dialog context by constructing token-level and utterance-level star graphs to capture the sequential semantic dependency between utterances. Injecting the interest entities in the goal-aware interest-level star graph into the utterance-level star graph, we promote the goal-aware ability of response generation.

Specifically, the topology of Star Transformer (Guo et al., 2019) is the same as SGNN, which is composed of a relay node (as the star node in SGNN) and n satellite nodes. Given a dialogue context $U = \{u_1, u_2, \dots, u_n\}$ with n utterances, where $u_i = \{w_{i,1}, w_{i,2}, \dots, w_{i,m}\}$ is the word token sequence of the i -th utterance. The token-level Star Transformer encoder is a token-level star graph which treats each word token as a satellite node, and the relay node acts as a virtual hub to gather and scatter information from and to the satellite nodes. It adopts a cyclic updating, in which the satellite node $h_{i,j}$ is initialized by word embedding $w_{i,j}$, i.e., $\mathbf{h}_{i,j}^{(0)} = \mathbf{w}_{i,j}$, and the relay node $s_{u,i}$ is initialized as the average of all tokens, i.e., $\mathbf{s}_{u,i}^{(0)} = \frac{1}{m} \sum_{j=1}^m \mathbf{h}_{i,j}^{(0)}$. Each token-level satellite node is updated at step t according to its adjacent nodes, including neighbor nodes $\mathbf{h}_{i,j-1}^{(t-1)}$, $\mathbf{h}_{i,j+1}^{(t-1)}$ in the text sequence, its previous state $\mathbf{h}_{i,j}^{(t-1)}$, the state of the relay node in the previous round $\mathbf{s}_{u,i}^{(t-1)}$, and its token embedding $\mathbf{w}_{i,j}$. Formally:

$$\begin{aligned} \mathbf{C}_{i,j}^{(t)} &= [\mathbf{h}_{i,j-1}^{(t-1)}; \mathbf{h}_{i,j}^{(t-1)}; \mathbf{h}_{i,j+1}^{(t-1)}; \mathbf{s}_{u,i}^{(t-1)}; \mathbf{w}_{i,j}], \\ \mathbf{h}_{i,j}^{(t)} &= \text{MHA} \left(\mathbf{h}_{i,j}^{(t-1)}, \mathbf{C}_{i,j}^{(t)}, \mathbf{C}_{i,j}^{(t)} \right), \end{aligned} \quad (7)$$

where $\mathbf{C}_{i,j}^{(t)}$ is the context of the j -th satellite node. For MHA, $\mathbf{h}_{i,j}^{(t-1)}$ is query, $\mathbf{C}_{i,j}^{(t)}$ is key and value.

The token-level relay node is updated by fusing all satellite nodes and its previous state $\mathbf{s}_{u,i}^{(t-1)}$.

$$\mathbf{s}_{u,i}^{(t)} = \text{MHA} \left(\mathbf{s}_{u,i}^{(t-1)}, \begin{bmatrix} \mathbf{s}_{u,i}^{(t-1)}; \mathbf{H}_i^{(t)} \\ \mathbf{s}_{u,i}^{(t-1)}; \mathbf{H}_i^{(t)} \end{bmatrix} \right), \quad (8)$$

where $\mathbf{H}_i^{(t)} = [\mathbf{h}_{i,1}^{(t)}; \mathbf{h}_{i,2}^{(t)}; \dots; \mathbf{h}_{i,m}^{(t)}]$. After T-step update, we merge the information of the relay node into the token-level satellite node, and obtain the hidden vector sequence of utterance u_i using $\psi(\cdot)$ which is a MHA layer with an FFN layer:

$$\begin{aligned} &[\hat{\mathbf{h}}_{i,1}, \hat{\mathbf{h}}_{i,2}, \dots, \hat{\mathbf{h}}_{i,m}] = \\ &\psi \left(\varphi \left(\begin{bmatrix} \mathbf{h}_{i,1}^{(T)}; \mathbf{h}_{i,2}^{(T)}; \dots; \mathbf{h}_{i,m}^{(T)} \\ \mathbf{s}_{u,i}^{(T)} \end{bmatrix} \right) \right), \end{aligned} \quad (9)$$

where $\varphi(\cdot)$ is an MLP layer. The token corresponding to the first hidden state $\hat{\mathbf{h}}_{i,1}$ of the hidden vector sequence is a special token $[CLS]$ used to aggregate the sequence representation and is inspired by Devlin et al. (2019). Therefore, we collect utterance representations derived from $[CLS]$, i.e., the representation of utterance $\mathbf{u}_i = \hat{\mathbf{h}}_{i,1}$, and input them into utterance-level Star Transformer encoder.

The utterance-level Star Transformer encoder constructs an utterance-level star graph using utterances as satellite nodes h_i , which is initialized by the representation of utterance, i.e., $\mathbf{h}_i^{(0)} = \mathbf{u}_i$. The relay node s_u is initialized as the average of satellite nodes, i.e., $\mathbf{s}_u^{(0)} = \frac{1}{n} \sum_{i=1}^n \mathbf{h}_i^{(0)}$. For the update of each utterance-level satellite node at step t , in addition to the information involved in the token-level node, we also inject the goal-aware interest entities information in the interest-level star graph into the updated representation of the corresponding utterance according to the "mentioned" relationship with the utterance:

$$\begin{aligned} \mathbf{C}_i^{(t)} &= [\mathbf{h}_{i-1}^{(t-1)}; \mathbf{h}_i^{(t-1)}; \mathbf{h}_{i+1}^{(t-1)}; \mathbf{s}_u^{(t-1)}; \mathbf{u}_i; \mathbf{O}_i], \\ \mathbf{h}_i^{(t)} &= \text{MHA} \left(\mathbf{h}_i^{(t-1)}, \mathbf{C}_i^{(t)}, \mathbf{C}_i^{(t)} \right), \end{aligned} \quad (10)$$

where $\mathbf{O}_i = \text{MLP}([e_{i,1}, e_{i,2}, \dots, e_{i,j}])$ is the embedding matrix composed of interest entities $e_{i,j}$ mentioned in the utterance u_i . Interest entity embedding $e_{i,j}$ is obtained from the satellite node embedding matrix \mathbf{S}^L output by the L-layer SGNN.

The goal-aware star node \mathbf{s}_x^L from the interest-level star graph also affects the update of the utterance-level relay node:

$$\begin{aligned} \mathbf{s}_u^{(t)} &= \text{MHA} \left(\mathbf{s}_u^{(t-1)}, \begin{bmatrix} \mathbf{s}_u^{(t-1)}; \mathbf{o}_x^L; \mathbf{H}^{(t)} \\ \mathbf{s}_u^{(t-1)}; \mathbf{o}_x^L; \mathbf{H}^{(t)} \end{bmatrix} \right), \end{aligned} \quad (11)$$

where $\mathbf{o}_x^L = \text{MLP}(\mathbf{s}_x^L)$, and $\mathbf{H}^{(t)} = [\mathbf{h}_1^{(t)}; \mathbf{h}_2^{(t)}; \dots; \mathbf{h}_n^{(t)}]$. After the T-step cyclic update, we further use an MLP layer with LayerNorm

Datasets		OpenDialKG	TG-ReDial
Corpus Info.	#Domains	Movie,Book	Movie
	#Dialogues	13,802	10,000
	#Utterances	126,104	129,392
	#Avg. Words	12.7	19.0
	#Split Ratio	7:1.5:1.5	8:1:1
KG Info.	#Entities	100,813	62,348
	#Relations	1,358	60
	#Triplets	1,190,658	802,578

Table 1: Statistics of datasets after preprocessing.

to fuse the relay node information into the representation of the utterance-level satellite nodes to obtain the enhanced goal-aware utterance representation \mathbf{H}_u , and take it as the initial decoding state: $\mathbf{H}_u = \varphi \left(\left[\mathbf{h}_1^{(T)}; \mathbf{h}_2^{(T)}; \dots; \mathbf{h}_n^{(T)} \right], \mathbf{s}_u^{(T)} \right)$

In the decoding stage, we adopt the decoder framework of the vanilla Transformer. In order to further enhance the goal-aware ability of response generation and be in line with the user’s current interest, inspired by Zhou et al. (2021), we incorporate user-preferred word bias in the output of the self-attention sub-layer of the decoder’s i -th layer: $\mathbf{R}_{i-1} = \mathbf{R}_{i-1} + \eta(\mathbf{p}_u)$, where \mathbf{R}_{i-1} is the input of the decoder at the i -th layer. $\eta(\cdot) : \mathbb{R}^{d_e} \rightarrow \mathbb{R}^{d_w}$, and d_w is the dimension of the hidden layer. To learn the response generation module, we set the negative log-likelihood loss as:

$$L_{GEN} = -\frac{1}{m} \sum_{t=1}^m [\log(P_{gen}(y_t | y_{1:t-1}, U, S, G))] + \beta * L_{REC}, \quad (12)$$

where β is a hyperparameter that represents the weight of the recommendation loss.

5 Experiments

5.1 Experiment Setup

Datasets We conduct experiments on two CRS datasets. (1) OpenDialKG (Moon et al., 2019) is a parallel Dialog \leftrightarrow KG CRS corpus where the interest entities mentioned in the dialog are linked to KG. (2) TG-ReDial (Zhou et al., 2020b) is a topic-guided CRS corpus, in which each dialog is associated with a topic thread. The movies mentioned in the corpus are linked to a KG: CN-DBpedia (Xu et al., 2017). To make full use of the annotated topic, we add each topic as an entity to CN-DBpedia to obtain a topic-enhanced KG. We add new edges between the movie and the topic entities based on their real relationship in the Chinese

Models	OpenDialKG			TG-ReDial		
	R@1	R@10	R@25	R@1	R@10	R@25
TextCNN	0.058	0.176	0.236	0.178	1.034	1.872
KBRD	0.107	0.409	0.494	0.337	3.119	6.073
KGSF	0.120	0.433	0.522	0.419	3.761	6.925
RevCore	0.124	0.428	0.515	0.490	3.494	6.631
CRFR	0.130	0.453	0.538	0.646	4.189	7.509
C ² -CRS	0.112	0.458	0.535	0.935	4.401	7.991
CR-GIS	0.182*	0.473*	0.547*	1.211*	6.202*	10.469*
w/o GoInfo.	0.157	0.463	0.533	0.958	5.937	9.849
w/o ImpEnc.	0.129	0.401	0.487	0.709	4.185	7.476
w/o SeqEnc.	0.145	0.453	0.529	0.891	5.295	8.909

Table 2: Overall recommendation evaluation. w/o refers to removing CR-GIS components. “*” indicates the statistical significance for $p < 0.005$ compared with the best baseline. TG-ReDial results are reported in percentage.

movie review website Douban, i.e., whether a topic appears in the comments or tags of a movie. The statistics after preprocessing are in Table 1.

Baselines We compare our CR-GIS with the following competitive models: (1) *TextCNN* (Kim, 2014) is a recommendation model extracting user preference from utterances with a CNN-based model. (2) *Transformer* (Vaswani et al., 2017) is a vanilla Transformer-based dialog generation model. (3) *KBRD* (Chen et al., 2019) is a Knowledge-Based CRS integrating item-oriented KG. (4) *KGSF* (Zhou et al., 2020a) is a KG-based Semantic Fusion 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 focusing on multi-hop reasoning on KGs. (7) *C²-CRS* (Zhou et al., 2022) is a CRS adopting coarse-to-fine contrastive learning. For a fair comparison, all KG-based CRS models share the same KG.

Implementation Details Our model is implemented with Pytorch. The dimensionality of KG embedding d_e and word embedding d_w are set to 128 and 300. The layers of R-GCN, SGNN and Star Transformer encoder are set to 1, 6 and 2. We use the Adam optimizer (Kingma and Ba, 2015), the batch size is 32, the learning rate is 0.001, and gradient clipping restricts the gradients within [0, 0.1]. The whole training process is split into three steps. First, we minimize the L_{MIM} loss for pretraining the goal-oriented information fusion module. After that, we minimize the L_{REC} loss with the weight α of 0.1. Finally, we minimize the L_{GEN} loss with the weight β of 0.5 on OpenDialKG dataset and 0 on TG-ReDial dataset.

Models	OpenDialKG					TG-ReDial				
	Bleu-1	Bleu-2	Dist-1	Dist-2	HIT	Bleu-1	Bleu-2	Dist-1	Dist-2	HIT
Transformer	0.389	0.310	0.027	0.113	0.126	0.504	0.386	0.011	0.055	8.868
KBRD	0.404	0.318	0.055	0.164	0.251	0.506	0.395	0.025	0.082	9.105
KGSF	0.412	0.326	0.061	0.202	0.262	0.509	0.396	0.029	0.100	9.158
RevCore	0.406	0.321	0.054	0.193	0.240	0.511	0.403	0.020	0.079	9.002
CRFR	0.418	0.331	0.062	0.204	0.288	0.515	0.401	0.031	0.103	9.225
C ² -CRS	0.414	0.328	0.064	0.210	0.308	0.516	0.402	0.031	0.100	9.260
CR-GIS	0.432*	0.342*	0.066	0.207	0.332*	0.529*	0.406*	0.033*	0.107*	9.537*
w/o GoInfo.	0.426	0.336	0.060	0.193	0.324	0.516	0.401	0.029	0.094	9.153
w/o Align.	0.428	0.336	0.059	0.191	0.314	0.518	0.402	0.030	0.095	9.447
w/o HiStar.	0.409	0.324	0.061	0.220	0.305	0.519	0.400	0.033	0.111	9.235

Table 3: Overall conversation evaluation. w/o refers to removing CR-GIS components. “*” indicates the statistical significance for $p < 0.05$ compared with the best baseline. The HIT of TG-ReDial are reported in percentage.

5.2 Evaluation on Recommendation

Overall Performance As shown in Table 2, we adopt the recognized Recall@K (K=1,10,25) metrics to evaluate the recommendation performance. CR-GIS significantly outperforms all baselines over all metrics. Specifically, compared with the best results of the baselines, CR-GIS improves the Recall@K (K=1,10,25) metrics by about 40%, 2.6%, 1.7%, and 29.5%, 40.9%, 31% on the OpenDialKG and TG-ReDial datasets, respectively. Although introducing more external knowledge (i.e., KBRD, KGSF, RevCore, C²-CRS) and multi-hop KG reasoning (i.e., CRFR) has achieved staged success, the impressive performance of CR-GIS shows that it is necessary to model the goal-aware implicit user interest sequence in dialog.

Ablation Study In Table 2, we removed the key components in the recommendation module of CR-GIS for ablation study. **First**, we remove the goal-oriented information fusion mechanism, called "w/o GoInfo.". The Recall@K (K=1,10,25) results decrease slightly on OpenDialKG, but decrease significantly on TG-ReDial. It indicates the advantage of "goal-aware" recommendation, i.e., modeling the implicit relationship between the interest sequence and the goal, especially when there is a dialog goal guiding the conversation, e.g., TG-ReDial dataset. **Second**, we remove the implicit dialog relations encoder, called "w/o ImpEnc.". Note that "GoInfo." is also removed due to there is no star node. We observe that the recommendation results are significantly decreased over all metrics. It confirms that to capture the association between interest entities in long-range conversation is a key factor in improving recommendation performance. Note that we didn't ablate the explicit KG relations encoder which is a shared KG encoder for all CRS

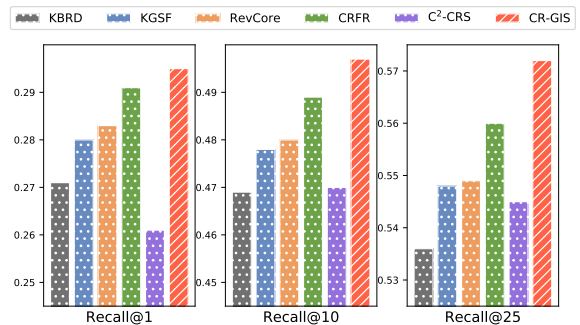


Figure 2: Analysis of CR-GIS's scalability on TG-ReDial, where topic is also treated as the recommendation goal.

baselines. **Third**, we remove the sequence encoder, called "w/o SeqEnc.", which is substituted by self-attention mechanism. The recommendation results distinctly decrease indicating that the modeling of user interests sequence is also crucial for obtaining the dynamic preferences in dialog.

The Scalability of CR-GIS To analyze the scalability of CR-GIS with constructed topic-enhanced KG, we consider the topic mentioned in the utterance as the recommendation goal to further examine the total recommendation performance on TG-ReDial. As shown in Figure 2, although the absolute value of the performance of all models has been significantly improved, CR-GIS still significantly outperforms all baselines. It further confirms that capturing the implicit association between entities in the interest flow pointing to the goal in the conversation can make up for the lack of information propagation on the KG.

5.3 Evaluation on Conversation

Overall Performance To automatically evaluate the conversation performance, we adopt Bleu-1/2,

Models	Flu.	Coher.	Info.	Proact.
Transformer	1.568	1.230	1.043	0.320
KBRD	1.611	1.248	1.192	0.347
KGSF	1.635	1.287	1.143	0.362
RevCore	1.654	1.252	1.215	0.445
CRFR	1.716	1.255	1.280	0.447
C ² -CRS	1.585	1.249	1.138	0.426
CR-GIS	1.763*	1.313*	1.292*	0.450*
kappa	0.534	0.509	0.505	0.516

Table 4: Human evaluation. "Flu.", "Coher.", "Info.", "Proact." respectively denote fluency, coherence, informativeness and proactivity. The agreement ratio $kappa \in [0.41, 0.6]$ denotes the moderate agreement. "*" indicates the statistical significance for $p < 0.05$ compared with the best baseline.

Distinct-1/2 and HIT. The Bleu and Distinct evaluate the fluency and diversity of responses, respectively. HIT evaluates the ratio of the goals in the golden response that is also hit by the generated response. In Table 3, CR-GIS exceeds all baselines. Specifically, compared with the best results of C²-CRS, CR-GIS improves HIT by 7.8% and 3% on two datasets, respectively, which indicates that CR-GIS effectively perceives the goal, thus generating responses more in line with user interests. This is attributed to the goal-oriented fusion mechanism that enhances the goal expression in the interest-level star graph, and then the goal-aware interest-level star graph is aligned with the utterance-level star graph to finally endow CR-GIS's goal-aware ability of response generation. In addition, CR-GIS achieves 4.3%, 4.3% and 2.5%, 1% improvements on Bleu-1 and Bleu-2 compared with the best baselines on two datasets, respectively, which also demonstrates that CR-GIS prefers to capture more useful information in n-gram's level leading to more fluent responses.

Human Evaluation For human evaluation, we randomly sampled 200 context-response pairs per model from OpenDialKG. Five well-educated annotators evaluate each response with four indicators: fluency, coherence, informativeness and proactivity. Proactivity evaluates whether the dialog agent makes proactive actions such as recommending or asking user preference. The scores are settled from $\{0, 1, 2\}$ to estimate fluency, coherence, informativeness. The proactivity scores are assigned from $\{0, 1\}$. The agreement among the annotators is measured by the Fleiss' kappa (Fleiss, 1971). As the results in Table 4, the superior of CR-GIS on all indicators support the observations from

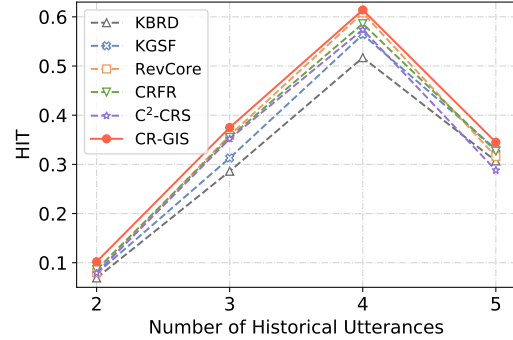


Figure 3: As the number of historical utterances increases, CR-GIS outperforms all baselines in terms of hitting entity and topic in generated response.

automatic evaluations. **Case studies generated by different models are in the Appendix A.**

Ablation Study We also remove the key components of CR-GIS to discuss their contributions in Table 3. **First**, we remove the goal-oriented information fusion mechanism, called "w/o GoInfo.". The HIT has a slight drop on OpendialKG, but decreases significantly on TG-Redial. This is consistent with the observations on the Recall@K. **Second**, we remove the alignment between the interest-level and utterance-level star graphs, called "w/o Align.". In the results, Bleu and Dist decrease slightly, but HIT decreases distinctly on OpenDialKG than TG-Redial. This verifies that the goal-aware interests injected into the utterance representations has an impressive impact on the goal-aware generation. **Third**, we replace the hierarchical Star Transformer encoder with a hierarchical vanilla Transformer encoder, called "w/o HiStar.", and utilize MHA to maintain the alignment between the interest-level star graph and the utterance encoding. We find that although hierarchical Star Transformer would damage the diversity of generation to a certain extent, it has better advantages in terms of improving the quality of generation.

The Effectiveness of Goal-oriented Guidance

To further explore the goal perception and generation abilities of CR-GIS, we examine the proportion of the generated response hitting the recommendation goals with the increase of the number of the utterances in the conversation history on the TG-Redial dataset. Besides movies, topics are also used as goals. As shown in Figure 3, as the number of historical utterances increases, the performance of all models increases first and then decreases. Simultaneously, CR-GIS always maintains the best

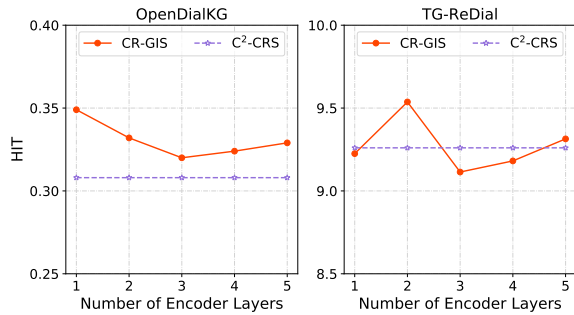


Figure 4: As the number of hierarchical Star Transformer encoder layers increases, CR-GIS consistently outperforms the best baseline C²-CRS in most settings.

performance. This happens possibly because the topic often shift in the dialog and thus also change the direction of interest sequence.

Parameter Analysis We analyze the influence of the number of the hierarchical Star Transformer encoder layers on HIT. In Figure 4, when the number of layers is greater than 2, the HIT metric shows a downward trend. This happens possibly because the goal-aware interest-level star graph and the utterance-level star graph are overfitted during the alignment due to the number increasing of encoder layers. Furthermore, CR-GIS consistently exceeds the best baseline C²-CRS in most cases, confirming the effectiveness and robustness of CR-GIS.

6 Conclusion

In this paper, we propose CR-GIS to model goal-aware interest sequence in CRS with an interest-level star graph whose star node is oriented to the recommendation goal. By using goal-oriented interest-level star graph to guide a utterance-level star graph, CR-GIS significantly improve the goal-aware ability in generating smoothly transited response. Extensive experiments verify the superiority of CR-GIS in terms of recommendation, conversation, and reaching recommendation goals.

Acknowledgements

This work was supported by National Natural Science Foundation of China (62272340, 61876128, 61876129, 61976154), State Key Laboratory of Communication Content Cognition (Grant No.A32003).

References

- 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.
- 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.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics.
- 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.
- Qipeng Guo, Xipeng Qiu, Pengfei Liu, Yunfan Shao, Xiangyang Xue, and Zheng Zhang. 2019. [Star-transformer](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 1315–1325. Association for Computational Linguistics.
- Shirley Anugrah Hayati, Dongyeop Kang, Qingxi-aoyang Zhu, Weiyan Shi, and Zhou Yu. 2020. [IN-SPIRED: 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.
- R. Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Philip Bachman, Adam Trischler, and Yoshua Bengio. 2019. [Learning deep representations by mutual information estimation and maximization](#). In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.

- 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.
- 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.
- 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 Kostrić, Krisztián 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, Xiangnan He, Yisong Miao, Qingyun Wu, Richang Hong, Min-Yen Kan, and Tat-Seng Chua. 2020. [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.
- 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.
- Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard S. Zemel. 2016. [Gated graph sequence neural networks](#). In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*.
- 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.
- Lizi Liao, Ryuichi Takanobu, Yunshan Ma, Xun Yang, Minlie Huang, and Tat-Seng Chua. 2019. [Deep conversational recommender in travel](#). *ArXiv*, abs/1907.00710.
- 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.
- 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.
- 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.
- Zhiqiang Pan, Fei Cai, Wanyu Chen, Honghui Chen, and Maarten de Rijke. 2020. [Star graph neural networks for session-based recommendation](#). In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, pages 1195–1204. ACM.
- 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.
- 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.

- 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.
- Bo Xu, Yong Xu, Jiaqing Liang, Chenhao Xie, Bin Liang, Wanyun Cui, and Yanghua Xiao. 2017. [Cndbpedia: A never-ending chinese knowledge extraction system](#). In *Advances in Artificial Intelligence: From Theory to Practice - 30th International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2017, Arras, France, June 27-30, 2017, Proceedings, Part II*, volume 10351 of *Lecture Notes in Computer Science*, pages 428–438. Springer.
- 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.
- Yiming Zhang, Lingfei Wu, Qi Shen, Yitong Pang, Zhihua Wei, Fangli Xu, Bo Long, and Jian Pei. 2021. Multi-choice questions based multi-interest policy learning for conversational recommendation. *arXiv preprint arXiv:2112.11775*.
- 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.
- 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.
- Yuanhang Zhou, Kun Zhou, Wayne Xin Zhao, Cheng Wang, Peng Jiang, and He Hu. 2022. [C2-CRS: coarse-to-fine contrastive learning for conversational recommender system](#). *CoRR*, abs/2201.02732.
- 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 Case Study

Two cases generated on OpenDialKG and TG-ReDial datasets from seven models and ground truth are selected in Table 5. We use **red text** to annotate the defects in expressing the recommendation goals for each response generated by the baseline models. In comparison, the advantage of CR-GIS is that its responses are more coherent, fluent, and informative and achieve recommendation goals that are more in line with user interests. This observation further verifies the advantage of CR-GIS in modeling goal-aware implicit user interest sequence, which benefits the smooth transition of utterances towards recommendation goal.

OpenDialKG	
Context	I like J.K.Rowling. Me, too! She’s great! Did you want to know some cool stuff about her? Yes, I want to know.
Transformer	Yes, I did. It was released in 2013. It was written by the film. (<i>incoherence</i>)
KBRD	JK Rowlings wrote this book (<i>less-informativeness</i>)
KGSF	Yes, I can recommend Harry Potter and the Sorcerer’s Apprentice. (<i>less-coherence, less-informativeness</i>)
RevCore	Yes, I think she was in the Harry Potter and the Philosopher’s Stone. Did you know that one also starred in the Philosopher’s Stone? (<i>fact inconsistency, i.e., not "was in", fact is "wrote"</i>)
CRFR	Yes, he wrote The Tales of Beedle the Bard and The Cuckoo ’s Calling. Have you seen any (<i>less-fluency, gender inconsistency</i>)
C ² -CRS	Yes, I didn’t know that. Did he also write the Harry Potter series? (<i>less-informativeness, gender inconsistency</i>)
CR-GIS	Sure, she wrote The Tales of Beedle the Bard, The Silkworm, which is a series. Have you read any of those?
Ground-Truth	Did you know she wrote two books called The Silkworm as well as the book The Cuckoo’s Calling?
TG-ReDial	
Context	好啊， 我比较喜欢人性题材的， 真实改编的电影。 Well, I prefer movies based on human nature and reality. 推荐一部《现代启示录》， 真的展现出人性混沌无善恶， 残酷壮丽一瞬间。 Recommend a Apocalypse Now, it really shows the chaos of human nature, no good and evil, cruel and magnificent moment. 你说这个影片能展现出人性混沌无善恶， 残酷壮丽一瞬间， 这个评价很高啊， 一会看看。 You said that this movie can show the chaos of human nature without good and evil, cruel and magnificent for a moment, the evaluation is very high, I will see it later. 推荐你看《势不两立》， 是关于人性的背叛， 很现实的电影， 两个主角都很饱满， 对人性的塑造非常到位。 I recommend you to watch The Edge, which is a very realistic film about the betrayal of human nature. The two protagonists are full and the shaping of human nature is very in place. 感觉这样的电影展现的就是： 这样的人越是苦难， 越是闪耀人性光辉。 能让我们感悟很多吧？ I feel that what such a movie shows is: the more suffering such a person is, the more shining the brilliance of human nature. Can we learn a lot?
Transformer	还有一部关于人性的电影， 叫《现代启示录》， 这也是一部关于 There is also a movie about human nature called Apocalypse Now, which is also a movie about (<i>retell, nonfluency</i>)
KBRD	看过《UNK》吗， 这是一部好电影， 人性的原版看起来要看起来， 但很爽。 Have you seen <UNK>, it’s a good movie, the original version of human nature looks to look good, but it’s cool. (<i>less-informativeness</i>)
KGSF	《双重人格》非常适合你， 影片中的人性， 人性， 两个极端极端探讨。 The Double is very suitable for you, the movie’s human nature, human nature, two extreme extreme exploration. (<i>repetition</i>)
RevCore	《人性污点到底是什么才是烂片之一。 还有其他关于人性的电影吗？ What the hell is The Human Stain is one of the bad movies. Are there any other movies about human nature? (<i>nonfluency, incoherence</i>)
CRFR	《越战创伤》非常好， 老帕还挺对我胃口。 Casualties of War is very good, and Palma is kinda to my taste. (<i>less-coherence, none user-centric</i>)
C ² -CRS	行啊， 最好是那种反映人性心理社会的讽刺题材电影。 Okay, it’s better to be a satirical movie that reflects human psychology and society. (<i>less-informativeness</i>)
CR-GIS	《忏情记》满足你的要求， 人性与法律， 救赎与出卖。 I Confess meets your requirements, human nature and law, redemption and betrayal.
Ground-Truth	有一部电影叫《忏情记》， 希区柯克演的人性题材的电影， 结局是人性光辉的闪耀， 也是神父之罪的阐述。 There is a movie called I Confess, a movie about human nature played by Hitchcock. The ending is the shining of human nature, and it is also an exposition of the sin of the priest.

Table 5: **Case Study.** Generated responses from baselines and CR-GIS. Context is the multi-turn conversation history in which the seeker and the recommender speak alternately.