

# FacetCRS: Multi-Faceted Preference Learning for Pricking Filter Bubbles in Conversational Recommender System

Yongsen Zheng<sup>1</sup>, Ziliang Chen<sup>2</sup>, Jinghui Qin<sup>3</sup>, Liang Lin<sup>1\*</sup>

<sup>1</sup>Sun Yat-sen University

<sup>2</sup>Jinan University

<sup>3</sup>Guangdong University of Technology

{z.yongsensmile, scape1989}@gmail.com, c.ziliang@yahoo.com, linliang@ieee.org

## Abstract

The filter bubble is a notorious issue in Recommender Systems (RSs), which describes the phenomenon whereby users are exposed to a limited and narrow range of information or content that reinforces their existing dominant preferences and beliefs. This results in a lack of exposure to diverse and varied content. Many existing works have predominantly examined filter bubbles in static or relatively-static recommendation settings. However, filter bubbles will be continuously intensified over time due to the feedback loop between the user and the system in the real-world online recommendation. To address these issues, we propose a novel paradigm, Multi-Facet Preference Learning for Pricking Filter Bubbles in Conversational Recommender System (FacetCRS), which aims to burst filter bubbles in the conversational recommender system (CRS) through timely user-item interactions via natural language conversations. By considering diverse user preferences and intentions, FacetCRS automatically model user preference into multi-facets, including entity-, word-, context-, and review-facet, to capture diverse and dynamic user preferences to prick filter bubbles in the CRS. It is an end-to-end CRS framework to adaptively learn representations of various levels of preference facet and diverse types of external knowledge. Extensive experiments on two publicly available CRS-based datasets demonstrate that our proposed method achieves state-of-the-art performance in mitigating filter bubbles and enhancing recommendation quality in CRS.

## Introduction

Conversational Recommendation Systems (CRSs) have emerged as powerful tools for providing personalized recommendations through natural language conversations and dialogue interactions (Qin et al. 2023; Li et al. 2023a; Mishra, Priya, and Ekbal 2023), which are broadly adopted in various domains, including online e-commerce (Liu et al. 2023), music recommendation (Epure and Hennequin 2023), health counseling (Wahbeh et al. 2023), *etc.* However, CRSs frequently encounter the challenge of filter bubbles, wherein users are consistently presented with a restricted range of information and suggestions that are aligned with their dominant preferences throughout their interactions with the system (Gao et al. 2022). Therefore, it is

crucial to prick filter bubbles for improving performance of the CRSs.

Recently, many research efforts have been dedicated to exploring filter bubbles in offline recommendation without considering user-item interactions (Hussein, Juneja, and Mitra 2020; Liu et al. 2021; Nguyen et al. 2014; Ribeiro et al. 2020; Spinelli and Crovella 2020; Tomlein et al. 2021). These offline recommendation approaches primarily focus on investigating the underlying causes of filter bubbles. Through extensive experiments conducted on large-scale recommender systems, two primary factors have been discovered. The first primary factor is that users with less diverse preferences are more susceptible to becoming entrapped within filter bubbles. Another primary factor is that the learning mechanisms employed by the recommender amplify the filter bubble phenomenon as they tend to prioritize and accentuate a user’s dominant interests. Although these studies provide valuable insights into the phenomenon of filter bubbles, they have not adequately addressed the negative impacts of the feedback loop between the user and the system on the exacerbation of filter bubbles. More recently, Gao et al. (Gao et al. 2022) have made attempts to mitigate filter bubbles in Interactive Recommender Systems (IRSs) by incorporating user-system interactions over time through offline reinforcement learning, but this strategy cannot allow users to interact with the system through natural language to express their true thoughts.

Despite their effectiveness, most existing methods still suffer from two major limitations: 1) *Interactive Strategy*. Most methods primarily focus on addressing filter bubbles in offline recommendations in the static settings while overlooking the effect of the user-system feedback loop on filter bubbles. In reality, filter bubbles will be exacerbated when the user chats with the system over time (Steck 2018; Gao et al. 2022). Even though a recent study CIRS (Gao et al. 2022) has emerged to combat filter bubbles interactively, its interaction strategy remains limited largely to rigid and single forms (*e.g.*, click or skip, like or dislike). This fails to allow users to express their diverse, dynamic, and complicated preferences via natural language utterances, making it not applicable to real-world conversational recommendations. 2) *Preference Exploration*. Most previous works (Hussein, Juneja, and Mitra 2020; Liu et al. 2021; Nguyen et al. 2014; Ribeiro et al. 2020) have found that modeling users’ diverse

\*Liang Lin is the corresponding author.

preferences can effectively prick filter bubbles. Based on this foundation, many methods strive to learn multi-level preferences to alleviate the filter bubble problem. But these approaches heavily rely on a limited set of sparse data sources such as historical user-item interactions, interest tags, and implicit and explicit ratings, in which a large portion of the available entries are empty or missing. This reliance on homogeneous and sparse data poses a significant limitation in excavating diverse and personalized preferences as it is still restricted by past preferences. Thus, it is hard to prevent filter bubbles from these scarce data sources.

To address these problems, we propose a novel framework, **Multi-Faceted Preference Learning for Pricking Filter Bubbles in Conversational Recommender System (FacetCRS)**. It models diverse user preferences from multiple facets adaptively, including entity-, word-, context-, and review-facet, to mitigate filter bubbles in the CRS. In these facets, the entity facet is extracted from item-oriented Knowledge Graph (KG) DBpedia (Auer et al. 2007). It contains all item entities appearing in the conversation based on the paths in such KG. This facet not only models user-item relationships better but also allows the recommender to discover new correlated items by the connections over KG, thus enabling the provision of diverse recommendations. Meanwhile, analyzing the prominent words employed by users during their interactions allows for identifying niche or specific preferences that might not be readily apparent through other facets of user profiling. Thus, we incorporate a word-oriented KG ConceptNet (Speer, Chin, and Havasi 2017) to model the relations such as synonyms and co-occurrence between words. Moreover, the ongoing dialogue (*i.e.*, context-facet preference) provides an opportunity to understand the evolving interests of users, so the CRSs can adapt their recommendations to align with current user preferences. Lastly, the reviews facet provides a direct and authentic information source reflecting the firsthand experiences of users. They often contain descriptive narratives, ratings, and opinions that shed light on various aspects of the items, such as quality, functionality, and performance. Analyzing these reviews allows recommender systems to identify common patterns, sentiment trends, and user sentiments towards specific features, helping them better understand user interests. We conduct experiments on two public CRS-based benchmarks, where our superior performance demonstrates the appealing reliability of our method.

Overall, our main contributions are included as follows:

- To the best of our knowledge, this is the first work to model user preference into multi-facets in the CRS, including entity-, word-, context-, and review-facet, to capture diverse user preferences to prick filter bubbles in the CRS.
- We proposed a novel paradigm, FacetCRS, which models multi-faceted user preference into an end-to-end CRS framework. It will adaptively learn representations of various levels of preference facet and multi-aspect knowledge.
- Quantitative and qualitative experimental results on two CRS-based datasets exhibits superior performance of our proposed method, which serve as compelling evidence of the effectiveness of our FacetCRS in mitigating the filter bubble in the CRS.

## Related Work

### Conversational Recommender System

With the widespread adoption of intelligent agents in diverse domains (Chen et al. 2017), conversational recommender systems have attracted a lot of attention. These CRSs can be discerned into two primary categories: attribute-based CRS and human-like CRS. The former (Christakopoulou, Radlinski, and Hofmann 2016; Sun and Zhang 2018; Zhou et al. 2020b; Zhang et al. 2022) focus on eliciting user preferences by prompting them to express their likes or dislikes with respect to various attributes, leveraging predefined actions such as item attributes and intent slots. Conversely, the latter (Lu et al. 2021; Zhou et al. 2022, 2020c; Hayati et al. 2020; Li et al. 2018) endeavors to furnish more lifelike recommendations through human-like responses. Human-like CRS typically comprises a conversation module responsible for generating appropriate responses and a recommendation module for delivering recommendations. Despite their sophistication, these approaches still suffer from the scarcity and insufficiency of contextual information present in initial conversational utterances. To address these challenges, existing methods often incorporate additional information from external sources, such as structured data like knowledge graphs (Zhou et al. 2022, 2020a), or unstructured data like reviews (Lu et al. 2021), to enhance the conversation utterances. Nevertheless, they still grapple with drawbacks associated with a limited scope of knowledge facets in external sources. Our work adopts a human-like CRS approach and models user preferences across multiple facets, encompassing entities, words, contexts, and reviews.

### Filter Bubbles in Recommendation

The filter bubble has emerged as a widely recognized predicament in the field of recommendation systems. When users find themselves ensnared within a filter bubble, the system tends to present information that aligns closely with their existing beliefs or interests, resulting in a curtailed exposure to a broader range of information (Liu et al. 2021; Nguyen et al. 2014). Recent research endeavors have delved into investigating the primary factors contributing to the formation of filter bubbles. From a user-centric perspective, individuals with narrower and less diverse preferences are more susceptible to becoming trapped within these bubbles (Ribeiro et al. 2020; Spinelli and Crovella 2020). From a system perspective, the learning process can inadvertently amplify a user’s dominant interests. Furthermore, the system typically assumes that user satisfaction equates to intrinsic interest by implicitly assuming that excessive exposure to items of interest will not affect user satisfaction (Gao et al. 2022; Li et al. 2023b). For example, Wang et al. (Wang et al. 2022) introduce a novel recommender paradigm that empowers users to actively control the mitigation of filter bubbles. However, existing strategies predominantly concentrate on the static recommendation setting, which poses challenges in effectively modeling the dynamic nature of filter bubbles. In contrast, our approach models multifaceted user preferences when the user interacts with the system via natural language conversations to prick filter bubbles.

## FacetCRS

Filter bubble is a challenging issue in CRSs, and it will be increasingly severe over time due to the feedback loop between the user and the system in the online recommendation. To address these issues, we propose a novel framework, FacetCRS, which comprises Multi-Facet Preference Learning and Multi-Facet Conversational Recommender System. The pipeline of our FacetCRS is shown in Fig. 1.

### Multi-Facet Preference Learning

Numerous extensive experiments conducted by most existing recommendation methods (Hussein, Juneja, and Mitra 2020; Liu et al. 2021; Nguyen et al. 2014) have consistently demonstrated that individuals with narrower preferences tend to be more susceptible to the confinement of filter bubbles. Along this line, we devise the Multi-Facet Preference Learning to model diverse user preferences into multi facets, including entity-, word-, context-, and review-facet, to improve recommendation diversity.

**Entity-Facet User Preference.** The KG entities facet not only enhance the user-item relationship modeling but also allows the recommender to discover new items by leveraging the connections within the KG, thus enabling the provision of diverse recommendations. To do this, we adopt the large-scale item-oriented KG DBpedia (Auer et al. 2007) to extract the entities of the paths in the KG, in which those entities usually appear in conversations as items, inspired by (Lu et al. 2021; Li et al. 2022; Zhou et al. 2022). The entities always present in the form of head entities  $e_{\text{head}}$  and tail one  $e_{\text{tail}}$  in the fruitful knowledge triples  $(e_{\text{head}}, r, e_{\text{tail}})$ , where  $r$  is the relation between  $e_{\text{head}}$  and  $e_{\text{tail}}$ . Given a conversation context  $\mathcal{C} = \{s_t\}_{t=1}^n$ , we first extract  $k$  entities  $\mathcal{E}_k^c = \{e_1, e_2, \dots, e_k\}$  via DBpedia, where  $s_t = \{w_j\}_{j=1}^m$  denotes one sentence (*i.e.*, each utterance) and  $w$  represents a word from vocabulary  $\mathcal{V}$ . Then, we utilize the Relational Graph Convolutional Networks (RGCN) to learn entity embeddings. Concretely, we utilize RGCN to learn entity representations on the extracted subgraph, and thus the representation of entity  $e$  at  $(l+1)^{\text{th}}$  layer is calculated as:

$$e^{l+1} = \sigma\left(\sum_{r \in \mathcal{R}} \sum_{\hat{e} \in \mathcal{H}_e^r} \frac{1}{z_{l+1}} \mathbf{W}_r^l \hat{e}^l + \mathbf{W}^l e^l\right). \quad (1)$$

Here  $e^l$  means the  $l^{\text{th}}$  layer's representation of entity  $e$ ,  $\sigma$  is the sigmoid function,  $\hat{e}$  is the entities from one-hop neighbor set  $\mathcal{H}_e^r$  of entity  $e$  under the relation  $r$ ,  $z_{l+1}$  is the hyperparameter, *i.e.*, the normalization factor.  $\mathbf{W}_r^l$  and  $\mathbf{W}^l$  can be trained during model training. With the spirit of previous works (Li et al. 2022; Lu et al. 2021), the last layer's representation  $e^L$  is employed as the entity representation  $e$  of the entity  $e$ . Therefore, the entity-facet embeddings  $\mathbf{X}_e$  can be described as:

$$\mathbf{X}_e = \text{RGCN}(\mathcal{E}_k^c) = \{e_1^T, e_2^T, \dots, e_k^T\}, \quad (2)$$

where  $e_i^T$  is the embedding of the entity  $e_i$  via RGCN.

**Word-Facet User Preference.** Keywords used in conversations serve as critical indicators to gain valuable insights

into the user's specific needs. Analyzing the prominent words employed by users during their interactions allows for the identification of niche or specific preferences that might not be readily apparent through other facets of user profiling. Thus, word-facet information enables the delivery of more personalized and tailored recommendations. To model user preference in word facet, the external lexical word-oriented KG ConceptNet (Speer, Chin, and Havasi 2017) is utilized to establish semantic relationships among words, encompassing diverse associations such as synonymy, antonyms, and co-occurrence. Given a context  $\mathcal{C}$ , the first step is to draw  $m$  words  $\mathcal{W}_m^c = \{w_1, w_2, \dots, w_m\}$ , and then utilize Graph Convolutional Networks (GCN) to learn the representations of current words. In GCN, the  $(l+1)^{\text{th}}$  layer's representation of word  $w$  can be expressed as:

$$\mathbf{w}^{l+1} = \sigma(\hat{\mathbf{A}}\mathbf{w}^l\mathbf{W}^l). \quad (3)$$

where  $\mathbf{w}^l$  is the  $l^{\text{th}}$  layer's representation of word  $w$ ,  $\hat{\mathbf{A}} = \mathcal{D}^{-\frac{1}{2}}\mathcal{A}\mathcal{D}^{-\frac{1}{2}}$  is the normalized adjacency matrix,  $\mathcal{A}$  is the adjacency matrix by retrieving semantic relationships of words from ConceptNet, and  $\mathcal{D}$  is a diagonal matrix of size  $l_A \times l_A$  where the elements on the diagonal represent the degrees of the nodes, *i.e.*,  $\mathcal{D}_{ii} = \sum_j^l \mathcal{A}_{ij}$ .  $\mathbf{W}^l$  is the trainable weights. Similar to RGCN, we use the last layer's representation  $\mathbf{w}^L$  as the word representation  $w$  of the word  $w$ . Thus, the word-facet embeddings  $\mathbf{X}_w$  is given as:

$$\mathbf{X}_w = \text{GCN}(\mathcal{W}_m^c) = \{w_1^T, w_2^T, \dots, w_m^T\}, \quad (4)$$

here  $w_i^T$  is the embedding of the word  $w_i$  by GCN.

**Context-Facet User Preference.** Unlike static user profiles or explicit feedback, conversation contents (*i.e.*, context-facet preference) provide a more nuanced view of user preferences by capturing the dynamic nature of their interests. The ongoing dialogue gives an opportunity to understand the evolving interests of users. By considering the conversational context, sentiment, and topics discussed, the CRS can adapt its recommendations to align with the user's current interests. Along this line, we employ a Transformer (Vaswani et al. 2017a) as the encoder to learn the representations of the conversation contexts. For the context  $\mathcal{C}$ , suppose the output embeddings of the previous transformer layer is  $\mathcal{T}^{(l-1)}(\mathcal{C})$ , then the current one  $\mathcal{T}^l(\mathcal{C})$  can be defined by *Multi-head Attention* function  $\text{MHA}(\cdot, \cdot, \cdot)$  as follows:

$$\begin{aligned} \mathcal{T}^l(\mathcal{C}) &= \text{MHA}(\mathcal{T}^{l-1}(\mathcal{C}), \mathcal{T}^{l-1}(\mathcal{C}), \mathcal{T}^{l-1}(\mathcal{C})), \\ \text{MHA}(\mathbf{K}, \mathbf{Q}, \mathbf{V}) &= [\text{head}_1^l; \dots; \text{head}_h^l] \mathbf{W}_j^l, \end{aligned} \quad (5)$$

where  $h$  is the number of heads,  $\mathbf{W}_j^l$  is the training parameters, and each head  $\text{head}_j^l$  is computed by *Scaled Dot-Product Attention* (Vaswani et al. 2017b) functions  $\text{SDA}(\cdot, \cdot, \cdot)$  as:

$$\begin{aligned} \text{head}_j^l &= \text{SDA}(\mathcal{T}^{l-1}(\mathcal{C})\mathbf{W}_j^k, \mathcal{T}^{l-1}(\mathcal{C})\mathbf{W}_j^q, \mathcal{T}^{l-1}(\mathcal{C})\mathbf{W}_j^v), \\ \text{SDA}(\mathbf{K}, \mathbf{Q}, \mathbf{V}) &= \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d/h}}\right)\mathbf{V}. \end{aligned} \quad (6)$$

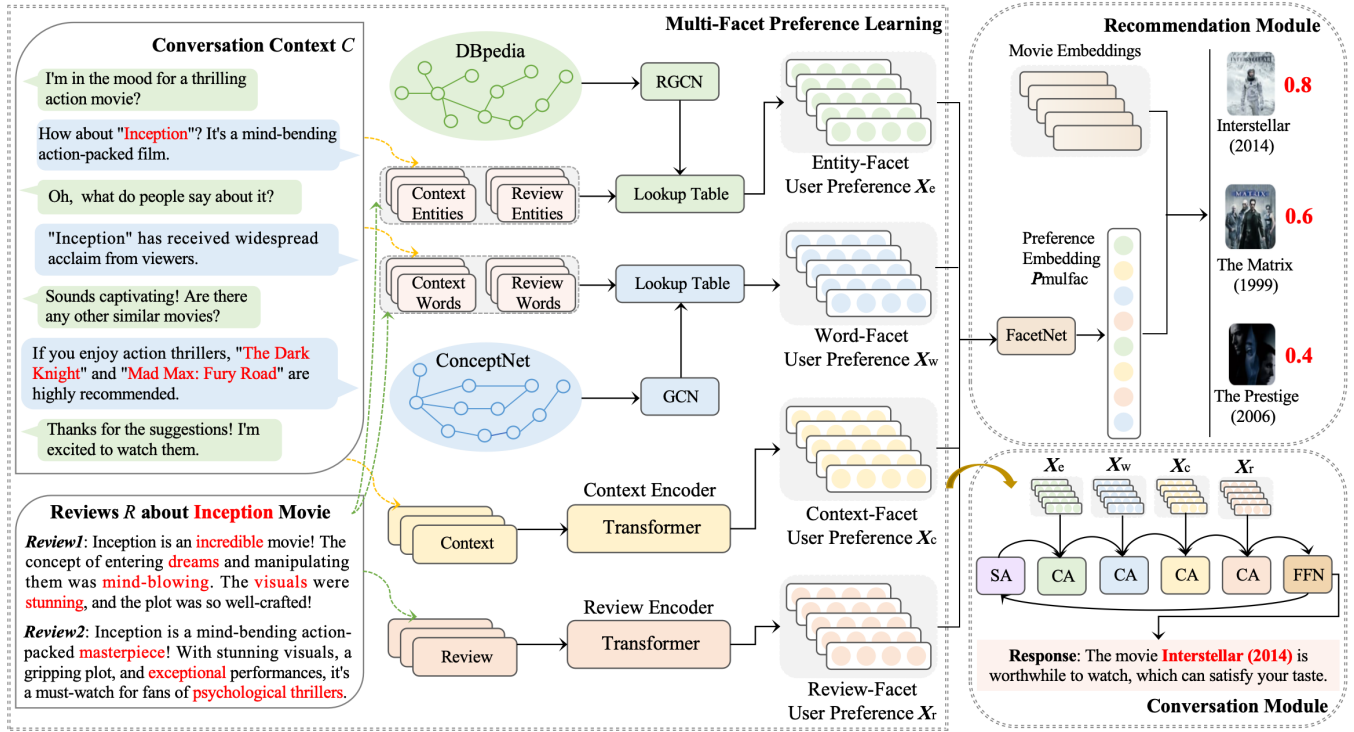


Figure 1: Overview of the proposed framework, FacetCRS, which consists of Multi-Facet Preference Learning and Multi-Facet Conversational Recommender System. The former aims to adaptively model diverse user preferences into multi-facets, including entity-, word-, context-, and review-facet, while the latter contains the recommendation module to select accurate items for the user and the conversation module to generate proper responses.

Note that  $\forall j \in [h]$ ,  $\mathbf{K}, \mathbf{Q}$  and  $\mathbf{V}$  denote the key, query, and value matrices, respectively.  $\text{Softmax}$  is the softmax function.  $\mathbf{W}_j^k, \mathbf{W}_j^q$  and  $\mathbf{W}_j^v$  are the parameters that can be learned. For convenience, we employ the output embedding of the last transformer layer to be context-facet embeddings  $\mathbf{X}_c$ . Based on Eq.(5), it can be written as:

$$\mathbf{X}_c = \text{MHA}(\mathcal{T}^{L-1}(\mathcal{C}), \mathcal{T}^{L-1}(\mathcal{C}), \mathcal{T}^{L-1}(\mathcal{C})). \quad (7)$$

Here,  $L$  denotes the number of transformer layers.

**Review-Facet User Preference.** Item reviews effectively reflect the firsthand experiences of users by describing the authentic user experiences and feelings. They often contain descriptive narratives, ratings, and opinions that shed light on various aspects of the items, such as quality, functionality, and performance. Analyzing these reviews is beneficial to identify common patterns, sentiment trends, and user sentiments towards specific features, helping them better understand user interests. Due to the advantages of reviews in exploring user preferences, some works (Lu et al. 2021; Zhou et al. 2022) have introduced reviews to improve CRS. Inspired by them, we also incorporate reviews to explore user preferences in the review facet. Given a review  $\mathcal{R}$  that is retrieved from a set of reviews based on (Lu et al. 2021), we also adopt the Transformer to model the review representations to complement other three facets of modeling user preference. Similar to context-facet embeddings modeling,

the output embeddings of the last transformer layer are considered as the review-facet user embeddings. It can be defined as follows according to Eq.(5):

$$\mathbf{X}_r = \text{MHA}(\mathcal{T}^{L-1}(\mathcal{R}), \mathcal{T}^{L-1}(\mathcal{R}), \mathcal{T}^{L-1}(\mathcal{R})). \quad (8)$$

**Fusing Multi-Facet User Preference.** In the preceding sections, we have formulated user preferences at various levels of detail, encompassing entity-, word-, context-, and review-facet considerations, gradually progressing from a broader perspective to a more nuanced one to model diverse user preference. These four types of preference facets differ in their definitions but are complementary to each other, allowing for diverse preference modeling and effectively mitigating filter bubbles in the CRS. We contend that user preferences may be collectively influenced by these facets, albeit with varying magnitudes. Consequently, the ultimate user preference representation  $\mathbf{P}_{\text{multifac}}$  can be achieved by our proposed multifaceted fusion function  $\text{FacetNet}(\cdot)$  as follows:

$$\text{FacetNet}(\cdot) = \begin{cases} \mathcal{T}_{\text{enwo}}^L = \mathcal{T}^L(\mathbf{X}_e) \otimes \mathcal{T}^L(\mathbf{X}_w), \\ \mathcal{T}_{\text{core}}^L = \mathcal{T}^L(\mathbf{X}_c) \otimes \mathcal{T}^L(\mathbf{X}_r), \\ \mathcal{T}_{\text{mfac}}^L = \mathcal{T}_{\text{enwo}}^L \oplus \mathcal{T}_{\text{core}}^L, \\ \mathbf{P}_{\text{multifac}} = \text{Softmax}(\text{ReLU}(\mathcal{T}_{\text{mfac}}^L)). \end{cases} \quad (9)$$

Here, the symbol  $\otimes$  represents the matrix multiplication operation, while  $\oplus$  signifies vector concatenation.  $\text{ReLU}$  is the

Rectified Linear Unit. Subsequently, the multifaceted user preference  $P_{\text{multifac}}$  is employed to provide accurate recommendations and generate suitable responses in the CRS.

### Multi-Facet Conversational Recommender System

To prick filter bubbles in the CRS, we adopt the multifaceted user preference  $P_{\text{multifac}}$  to effectively make item predictions in the recommendation module and accurately predict the next utterances in the conversation module.

**Recommendation Module.** The main goal of the recommendation module is to deliver precise item suggestions to users by leveraging their interactions with the system through natural conversations. In order to mitigate the impact of filter bubbles, we aim to explore the breadth of user preferences and encompass a diverse range of recommendations. Therefore, the multifaceted user preference  $P_{\text{multifac}}$  is used to excavate diverse user interests from multiple perspectives to enhance recommendation performance. Specifically,  $P_{\text{multifac}}$  is first fed into the Multilayer Perceptron Layer (MPLNet) followed by a softmax layer to generate the recommendation prediction  $P_r$ . This process is:

$$P_r = \text{Softmax}(\text{MPLNet}(P_{\text{multifac}})). \quad (10)$$

To train the model, we adopt the cross-entropy loss (Lu et al. 2021) to achieve our learning goals. The cross-entropy loss  $\mathcal{L}_r$  measures the dissimilarity between the prediction  $P_r$  and the target item category and can be computed as:

$$\mathcal{L}_r = -\frac{1}{N_r} \sum_{j=1}^{N_r} \log P_r^j, \quad (11)$$

where  $N_r$  denotes total number of recommendations, and  $P_r^{(j)}$  is the target category in the  $j$ -th recommendation.

**Conversation Module.** The goal of the conversation task is to generate suitable utterances to respond to the user. In our conversation module, we adopt the entity-facet embeddings  $X_e$ , word-facet embeddings  $X_w$ , context-facet embeddings  $X_c$ , and review-facet embeddings  $X_r$  to the decoder network for simultaneously predicting the next responses. Concretely, these multi-facet user preference embeddings are fed into each cross-attention layer to fuse their useful knowledge information to produce diverse response contents. Formally, we adopt  $Y^{(i-1)}$  to denote the output of the last time unit, then the current one  $Y^{(i)}$  is:

$$\begin{aligned} A_0^i &= \text{MHA}(Y^{i-1}, Y^{i-1}, Y^{i-1}), \\ A_1^i &= \text{MHA}(A_0^i, X_e, X_e), \\ A_2^i &= \text{MHA}(A_1^i, X_w, X_w), \\ A_3^i &= \text{MHA}(A_2^i, X_c, X_c), \\ A_4^i &= \text{MHA}(A_3^i, X_r, X_r), \\ Y^i &= \text{FFN}(A_4^i). \end{aligned} \quad (12)$$

Here, MHA can be computed by Eq. (5), and the FFN( $\cdot$ ) is the fully-connected feed-forward network, which can be modelled as follows:

$$\text{FFN}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2, \quad (13)$$

where  $W_1$  and  $W_2$  represent the learnable weight matrices, and  $b_1$  and  $b_2$  denote the bias terms corresponding to the first and second layers, respectively. As illustrated above, multifaceted user preference information is progressively integrated into the decoding stage. Initially, the entity-facet user preference is incorporated, then the word-facet user preference, followed by the context-facet user preference, and lastly, the review-facet user preference. Integrating multiple facets of information enriches users' understanding of items, resulting in a comprehensive and diverse range of user interests, which can effectively prick filter bubbles in CRS. To train the conversation module, we also adopt the cross-entropy loss inspired by (Lu et al. 2021) to learn the response generation. The conversation loss  $\mathcal{L}_c$  can be illustrated as:

$$\mathcal{L}_c = -\frac{1}{M} \sum_{t=1}^M \log(\tilde{P}(s_t|\{s_{t-1}\})),$$

$$\tilde{P}(s_t|\{s_{t-1}\}) = \tilde{P}_v(s_t|Y_i) + \tilde{P}_g(s_t|Y_i, \mathcal{G}) + \tilde{P}_r(s_t|Y_i, \mathcal{R}), \quad (14)$$

where  $M$  is the number of conversation turns,  $s_t$  represents the  $t^{\text{th}}$  utterance in the conversation,  $\{s_{t-1}\}$  represent the previously generated sub-sequence, which can be denoted as  $\{s_{t-1}\} = s_1, s_2, \dots, s_{t-1}$ .  $\tilde{P}(s_t|\{s_{t-1}\})$  refers to the generation probability of the next token  $s_t$ . We utilize  $\tilde{P}_v(\cdot)$ ,  $\tilde{P}_g(\cdot)$ , and  $\tilde{P}_r(\cdot)$  as probability functions over the vocabulary  $\mathcal{V}$ , entities from KG  $\mathcal{G}$ , and reviews  $\mathcal{R}$  with  $Y_i$  as the input.

## Experiments and Analyses

In this section, we conduct experiments to evaluate the performance of FacetCRS and answer the following questions:

- **RQ1:** How does FacetCRS perform compared with state-of-the-art methods in the recommendation task?
- **RQ2:** How does FacetCRS perform compared with state-of-the-art methods in the conversation task?
- **RQ3:** How does FacetCRS prick filter bubbles in the CRS?
- **RQ4:** How do the entity-facet user preference  $X_e$ , word-facet  $X_w$ , context-facet  $X_c$ , and review-facet  $X_r$  contribute to the performance?

### Experimental Protocol

**Datasets.** We evaluate our FacetCRS on two challenging CRS-based datasets **REDIAL** (Li et al. 2018) and **TG-REDIAL** (Zhou et al. 2020c). **REDIAL** encompasses a collection of 10,006 conversations associated with 51,699 movies. It is partitioned into three training, validation, and testing datasets, respectively. Additionally, this dataset incorporates a review database sourced from the IMDb<sup>1</sup> website, containing 30 reviews for each movie. **TG-REDIAL** is a Chinese dataset consisting of 10,000 dialogues involving two parties, which collectively comprise a total of 129,392 utterances relevant to 33,834 movies. Each conversation commences with an initial sentence and unfolds chronologically, giving rise to subsequent response utterances and recommendations. The review database integrated within TG-REDIAL is obtained from the Douban<sup>2</sup> website.

<sup>1</sup><https://www.dbpedia.org/>

<sup>2</sup><https://movie.douban.com/>

**Metrics.** To make a fair comparison, we use same metrics as most works (Lu et al. 2021; Zhou et al. 2022). For the recommendation task, we adopt Recall@ $k$  (R@ $k$ , with  $k = 1, 10, 50$ ). For the conversation task, we adopt both automatic evaluation (*i.e.*, Distinct  $n$ -gram (D- $n$ , with  $n = 2, 3, 4$ )) and human evaluation (*i.e.*, *Fluency* and *Informativeness*).

**Baselines.** We conduct a comparative analysis with several baselines in both recommendation and conversation tasks. The compared methods include **Popularity**, **TextCNN** (Kim 2014), **Trans** (Vaswani et al. 2017a), **ReDial** (Li et al. 2018), **KBRD** (Chen et al. 2019), **KGSF** (Zhou et al. 2020a), **KECRS** (Zhang et al. 2021). By conducting a thorough comparison with these state-of-the-art models, we can effectively evaluate the performance and effectiveness of our FacetCRS in the recommendation and conversation tasks.

### Evaluation on Recommendation Task (RQ1)

We utilize the Recall@ $k$  to evaluate our FacetCRS in the recommendation task. Table 1 summarizes the experimental results. We can see that our model is superior to all the compared methods. The best is labeled in boldface.

The improvement of FacetCRS over these baselines can be attributed to three aspects: (1) In the real-world, the data of user interactions is rather sparse, and thus external data is crucial to complement the user behavior data to explore user preference. Thus, we adopt multi-aspect external data, such as DBpedia, ConceptNet, and item reviews, to model diverse user interests; (2) Besides these external data, the dialogue contexts between the user and the system in the CRS also play an important role in excavating users’ dynamic interests and evolving needs. Hence, modeling multifaceted user preferences via this multi-dimension knowledge is beneficial to explore diverse user preferences, which can effectively make recommendations; (3) FacetCRS is designed to prick filter bubbles in the CRS when the user chats with the system over time via natural language conversations. To do this, we adopt the multifaceted user preference  $P_{\text{multifac}}$  to select the important candidate items to recommend to the user in the recommendation module. In this manner, leveraging multifaceted user preferences can effectively prick filter bubbles in the CRS via the user-system feedback loop.

### Evaluation on Conversation Task (RQ2)

**Automatic Evaluation.** Table 2 presents the D@ $n$  results in our FacetCRS on the conversation task. The experimental results unequivocally demonstrate that our proposed FacetCRS attains superior performance compared to all competing methods, consistently outperforming them across various metrics. First, we can see that ReDial outperforms Popularity and TextCNN because it introduces a pre-trained RNN module. Meanwhile, KBRD is superior to KECRS since it incorporates external data (*e.g.*, DBpedia) to align the item and word representations. Additionally, KGSF performs better than other baselines. This is attributed that KGSF not only aligns the conversation contexts and items but also further enhances their representations.

Compared with these baselines, the improvement of FacetCRS can be attributed to several main reasons: (1) We adopt the entity-facet user preference  $X_e$ , word-facet  $X_w$ ,

Datasets	REDIAL			TG-REDIAL		
	R@1	R@10	R@50	R@1	R@10	R@50
Popularity	0.011	0.054	0.183	0.0004	0.003	0.014
TextCNN	0.013	0.068	0.191	0.003	0.010	0.024
ReDial	0.024	0.140	0.320	0.000	0.002	0.013
KBRD	0.031	0.150	0.336	0.005	0.032	0.077
KGSF	0.039	0.183	0.378	0.005	0.030	0.074
KECRS	0.021	0.143	0.340	0.002	0.026	0.069
<b>FacetCRS</b>	<b>0.041*</b>	<b>0.202*</b>	<b>0.386*</b>	<b>0.006*</b>	<b>0.034*</b>	<b>0.080*</b>

Table 1: Results on the recommendation task. Numbers marked with \* indicate that the improvement is statistically significant compared with the best baseline ( $t$ -test with  $p$ -value  $< 0.05$ ).

Datasets	REDIAL			TG-REDIAL		
	D-2	D-3	D-4	D-2	D-3	D-4
Trans	0.067	0.139	0.227	0.053	0.121	0.204
ReDial	0.082	0.143	0.245	0.055	0.123	0.215
KBRD	0.086	0.153	0.265	0.045	0.096	0.233
KGSF	0.114	0.204	0.282	0.086	0.186	0.297
KECRS	0.040	0.090	0.149	0.047	0.114	0.193
<b>FacetCRS</b>	<b>0.126*</b>	<b>0.209*</b>	<b>0.305*</b>	<b>0.113*</b>	<b>0.228*</b>	<b>0.312*</b>

Table 2: Results on the conversation task via automatic evaluation. Numbers marked with \* indicate that the improvement is statistically significant compared with the best baseline ( $t$ -test with  $p$ -value  $< 0.05$ ).

context-facet  $X_c$ , and review-facet  $X_r$  to model feature preferences at different latitudes, which is helpful of modeling user preference into multi-facet, thereby mitigating the filter bubbles in the CRS and make diverse recommendation. (2) To accurately predict the next utterances, we take these four facet user preferences into the cross-attention to obtain the respective context representations to complement each other for generating suitable responses.

**Human Evaluation.** Table 3 shows the experimental results in FacetCRS via human evaluation on the conversation task. From the experimental results, we can observe that: 1) ReDial performs better than Transformer since it has a pre-trained RNN encoder; 2) KGSF outperforms multiple baselines such as ReDial, KBRD, KECRS in terms of *informativeness* metric because it incorporates the external KG to align the word and item representations; 3) From the view of *informativeness*, KGSF also performs better than all other baselines attributed to the fact that it enhances representations via KG.

FacetCRS consistently achieves superior performance in both metrics compared to all the baselines. Our model incorporates various facets, including entity facet, word facet, context facet, and review facet, thereby effectively mitigating filter bubbles in the CRS as the user interacts with the system over time. Additionally, our model generates conversation utterances that are both more informative and fluent via these four facets of knowledge information.

Models	Fluency	Informativeness
Trans	0.97	0.92
ReDial	1.35	1.04
KBRD	1.23	1.15
KGSF	1.48	1.37
KECRS	1.39	1.19
<b>FacetCRS</b>	<b>1.50*</b>	<b>1.39*</b>

Table 3: Results on the conversation task via human evaluation. Numbers marked with \* indicate that the improvement is statistically significant compared with the best baseline ( $t$ -test with  $p$ -value  $< 0.05$ ).

Datasets	REDIAL		TG-REDIAL	
	Iso-Index ↓	Coverage ↑	Iso-Index ↓	Coverage ↑
Popularity	0.1281	5.9110	0.1431	7.0332
TextCNN	0.1127	6.9041	0.1301	8.9254
ReDial	0.1185	6.8202	0.1242	8.9418
KBRD	0.1149	5.8931	0.1222	9.0810
KGSF	0.1055	6.7920	0.1198	9.1373
KECRS	0.1072	6.8157	0.1075	9.0942
<b>FacetCRS</b>	<b>0.0856</b>	<b>7.6436</b>	<b>0.0902</b>	<b>10.8172</b>

Table 4: Results on *Iso-Index* and *Coverage* metrics.

### Study on Filter Bubbles (RQ3)

Since we aim to prick filter bubbles in the CRS, we examine the recommendation results compared with the strongest baselines to verify whether FacetCRS can mitigate the filter bubbles. To this end, we adopt multiple metrics to fully evaluate the severity of filter bubbles.

1) *Isolation-Index (Iso-Index) & Coverage*: In Table 4, we present the experimental results for two widely-used metrics, namely *Iso-Index* and *Coverage*. It is evident that FacetCRS consistently demonstrates the lowest *Iso-Index* values and the highest *Coverage* values across both datasets. The reduced *Iso-Index* values indicate a greater recommendation diversity, encompassing a broader range of facets and topics. Moreover, the higher *Coverage* values confirm its ability to effectively cover a larger portion of the recommendation space. These observations highlight the capability of FacetCRS to effectively mitigate filter bubbles by enhancing the diversification of recommendations.

2) *Item Similarity*: In Figure 2, we evaluate the similarity among the representations of the top 50 recommended items generated by each method on the REDIAL dataset. Heat map with brighter color denotes increased recommendation diversity while darker hues suggest a contrasting scenario. Besides, the presence of more \* markers indicates higher similarity and lower diversity. We observe that KECRS exhibits a darker color and the highest \* magnitudes, indicating lower diversity in its recommended items. Similar characteristics are observed in the KBRD and KGSF methods. Conversely, our approach, FacetCRS, demonstrates superior performance with a noticeably lighter heat map and fewer \* markers. This validates the effectiveness of FacetCRS in pricking filter bubbles with diverse recommendations.

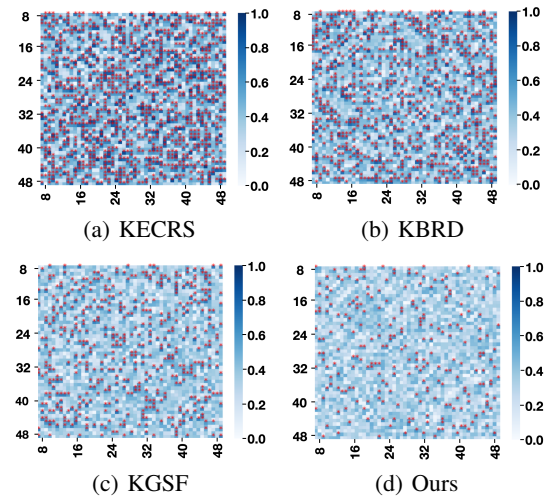


Figure 2: The heat maps depicting item similarity.

Models	D-2	D-3	D-4
<b>FacetCRS</b>	<b>0.126</b>	<b>0.209</b>	<b>0.305</b>
w/o $X_e$	0.118	0.201	0.298
w/o $X_w$	0.122	0.202	0.300
w/o $X_c$	0.119	0.197	0.296
w/o $X_r$	0.123	0.205	0.301

Table 5: Ablation studies on the conversation task.

### Ablation Studies (RQ4)

In this part, we conduct ablation experiments with different variants of our FacetCRS on the REDIAL dataset to verify the contributions of each component, including: 1) w/o  $X_e$ : we remove entity-facet embeddings  $X_e$ ; 2) w/o  $X_w$ : we remove word-facet embeddings  $X_w$ ; 3) w/o  $X_c$ : we remove context-facet embeddings  $X_c$ ; 4) w/o  $X_r$ : we remove review-facet embeddings  $X_r$ . Table 5 shows that removing any kind of various levels of preference facet leads to performance degradation. This validates that these various data are beneficial for enhancing data representations to excavate diverse user preferences, thereby mitigating filter bubbles in the CRS to fulfill the recommendation goal.

## Conclusion

To prick filter bubbles in the CRS, we propose a novel framework FacetCRS, which models multi-faceted user preferences in the CRS, including entity-, word-, context-, and review-facet, to capture diverse user preferences to mitigate filter bubbles. Meanwhile, FacetCRS is an end-to-end framework to automatically learn representations of various levels of preference facet and diverse types of external knowledge. Through extensive experiments, our method consistently outperforms several competitive baselines, which demonstrate the effectiveness of our FacetCRS.

## Acknowledgments

This work is supported in part by the National Key R&D Program of China under Grant No. 2021ZD0111601, National Natural Science Foundation of China (NSFC) under Grant No. 61836012, 62325605, U21A20470, 62206110, 62206314, GuangDong Basic and Applied Basic Research Foundation under Grant No. 2023A1515011374, 2022A1515011835, China Postdoctoral Science Foundation funded project under Grant No. 2021M703687.

## References

- Auer, S.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; and Ives, Z. G. 2007. DBpedia: A Nucleus for a Web of Open Data. In *International Semantic Web Conference/Asian Semantic Web Conference*, volume 4825, 722–735.
- Chen, H.; Liu, X.; Yin, D.; and Tang, J. 2017. A Survey on Dialogue Systems: Recent Advances and New Frontiers. *ACM SIGKDD Explorations*, 19(2): 25–35.
- Chen, Q.; Lin, J.; Zhang, Y.; Ding, M.; Cen, Y.; Yang, H.; and Tang, J. 2019. Towards knowledge-based recommender dialog system. *arXiv preprint arXiv:1908.05391*.
- Christakopoulou, K.; Radlinski, F.; and Hofmann, K. 2016. Towards Conversational Recommender Systems. In *International Conference on Knowledge Discovery and Data Mining*, 815–824.
- Epure, E. V.; and Hennequin, R. 2023. A Human Subject Study of Named Entity Recognition in Conversational Music Recommendation Queries. In *European Chapter of the Association for Computational Linguistics*, 1273–1288.
- Gao, C.; Lei, W.; Chen, J.; Wang, S.; He, X.; Li, S.; Li, B.; Zhang, Y.; and Jiang, P. 2022. CIRS: Bursting Filter Bubbles by Counterfactual Interactive Recommender System. *ACM Transactions on Information Systems*.
- Hayati, S. A.; Kang, D.; Zhu, Q.; Shi, W.; and Yu, Z. 2020. INSPIRED: Toward Sociable Recommendation Dialog Systems. In *Empirical Methods in Natural Language Processing*, 8142–8152.
- Hussein, E.; Juneja, P.; and Mitra, T. 2020. Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube. *ACM on Human-Computer Interaction*, 4(CSCW): 048:1–048:27.
- Kim, Y. 2014. Convolutional Neural Networks for Sentence Classification. In *Empirical Methods in Natural Language Processing (Demonstrations)*, 1746–1751.
- Li, R.; Kahou, S. E.; Schulz, H.; Michalski, V.; Charlin, L.; and Pal, C. 2018. Towards Deep Conversational Recommendations. In *Advances in Neural Information Processing Systems*, 9748–9758.
- Li, S.; Cheng, Q.; Li, L.; and Qiu, X. 2023a. Mitigating Negative Style Transfer in Hybrid Dialogue System. In *Association for the Advancement of Artificial Intelligence*, 13103–13111.
- Li, S.; Xie, R.; Zhu, Y.; Ao, X.; Zhuang, F.; and He, Q. 2022. User-Centric Conversational Recommendation with Multi-Aspect User Modeling. In *Conference on Research and Development in Information Retrieval*, 223–233.
- Li, Z.; Dong, Y.; Gao, C.; Zhao, Y.; Li, D.; Hao, J.; Zhang, K.; Li, Y.; and Wang, Z. 2023b. Breaking Filter Bubble: A Reinforcement Learning Framework of Controllable Recommender System. In *The Web Conference*, 4041–4049. ACM.
- Liu, P.; Shivaram, K.; Culotta, A.; Shapiro, M. A.; and Bilgic, M. 2021. The Interaction between Political Typology and Filter Bubbles in News Recommendation Algorithms. In *The Web Conference*, 3791–3801.
- Liu, Y.; Zhang, W.; Dong, B.; Fan, Y.; Wang, H.; Feng, F.; Chen, Y.; Zhuang, Z.; Cui, H.; Li, Y.; and Che, W. 2023. U-NEED: A Fine-grained Dataset for User Needs-Centric E-commerce Conversational Recommendation. In *Conference on Research and Development in Information Retrieval*, 2723–2732. ACM.
- Lu, Y.; Bao, J.; Song, Y.; Ma, Z.; Cui, S.; Wu, Y.; and He, X. 2021. RevCore: Review-Augmented Conversational Recommendation. In *Findings of the Association for Computational Linguistics*, 1161–1173.
- Mishra, K.; Priya, P.; and Ekbal, A. 2023. Help Me Heal: A Reinforced Polite and Empathetic Mental Health and Legal Counseling Dialogue System for Crime Victims. In *Association for the Advancement of Artificial Intelligence*, 14408–14416.
- Nguyen, T. T.; Hui, P.; Harper, F. M.; Terveen, L. G.; and Konstan, J. A. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *The Web Conference*, 677–686. ACM.
- Qin, L.; Li, Z.; Yu, Q.; Wang, L.; and Che, W. 2023. Towards Complex Scenarios: Building End-to-End Task-Oriented Dialogue System across Multiple Knowledge Bases. In *Association for the Advancement of Artificial Intelligence*, 13483–13491.
- Ribeiro, M. H.; Ottoni, R.; West, R.; Almeida, V. A. F.; and Jr., W. M. 2020. Auditing radicalization pathways on YouTube. In *Conference on Fairness, Accountability, and Transparency*, 131–141. ACM.
- Speer, R.; Chin, J.; and Havasi, C. 2017. ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. In *Association for the Advancement of Artificial Intelligence*, 4444–4451.
- Spinelli, L.; and Crovella, M. 2020. How YouTube Leads Privacy-Seeking Users Away from Reliable Information. In *Conference on User Modeling, Adaptation, and Personalization*, 244–251. ACM.
- Steck, H. 2018. Calibrated recommendations. In *Conference on Recommender Systems*, 154–162.
- Sun, Y.; and Zhang, Y. 2018. Conversational Recommender System. In *Conference on Research & Development in Information Retrieval*, 235–244.
- Tomlein, M.; Pecher, B.; Simko, J.; Srba, I.; Móro, R.; Stefančová, E.; Kompan, M.; Hrcková, A.; Podrouzek, J.; and Bieliková, M. 2021. An Audit of Misinformation Filter Bubbles on YouTube: Bubble Bursting and Recent Behavior Changes. In *Conference on Recommender Systems*, 1–11. ACM.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017a. Attention is All you Need. In *Advances in Neural Information Processing Systems*, 5998–6008.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017b. Attention is All you Need. In *Advances in Neural Information Processing Systems*, 5998–6008.

Wahbeh, A.; Al-Ramahi, M.; El-Gayar, O. F.; Elnoshokaty, A.; and Nasrallah, T. 2023. Conversational Agents for Mental Health and Well-being: Discovering Design Recommendations Using Text Mining. In *Hawaii International Conference on System Sciences*, 3184–3193.

Wang, W.; Feng, F.; Nie, L.; and Chua, T. 2022. User-controllable Recommendation Against Filter Bubbles. In *Conference on Research and Development in Information Retrieval*, 1251–1261. ACM.

Zhang, T.; Liu, Y.; Zhong, P.; Zhang, C.; Wang, H.; and Miao, C. 2021. KE CRS: Towards Knowledge-Enriched Conversational Recommendation System. *CoRR*, abs/2105.08261.

Zhang, Y.; Wu, L.; Shen, Q.; Pang, Y.; Wei, Z.; Xu, F.; Long, B.; and Pei, J. 2022. Multiple Choice Questions based Multi-Interest Policy Learning for Conversational Recommendation. In *The Web Conference*, 2153–2162.

Zhou, K.; Zhao, W. X.; Bian, S.; Zhou, Y.; Wen, J.-R.; and Yu, J. 2020a. Improving conversational recommender systems via knowledge graph based semantic fusion. In *International Conference on Knowledge Discovery and Data Mining*, 1006–1014.

Zhou, K.; Zhao, W. X.; Wang, H.; Wang, S.; Zhang, F.; Wang, Z.; and Wen, J. 2020b. Leveraging Historical Interaction Data for Improving Conversational Recommender System. In *Conference on Information and Knowledge Management*, 2349–2352.

Zhou, K.; Zhou, Y.; Zhao, W. X.; Wang, X.; and Wen, J. 2020c. Towards Topic-Guided Conversational Recommender System. In *International Conference on Computational Linguistics*, 4128–4139.

Zhou, Y.; Zhou, K.; Zhao, W. X.; Wang, C.; Jiang, P.; and Hu, H. 2022. C<sup>2</sup>-CRS: Coarse-to-Fine Contrastive Learning for Conversational Recommender System. In *Web Search and Data Mining*, 1488–1496. ACM.