Why Multi-Interest Fairness Matters: Hypergraph Contrastive Multi-Interest Learning for Fair Conversational Recommender System

Yongsen Zheng[1,2], **Zongxuan Xie**[3], **Guohua Wang**[4†],
Ziyao Liu[1,2], **Liang Lin**[3,5], **Kwok-Yan Lam**[1,2]

[1]Nanyang Technological University, Singapore [2]Digital Trust Centre Singapore
[3]Sun Yat-sen University [4]South China Agricultural University [5]Peng Cheng Laboratory
{yongsen.zheng, liuziyao, kwokyan.lam}@ntu.edu.sg
xiezx25@mail2.sysu.edu.cn, wangguohua@scau.edu.cn, linliang@ieee.org

Abstract

Unfairness is a well-known challenge in Recommender Systems (RSs), often resulting in biased outcomes that disadvantage users or items based on attributes such as gender, race, age, or popularity. Although some approaches have started to improve fairness recommendation in offline or static contexts, the issue of unfairness often exacerbates over time, leading to significant problems like the Matthew effect, filter bubbles, and echo chambers. To address these challenges, we proposed a novel framework, **Hy**pergraph Contrastive Multi-Interest Learning for **Fair C**onversational **R**ecommender **S**ystem (**HyFairCRS**), aiming to promote multi-interest diversity fairness in dynamic and interactive Conversational Recommender Systems (CRSs). HyFairCRS first captures a wide range of user interests by establishing diverse hypergraphs through contrastive learning. These interests are then utilized in conversations to generate informative responses and ensure fair item predictions within the dynamic user-system feedback loop. Experiments on two CRS-based datasets show that HyFairCRS achieves a new state-of-the-art performance while effectively alleviating unfairness. Our code is available at https://github.com/zysensmile/HyFairCRS.

1 Introduction

Conversational Recommender Systems (CRSs) have become an important research area (Zheng et al., 2024c,d; Qian et al., 2023), focusing on understanding user preferences through multi-turn dialogues. However, issues like data imbalance and inconsistent user-item interactions can result in unfair recommendations for disadvantaged users, leading to unsatisfactory suggestions. Thus, it is rather crucial to alleviate the unfairness in the dynamic and interactive CRS.

Recently, there has been a growing interest in examining fairness in Recommender Systems (RSs), which cater to multiple stakeholders (Patro et al., 2020), including customers (user side) and providers (item side). These fairness-aware recommendation methods can be categorized into three main areas. The first category (Bobadilla et al., 2021; Liu et al., 2024; Wu et al., 2021a; Burke et al., 2018) focuses on customer-side fairness, aiming to ensure equitable treatment of users across different groups (*e.g.*, gender, race) to eliminate discrimination in recommendations. The second category (Beutel et al., 2019; Lei et al., 2020a; Ge et al., 2021; Zehlike and Castillo, 2020) addresses provider-side (item-side) fairness, ensuring relatively fair recommendation opportunities for diverse item groups (*e.g.*, popular vs. unpopular) to create a fair platform for all providers. The third category (Xu et al., 2024; Wang et al., 2024; Patro et al., 2020; Wu et al., 2021b) investigates two-sided fairness, which seeks to mitigate bias in recommendations by ensuring that both users and items are treated equitably. This highlights the significance of balanced representation.

Despite their effectiveness, most existing methods still encounter two major challenges. *1) Unfairness Evolution.* Many methods aim to alleviate unfairness in offline or relatively static recommendation settings, often failing to account for the dynamic nature of unfairness in real-world scenarios. In practice, a feedback loop exists between users and the system, which can intensify unfairness over time as users engage with the recommendations. This escalation can result in significant issues, such as the Matthew effect (Liu and Huang, 2021), filter bubbles (Steck, 2018), and echo chambers (Ge et al., 2020). Hence, it is crucial to address unfairness in the dynamic CRS. *2) Diversity Fairness.* Some studies (Zhao et al., 2023a; Wu et al., 2022) have indicated that when RSs negative fairness in user interest diversity and focus exclusively on rele-

---

†Corresponding author.

vance scores, they often produce recommendations that are overly similar or homogeneous. This lack of diversity can have negative consequences for both customers and providers. Customers may experience fatigue and boredom due to a surplus of similar items, which can diminish their long-term satisfaction (Zhao et al., 2023a). On the provider side, less popular items may struggle to gain visibility, overshadowed by a few dominant choices. Thus, investigating diversity fairness related to multiple interests is vital for fostering more balanced and equitable recommendations.

To address these issues, we propose a novel framework, **Hy**pergraph Contrastive Multi-Interest Learning for **Fair C**onversational **R**ecommender **S**ystem (**HyFairCRS**), aimed at exploring diversity fairness related to users' multiple interests in the interactive. HyFairCRS consists of two key components: Hypergraph Contrastive Multi-Interest Learning and Fair CRS. The former seeks to capture a wide range of user interests by creating diverse hypergraphs through contrastive learning. It starts by generating various hypergraphs, *i.e.*, entity-, item-, word-, and review-guided, using different external knowledge sources (*e.g.*, DBpedia (Auer et al., 2007), ConceptNet (Speer et al., 2017), item reviews (Lu et al., 2021)) along with historical conversation data. Each hypergraph is decoupled into a hypergraph and a line graph, which are processed through hypergraph convolution and graph convolution layers to produce multiple representations of user interests. Finally, contrastive learning is used to refine these multi-interest embeddings, enhancing their diversity. The latter leverages these diverse interests to promote fairness in the CRS. Specifically, it integrates varied user interests into conversational tasks to generate informative responses and into recommendation tasks to boost result diversity, ensuring fairness across multi-turn dialogues. Extensive experiments on two widely-used CRS-based datasets show that HyFairCRS achieves a new state-of-the-art performance, validating its superiority of improving fairness.

Overall, our main contributions are included:

- To the best of our knowledge, this is the first work to investigate diversity fairness related to multiple interests in the dynamic CRS that involves a user-system feedback loop.
- We proposed HyFairCRS, a novel end-to-end paradigm that aims to learn multiple user interests to generate informative responses in conversational tasks and to ensure fair item predictions

in recommendation tasks.
- Extensive experimental results on two CRS-based datasets show the effectiveness of HyFairCRS and its superiority of alleviating unfairness.

2 Related Work

2.1 Conversational Recommender System

Conversational Recommender System (Zheng et al., 2024a,b; Qian et al., 2023; Zheng et al., 2024c) assists users in finding products, services, or content through natural language interactions. It has two key components: the recommendation task, which matches items to user preferences, and the conversational task, which generates relevant responses. Currently, CRS-based methods fall into two groups: attributed-based CRS and generation-based CRS. The former (Deng et al., 2021a; Lei et al., 2020b,c; Ren et al., 2021) focuses on leveraging specific attributes of users and items to provide personalized recommendations. These systems typically use structured data, such as user profiles (demographics, preferences) and item features (genre, price, ratings). The latter (Qian et al., 2023; Zhou et al., 2020a; Zheng et al., 2024d; Shang et al., 2023) utilizes natural language generation techniques to create responses and recommendations dynamically. Instead of relying solely on predefined attributes, these systems can generate conversational responses based on context and dialogue history. While both approaches enhance user understanding, they struggle to alleviate unfairness in the CRS as user engagement evolves over time.

2.2 Fair Recommendation

In the context of recommendations (Jin et al., 2023a; Zheng et al., 2023b,a; Jin et al., 2023b; Zheng et al., 2021), fair recommendation aims to ensure equitable outcomes for all users and items, minimizing biases related to sensitive attributes like gender, race, age, and socioeconomic status. The goal is to create systems that provide relevant recommendations while promoting fairness and preventing the reinforcement of existing inequalities. Fairness-aware recommendation systems can be categorized based on their stakeholder focus. 1) Customer-side fairness (Bobadilla et al., 2021; Liu et al., 2024; Wu et al., 2021a; Burke et al., 2018); 2) Provider-side fairness (Beutel et al., 2019; Lei et al., 2020a; Zehlike and Castillo, 2020), and 3) Two-sided fairness (Xu et al., 2024; Wang et al., 2024; Wu et al., 2021b). For example, (Liu et al.,

2024) is proposed to enhance user embeddings with additional interaction data and implements a fairness strategy to improve recommendations for diverse user groups, and (Ge et al., 2021) is devised to improve item exposure equity across groups via a fairness-constrained reinforcement learning algorithm, modeling it as a constrained Markov decision process. However, these methods overlook the negative impact of user-system feedback loop on fairness.

3 HyFairCRS

Enhancing fairness in the CRS is challenging, as unfairness can worsen over time through dynamic interactions. To this end, we propose HyFairCRS, which combines Hypergraph Contrastive Multi-Interest Learning with fair CRS. The pipeline of HyFairCRS is shown in Fig.1.

3.1 Preliminaries

Hypergraph. A hypergraph is defined as $\mathcal{G}_*^{(t)} = (\mathcal{V}_*^{(t)}, \mathcal{E}_*^{(t)}, \mathbf{H}_*^{(t)})$, where $* \in \{e, i, w, r\}$ represents entity, item, word, or review. Here, $\mathcal{V}_*^{(t)}$, $\mathcal{E}_*^{(t)}$, and $\mathbf{H}_*^{(t)}$ are the sets of nodes, hyperedges, and the weighted adjacency matrix, respectively. Let $\mathbf{V}_*^{(t)} \in \mathbb{N}^{|\mathcal{V}_*^{(t)}| \times |\mathcal{V}_*^{(t)}|}$ and $\mathbf{E}_*^{(t)} \in \mathbb{N}^{|\mathcal{E}_*^{(t)}| \times |\mathcal{E}_*^{(t)}|}$ be diagonal matrices of vertex and edge degrees.

Line Graph. A line graph for each hypergraph is written as $\mathcal{L}_*^{(t)} = (\widetilde{\mathcal{V}}_*^{(t)}, \widetilde{\mathcal{E}}_*^{(t)}, \widetilde{\mathbf{H}}_*^{(t)})$ to capture hyperedge-level information. $\widetilde{\mathcal{V}}_*^{(t)}$ corresponds to hyperedges from the original hypergraph, connected if they share a common node. $\widetilde{\mathcal{E}}_*^{(t)}$ consists of pairs of hyperedges sharing at least one node, while $\widetilde{\mathbf{H}}_*^{(t)}$ is the incidence matrix representing relationships between hyperedges and their nodes.

3.2 Hypergraph Contrative Multi-Interest Learning

To model diverse preferences to address the unfairness issue (Zhao et al., 2023a; Liu et al., 2024), we devise hypergraphs to effectively capture high-order user relationship patterns as a hypergraph improves traditional graphs by linking multiple vertices through hyperedges, allowing us to uncover and learn diverse user interests.

3.2.1 Multi-Hypergraph Construction

To explore fairness related to multiple user interests, we establish a set of hypergraphs, including entity-, item-, word-, and review-guided hypergraphs.

1) Entity-Guided Hypergraph: We harness entities from historical sessions to extract multi-hop neighbors from the extensive DBpedia Knowledge Graph (KG) (Auer et al., 2007), enabling the construction of hyperedges linked by shared entities. Subsequently, the hyperedges are interconnected via common entities, culminating in the formation of an entity-guided hypergraph. *2) Item-Guided Hypergraph:* It centers on items from previous conversations, treating them as nodes and forming hyperedges that connect through shared items. All hyperedges are linked together by the items in common to build item-guided hypergraph. This enables us to explore diverse users' interests towards different items. *3) Word-Guided Hypergraph:* Keywords reflect users' dynamic needs, and we use items from conversations as keywords and identify their multi-hop neighbors in the ConceptNet KG (Speer et al., 2017). This forms interconnected hyperedges, linked by common words to create the hypergraph. *4) Review-Guided Hypergraph:* Item reviews highlight the importance of integrating user feedback and comments to improve recommendation performance. We utilize items from each conversation to search positive and negative words in the reviews, thereby creating distinct hyperedges. Formally, they can be given as:

$$\begin{cases} \mathcal{G}_e^{(t)} = (\mathcal{V}_i^{(t)} \cup \mathcal{V}_{n_d}^{(t)}, \mathcal{E}_e^{(t)}, \mathbf{H}_e^{(t)}); \\ \mathcal{G}_i^{(t)} = (\mathcal{V}_i^{(t)}, \mathcal{E}_i^{(t)}, \mathbf{H}_i^{(t)}); \\ \mathcal{G}_w^{(t)} = (\mathcal{V}_i^{(t)} \cup \mathcal{V}_{n_c}^{(t)}, \mathcal{E}_w^{(t)}, \mathbf{H}_w^{(t)}); \\ \mathcal{G}_r^{(t)} = (\mathcal{V}_i^{(t)} \cup \mathcal{V}_{n_r}^{(t)}, \mathcal{E}_r^{(t)}, \mathbf{H}_r^{(t)}). \end{cases} \tag{1}$$

Here $\mathcal{V}_i^{(t)}$, $\mathcal{V}_{n_d}^{(t)}$, $\mathcal{V}_{n_c}^{(t)}$, and $\mathcal{V}_{n_r}^{(t)}$ represent nodes from historical conversations, DBpedia KG, nodes from ConceptNet KG, and item reviews, respectively. Additionally, $\mathcal{E}_e^{(t)}$, $\mathcal{E}_i^{(t)}$, $\mathcal{E}_w^{(t)}$, and $\mathcal{E}_r^{(t)}$ denote the entity-, item-, word-, and review-guided hyperedge sets. The matrices $\mathbf{H}_*^{(t)}$ correspond to their incidence matrices, defined as follows:

$$\mathbf{H}_{*(v,h)}^{(t)} = \begin{cases} 1, & \text{if } v \in h \\ 0, & \text{if } v \notin h \end{cases} \tag{2}$$

For each node v, its degree is given by $d(v) = \sum_{h \in \mathcal{E}_*^{(t)}} \mathbf{H}_{*(v,h)}^{(t)}$; for each hyperedge h, its degree is defined as $\delta(h) = \sum_{v \in \mathcal{V}_*^{(t)}} \mathbf{H}_{*(v,h)}^{(t)}$.

3.2.2 Line Graph Induction

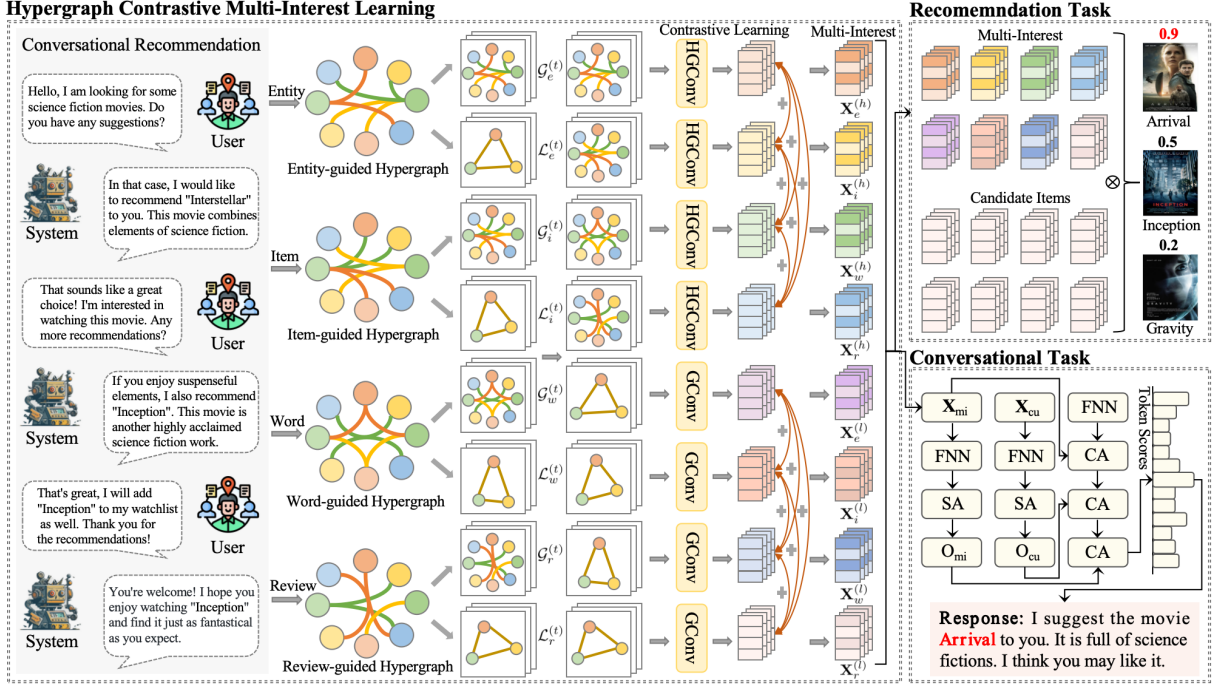Since the hypergraph captures node-level information (intra-hyperedge), we create a correspond-

Figure 1: Overview of our HyFairCRS, which consists of Hypergraph Contrastive Learning and Fair CRS. The former aims to learn multiple user interests to explore diversity fairness, while the latter adopts these diverse interests to generate responses in the conversation task, and to make fair item predictions in the recommendation task.

ing line graph for each hypergraph to explore hyperedge-level relationships (inter-hyperedge). This approach not only captures local messages but also explores global information. The interaction between these two types of information deepens our understanding of diverse user preferences. Formally, the line graphs associated with all hypergraphs can be expressed as:

$$
\begin{cases}
\mathcal{L}_e^{(t)} = (\widetilde{\mathcal{V}}_i^{(t)} \cup \widetilde{\mathcal{V}}_{n_d}^{(t)}, \widetilde{\mathcal{E}}_e^{(t)}, \mathbf{L}_e^{(t)}); \\
\mathcal{L}_i^{(t)} = (\widetilde{\mathcal{V}}_i^{(t)}, \widetilde{\mathcal{E}}_i^{(t)}, \mathbf{L}_i^{(t)}); \\
\mathcal{L}_w^{(t)} = (\widetilde{\mathcal{V}}_i^{(t)} \cup \widetilde{\mathcal{V}}_{n_c}^{(t)}, \widetilde{\mathcal{E}}_w^{(t)}, \mathbf{L}_w^{(t)}); \\
\mathcal{L}_r^{(t)} = (\widetilde{\mathcal{V}}_i^{(t)} \cup \widetilde{\mathcal{V}}_{n_r}^{(t)}, \widetilde{\mathcal{E}}_r^{(t)}, \mathbf{L}_r^{(t)}).
\end{cases}
\tag{3}
$$

Here $\widehat{\mathcal{V}}_*^{(t)}$ contains nodes v_e for hyperedges e in $\mathcal{E}_*^{(t)}$. The set $\widehat{\mathcal{E}}_*^{(t)}$ includes pairs (v_{e_p}, v_{e_q}) for hyperedges e_p and e_q that share at least one node. The incidence matrix $\mathbf{L}_*^{(t)}$ defines the structure of $\mathcal{L}_*^{(t)}$, and we assign each edge (v_{e_p}, v_{e_q}) a weight $W_{p,q}$, which can be expressed as:

$$
W_{p,q} = \frac{|e_p \cap e_q|}{|e_p \cup e_q|}.
\tag{4}
$$

3.2.3 Contrastive Multi-Interest Learning

Once we have generated a series of hypergraphs along with their corresponding line graphs, we shift our focus to examining the diverse interests of users through various contrastive learning paradigms. To streamline this exploration, we categorize user interests into two distinct types: hypergraph-oriented interests representations and line graph-oriented interest representations. To obtain these diverse representations, we input all hypergraphs (*i.e.*, $\mathcal{G}_*^{(t)}$) into the Hypergraph Convolution Network (HGConv) to learn the intra-hyperedge representations. Simultaneously, we feed all line graphs (*i.e.*, $\mathcal{L}_*^{(t)}$) into the Graph Convolution Network (GConv) to learn the inter-hyperedge representations. This dual strategy enables us to derive a set of primary feature representations for user interests, defined as:

$$
\begin{cases}
\mathbf{O}_*^{(l+1)} = \mathbf{V}_*^{-1} \mathbf{H}_*^{(t)} \mathbf{W}_*^{(l)} \mathbf{E}_*^{-1} \mathbf{H}_*^{(t)^T} \mathbf{O}_*^{(l)}; \\
\mathbf{X}_*^{(h)} = \mathbf{O}_*^{(L)} = \mathbf{V}_*^{-1} \mathbf{H}_*^{(t)} \mathbf{W}_*^{(L-1)} \mathbf{E}_*^{-1} \mathbf{H}_*^{(t)^T} \mathbf{O}_*^{(L-1)}.
\end{cases}
\tag{5}
$$

$$
\begin{cases}
\widehat{\mathbf{O}}_*^{(l+1)} = \sigma(\widehat{\mathbf{V}}_*^{-\frac{1}{2}} \widehat{\mathbf{N}}_* \widehat{\mathbf{V}}_*^{-\frac{1}{2}} \widehat{\mathbf{O}}_*^{(l)} \Theta_*^{(l)}); \\
\mathbf{X}_*^{(l)} = \widehat{\mathbf{O}}_*^{(\widehat{L})} = \sigma(\widehat{\mathbf{V}}_*^{-\frac{1}{2}} \widehat{\mathbf{N}}_* \widehat{\mathbf{V}}_*^{-\frac{1}{2}} \widehat{\mathbf{O}}_*^{(\widehat{L}-1)} \Theta_*^{(\widehat{L}-1)}).
\end{cases}
\tag{6}
$$

In HGConv, $\mathbf{O}_*^{(l)}$ and $\mathbf{O}_*^{(l+1)}$ denote the output embeddings for layers l and $l+1$, with $\mathbf{W}_*^{(l)}$ representing the trainable parameters. Information is aggregated from nodes to hyperedges using $\mathbf{H}_*^{(t)^T} \mathbf{O}_*^{(l)}$ and from hyperedges back to nodes through the

premultiplication of $\mathbf{H}_*^{(t)}$. In GConv, the input and output feature matrices for layer l are $\widehat{\mathbf{O}}_*^{(l)}$ and $\widehat{\mathbf{O}}_*^{(l+1)}$. The adjacency matrix of the line graph is defined as $\widehat{\mathbf{N}}_* = \widehat{\mathbf{H}}_* + \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. The degree matrix $\widehat{\mathbf{V}}_*^{-\frac{1}{2}}$ is computed with $\widehat{\mathbf{V}}_{*(p,p)} = \sum_{q=1}^{m} \widehat{\mathbf{N}}_{*(p,q)}$, where m is the number of nodes per hyperedge. For simplicity, we consider the outputs from the final layers L and $\widehat{L}$ as the hypergraph-oriented and line graph-oriented interest representations, respectively.

Next, we employ contrastive learning to refine both hypergraph-guided interest representations (*i.e.*, $\mathbf{X}_*^{(h)}$), and line graph-guided interest representations (*i.e.*, $\mathbf{X}_*^{(l)}$). The primary reason for adopting contrastive learning is its ability to enhance the robustness of learned representations, making them invariant to different views of the same instance by attracting positive pairs and separating negative pairs. Firstly, we employ the InfoNCE (Shang et al., 2023) contrastive loss to refine any pairs of hypergraph-guided interest representations as:

$$\mathcal{J}_{(i,e)}^{\mathcal{H}} = -\log \frac{\exp(\text{sim}(\mathbf{X}_e^{(h)}, \mathbf{X}_i^{(h)^+})/\tau)}{\sum_{k=0}^{K} \exp(\text{sim}(\mathbf{X}_e^{(h)}, \mathbf{X}_i^{(k)^-})/\tau)},$$
(7)

where $\text{sim}(\cdot)$ and τ are the cosine similarity function and temperature hyperparameter, respectively. $\mathbf{X}_i^{(h)^+}$ is the positive sample matched with $\mathbf{X}_i^{(h)}$, while $\mathbf{X}_i^{(k)^-}$ refers to k negative samples. Analogously, we define the contrastive losses for other pairs of hypergraph-guided representations, namely $(\mathbf{X}_e^{(h)}, \mathbf{X}_w^{(h)})$, $(\mathbf{X}_e^{(h)}, \mathbf{X}_r^{(h)})$, $(\mathbf{X}_i^{(h)}, \mathbf{X}_w^{(h)})$, $(\mathbf{X}_i^{(h)}, \mathbf{X}_r^{(h)})$, and $(\mathbf{X}_w^{(h)}, \mathbf{X}_r^{(h)})$ as $\mathcal{J}_{(e,w)}^{\mathcal{H}}$, $\mathcal{J}_{(e,r)}^{\mathcal{H}}$, $\mathcal{J}_{(i,w)}^{\mathcal{H}}$, $\mathcal{J}_{(i,r)}^{\mathcal{H}}$, and $\mathcal{J}_{(w,r)}^{\mathcal{H}}$, respectively. The total loss is obtained by summing the hypergraph contrastive losses across any two pairs:

$$\mathcal{J}_{CL}^{\mathcal{H}} = \mathcal{J}_{(e,i)}^{\mathcal{H}} + \mathcal{J}_{(w,i)}^{\mathcal{H}} + \mathcal{J}_{(r,i)}^{\mathcal{H}} + \mathcal{J}_{(i,w)}^{\mathcal{H}} + \mathcal{J}_{(i,r)}^{\mathcal{H}} + \mathcal{J}_{(w,r)}^{\mathcal{H}}.$$
(8)

Similarly, we can derive the final contrastive loss for the line graph-based interest representations using Eq.(7) and Eq.(8) as $\mathcal{J}_{CL}^{\mathcal{L}}$. To obtain the overall contrastive loss, we average the losses from both the hypergraph-based and line graph-based interest representations as:

$$\mathcal{J}_{CL} = \mathcal{J}_{CL}^{\mathcal{H}} + \mathcal{J}_{CL}^{\mathcal{L}}.$$
(9)

3.3 Fair CRS

To ensure fair recommendation, we next apply these interest representations into the CRS.

3.3.1 Multi-Interest Fairness

Most existing methods (Zhao et al., 2023a) indicate that exploring diversity fairness related to multiple user interests can mitigate unfairness. Thus, we integrate hypergraph- and line graph-guided interest representations into both recommendation and conversational tasks. We first concatenate all interest representations, then apply average pooling ($\text{Pooling}(\cdot)$) and Multi-Head Attention ($\text{MHA}(\cdot)$) (Vaswani et al., 2017) to obtain the fair representations $\mathbf{X}_{\text{fair\_reco}}$ and $\mathbf{X}_{\text{fair\_conv}}$ for item prediction and response generation, respectively. Formally, this procedure can be written as:

$$\mathbf{X}_{\text{fair}} = [\mathbf{X}_e^{(h)}; \mathbf{X}_i^{(h)}; \mathbf{X}_w^{(h)}; \mathbf{X}_r^{(h)}; \mathbf{X}_e^{(l)}; \mathbf{X}_i^{(l)}; \mathbf{X}_w^{(l)}; \mathbf{X}_r^{(l)}],$$
$$\mathbf{X}_{\text{fair\_reco}} = \text{Pooling}([\text{Pooling}(\mathbf{X}_{\text{fair}}); \mathbf{X}_{\text{curr}}]),$$
$$\mathbf{X}_{\text{fair\_conv}} = \text{MHA}(\mathbf{X}_{\text{curr}}, \mathbf{X}_{\text{fair}}, \mathbf{X}_{\text{fair}}),$$
(10)

where ; denote concatenation, and $\mathbf{X}_{\text{curr}}$ represent current conversation representations.

3.3.2 Recommendation Task

The recommendation task aims to suggest relevant items based on user preferences, enhancing satisfaction and engagement. To ensure fairness, we use $\mathbf{X}_{\text{fair\_reco}}$ to make fair item prediction. Let $\mathbf{I}_{\text{cand}}$ be the feature representations of candidate items in the set $\mathcal{I}$. The recommendation probabilities for these items are given as:

$$\mathbf{P}_{\text{rec}} = \text{Softmax}(\mathbf{X}_{\text{fair\_reco}} \cdot \mathbf{I}_{\text{cand}}).$$
(11)

Then, we define the cross-entropy loss $\mathcal{J}_R$ as our main objective, supported by the contrastive loss $\mathcal{J}_{\text{CL}}$ to enhance recommendations. This allows us to effectively learn user preferences while ensuring diverse suggestions, as illustrated below:

$$\mathcal{J}_R = -\sum_{j=1}^{\mathcal{B}} \sum_{i=1}^{|\mathcal{I}|} [y_{ij} \log(\mathbf{P}_{\text{rec}}^{(j)}(i)) + (1 - y_{ij}) \log(1 - \mathbf{P}_{\text{rec}}^{(j)}(i))];$$
(12)

$$\mathcal{J}_{\text{CL\_R}} = \alpha \mathcal{J}_{\text{CL}} + \mathcal{J}_R,$$
(13)

where $\mathcal{B}$ is the mini-batch size, $|\mathcal{I}|$ is the item set size, $y_{ij} \in [0, 1]$ is the target label, and α controls the contrastive loss impact.

3.3.3 Conversational Task

The conversational task is responsible for generating recommendation responses when the system chats with users via natural language dialogues. We

incorporate fair conversation-based representations $\mathbf{X}_{\text{fair\_conv}}$ in a Transformer-based encoder-decoder framework to generate informative and diverse responses. If $\mathbf{T}^{(n-1)}$ is the previous output, then the current output $\mathbf{T}^{(n)}$ is:

$$
\begin{aligned}
\mathbf{A}_0^n &= \text{MHA}(\mathbf{T}^{n-1}, \mathbf{T}^{n-1}, \mathbf{T}^{n-1}), \\
\mathbf{A}_1^n &= \text{MHA}(\mathbf{A}_0^n, \mathbf{X}_{\text{fair\_conv}}, \mathbf{X}_{\text{fair\_conv}}), \\
\mathbf{A}_2^n &= \text{MHA}(\mathbf{A}_1^n, \mathbf{X}_{\text{curr}}, \mathbf{X}_{\text{curr}}), \\
\mathbf{A}_3^n &= \text{MHA}(\mathbf{A}_2^n, \mathbf{X}_{\text{hist}}, \mathbf{X}_{\text{hist}}), \\
\mathbf{A}_4^n &= \gamma \cdot \mathbf{A}_2^n + (1 - \gamma) \cdot \mathbf{A}_3^n, \\
\mathbf{T}^n &= \text{FFN}(\mathbf{A}_4^n),
\end{aligned}
\tag{14}
$$

where $\mathbf{X}_{\text{curr}}$ and $\mathbf{X}_{\text{hist}}$ represent the features of the current and historical conversations, γ regulates the information flow, and $\text{FFN}(\cdot)$ denotes a fully-connected network. To enhance response diversity, the probability of generating the next token, given the previous sequence $\{s_{t-1}\} = s_1, s_2, \ldots, s_{t-1}$, can be defined as:

$$
\mathsf{P}_{\text{con}}(s_t | \{s_{t-1}\}) = \sum_{j=1}^{3} \mathcal{P}_j(s_t | \mathcal{X}_j); \tag{15}
$$

Here $\mathcal{X}_1 = T_i$, $\mathcal{X}_2 = \mathbf{X}_{\text{fair\_conv}}$, $\mathcal{X}_3 = (T_i, \mathbf{X}_{\text{fair\_conv}})$. T_i is the output generated by the decoder, while s_t denotes the t-th utterance in the sequence. $\mathcal{P}_1(\cdot)$ is vocabulary probability based on the input T_i, $\mathcal{P}_2(\cdot)$ means vocabulary bias from $\mathbf{X}_{\text{fair\_conv}}$, and $\mathcal{P}_3(\cdot)$ indicates copy probability, with non-item vocabularies scored as zero. We train the model using cross-entropy loss with a contrastive loss as:

$$
\mathcal{J}_{\text{C}} = -\sum_{b=1}^{\mathcal{B}} \sum_{t=1}^{\mathcal{T}} \log(\mathsf{P}_{\text{con}}(s_t | \{s_{t-1}\})); \tag{16}
$$

$$
\mathcal{J}_{\text{CL\_C}} = \beta \mathcal{J}_{\text{CL}} + \mathcal{J}_{\text{C}}. \tag{17}
$$

Here $\mathcal{T}$ is the truncated length of utterances, and β is to control the contrastive loss.

3.4 Model Complexity Analysis

The complexity of HyFairCRS involves four main tasks. 1) Hypergraph construction: $O(n)$; 2) Hypergraph convolution: $O(n)$; 3) Contrastive learning: $O(n \times d)$; 4) Line graph construction: $O(n^2 \times e)$ where n, d are batch size and feature dimension.

The first three tasks have **linear complexity**, and we optimized the fourth task to achieve $O(nk \log k)$ through: a) Set Operation Optimization: Using NumPy's 'intersect1d' and 'union1d' reduced 'Jaccard' index computation from $O(e)$ to

$O(k \log k)$; b) JIT Compilation: Numba's 'njit' enhances speed; c) Loop Parallelization: 'prange' enables multi-core execution. After optimization, the theoretical complexity of HyFairCRS is $O\left(\frac{n^2 k \log k}{p}\right)$, where p is the number of CPU cores. On multi-core machines, p can often approach n (*e.g.*, on our 128-core machine), simplifying the complexity to $O(nk \log k)$. Compared with the strongest baseline **HiCore** (Zheng et al., 2024d), has a complexity of $O(n^3)$ because it requires identifying all distinct triangle motifs to construct hyperedges, involving the enumeration of all triplets. Our method outperforms HiCore in both accuracy and efficiency.

4 Experiments and Analyses

To fully evaluate our HyFairCRS, we conduct experiments to answer the following questions:

- **RQ1:** How does HyFairCRS perform compared with all baselines in the recommendation task?
- **RQ2:** How does HyFairCRS perform compared with all baselines in the conversation task?
- **RQ3:** How does HyFairCRS alleviate the unfairness in the CRS?
- **RQ4:** How does each type of hypergraph contribute to the performance?
- **RQ5:** How does HyFairCRS perform on the cross-domain datasets?
- **RQ6:** How do parameters affect our HyFairCRS?

4.1 Experimental Protocol

Datasets. We evaluate our HyFairCRS on two challenging public CRS-based benchmarks: **REDIAL** (Li et al., 2018b) and **TG-REDIAL** (Zhou et al., 2020b). Reviews of REDIAL and TG-REDIAL are sourced from IMDb and Douban, respectively. To further evaluate our proposed HyFairCRS, we incorporate additional cross-domain datasets, including **OpenDialKG** (Moon et al., 2019) and **DuRecDial** (Liu et al., 2021). These datasets encompass a variety of domains, such as movies, music, sports, books, news, culinary experiences, and restaurants. **Baselines.** To fully evaluate our HyFairCRS, we compared with a series of state-of-the-art methods, including **TextCNN** (Kim, 2014), **SASRec** (Kang and McAuley, 2018), **BERT4Rec** (Sun et al., 2019), **Transformer** (Vaswani et al., 2017), **ReDial** (Li et al., 2018a), **KBRD** (Chen et al., 2019), **KGSF** (Zhou et al., 2020a), **KGConvRec** (Sarkar

---

https://www.dbpedia.org/
https://movie.douban.com/

Model	REDIAL						TG-REDIAL					
	R@10	R@50	M@10	M@50	N@10	N@50	R@10	R@50	M@10	M@50	N@10	N@50
TextCNN	0.0644	0.1821	0.0235	0.0285	0.0328	0.0580	0.0097	0.0208	0.0040	0.0045	0.0053	0.0077
SASRec	0.1117	0.2329	0.0540	0.0593	0.0674	0.0936	0.0043	0.0178	0.0011	0.0017	0.0019	0.0047
BERT4Rec	0.1285	0.3032	0.0475	0.0555	0.0663	0.1045	0.0043	0.0226	0.0013	0.0020	0.0020	0.0058
ReDial	0.1705	0.3077	0.0677	0.0738	0.0925	0.1222	0.0038	0.0165	0.0012	0.0017	0.0018	0.0045
TG-ReDial	0.1679	0.3327	0.0694	0.0771	0.0924	0.1286	0.0110	0.0174	0.0048	0.0050	0.0062	0.0076
KBRD	0.1796	0.3421	0.0722	0.0800	0.0972	0.1333	0.0201	0.0501	0.0077	0.0090	0.0106	0.0171
KGSF	0.1785	0.3690	0.0705	0.0796	0.0956	0.1379	0.0215	0.0643	0.0069	0.0087	0.0103	0.0194
BERT	0.1608	0.3525	0.0597	0.0688	0.0831	0.1255	0.0040	0.0194	0.0011	0.0017	0.0018	0.0050
XLNet	0.1569	0.3590	0.0583	0.0677	0.0811	0.1255	0.0040	0.0187	0.0011	0.0017	0.0017	0.0048
BART	0.1693	0.3783	0.0646	0.0744	0.0888	0.1350	0.0047	0.0187	0.0012	0.0017	0.0020	0.0048
KGConvRec	0.1819	0.3587	0.0711	0.0794	0.0969	0.1358	0.0220	0.0524	0.0088	0.0102	0.0119	0.0185
MHIM	0.1966	0.3832	0.0742	0.0830	0.1027	0.1440	0.0300	0.0783	0.0108	0.0129	0.0152	0.0256
HiCore	0.2192	0.4163	0.0775	0.0874	0.1107	0.1558	0.0270	0.0769	0.0088	0.0107	0.0152	0.0225
HyFairCRS*	**0.2237**	**0.4382**	**0.0798**	**0.0899**	**0.1297**	**0.1590**	**0.0358**	**0.0883**	**0.0156**	**0.0179**	**0.0174**	**0.0290**

Table 1: Recommendation results. * indicates statistically significant improvement ($p < 0.05$) over all baselines.

et al., 2020), **BERT** (Devlin et al., 2019), **XLNet** (Yang et al., 2019), **BART** (Lewis et al., 2020), **DialoGPT** (Zhang et al., 2020), **GPT-3** (Brown et al., 2020), **C2-CRS** (Zhou et al., 2022), **LOT-CRS** (Zhao et al., 2023b), **HiCore** (Zheng et al., 2024d), **UniCRS** (Deng et al., 2021b), **MHIM** (Shang et al., 2023), and **HyCoRec** (Zheng et al., 2024c).

4.2 Recommendation Performance (RQ1)

For the recommendation task, we use several metrics: Recall@K (R@K), Mean Reciprocal Rank@K (MRR@K), and Normalized Discounted Cumulative Gain@K (NDCG@K) for K=10 and 50 (Shang et al., 2023; Zheng et al., 2024d). HyFairCRS consistently outperforms all baselines, as shown in Table 1. Its superior performance is attributed to: 1) Diverse Hypergraphs: our HyFairCRS constructs hypergraphs that include various dimensions (emphi.e., entities, words, items, and reviews) for capturing the complexity of user interests and enabling the identification of intricate relationships for more relevant suggestions. 2) Decoupled Structure: each hypergraph is split into a hypergraph and a line graph. The hypergraph focuses on intra-hyperedge relationships, while the line graph examines inter-hyperedge connections, enhancing flexibility and effectively capturing complex user-item interactions. 3) Multiple Interest Representations : the model features eight distinct interest representations, each targeting specific user preferences, leading to more personalized recommendations and a better user experience.

4.3 Conversational Performance (RQ2)

For the conversation task, we also employ several metrics (Zheng et al., 2024b), including Distinct n-gram (Dist-n), BLEU-n for $n \in [2, 3, 4]$. Ta-

Model	REDIAL				TG-REDIAL			
	Dist-2	Dist-3	BLEU-2	BLEU-3	Dist-2	Dist-3	BLEU-2	BLEU-3
ReDial	0.0214	0.0659	0.0198	0.0054	0.2178	0.5136	0.0387	0.0094
Trans.	0.0538	0.1574	0.0164	0.0027	0.2362	0.7063	0.0335	0.0075
KBRD	0.0765	0.3344	0.0203	0.0061	0.8013	1.7840	0.0411	0.0107
KGSF	0.0572	0.2483	0.0211	0.0067	0.3891	0.8868	0.0442	0.0128
GPT-3	0.3604	0.6399	0.0225	0.0090	1.2255	2.5713	0.0438	0.0124
DialoGPT	0.3542	0.6209	0.0220	0.0085	1.1881	2.4269	0.0440	0.0126
UniCRS	0.2464	0.4273	0.0218	0.0080	0.6252	2.2352	0.0438	0.0126
C2-CRS	0.2623	0.3891	0.0223	0.0088	0.5235	1.9961	0.0434	0.0120
LOT-CRS	0.3312	0.6155	0.0222	0.0085	0.9287	2.4880	0.0428	0.0123
MHIM	0.3278	0.6204	0.0226	0.0089	1.1100	2.3520	0.0435	0.0118
CoMoRec	0.3492	0.6480	0.0230	0.0089	1.1273	2.5063	0.0437	0.0125
HyCoRec	0.3661	0.6434	0.0228	0.0092	1.2590	2.6000	0.0440	0.0126
HiCore	0.5871	1.1170	0.0231	0.0093	2.8610	5.7440	0.0441	0.0125
HyFairCRS*	**0.6022**	**1.2270**	**0.0233**	**0.0098**	**2.9390**	**5.8360**	**0.0445**	**0.0128**

Table 2: Conversation results. * indicates statistically significant improvement ($p < 0.05$) over all baselines.

ble 2 demonstrate that HyFairCRS consistently outperforms all baselines. Key observations include:1) LLM-guided and KG-based methods (*e.g.*, DialoGPT, GPT-3, MHIM, CoMoRec, HyCoRec, HiCore) generally perform better due to their ability to leverage rich knowledge for learning interest representations; 2) Token-sequence-aware approaches (*e.g.*, Transformer, ReDial) show the weakest results, as they focus solely on token sequences and miss user preferences embedded in entities. 3) HyFairCRS outperforms debiasing-based methods (*e.g.*, HyCoRec) by utilizing multiple external large-scale KGs and item reviews to create multidimensional hypergraphs (entity-, item-, word-, and review-guided). This approach effectively captures diverse user interests, enhancing recommendation diversity and enriching dialogue content and responses.

4.4 Study on Fairness (RQ3)

To assess if HyFairCRS reduces unfairness, we use several fairness metrics (Jin et al., 2023a): Average

Model	REDIAL								TG-REDIAL							
	A@5↓	A@10↓	A@15↓	A@20↓	G@5↓	G@10↓	G@15↓	G@20↓	A@5↓	A@10↓	A@15↓	A@20↓	G@5↓	G@10↓	G@15↓	G@20↓
KBRD	0.0062	0.0099	0.0131	0.0158	0.8282	0.8494	0.8533	0.8534	0.0181	0.0236	0.0291	0.0325	0.8265	0.8156	0.8020	0.7948
KGSF	0.0062	0.0103	0.0135	0.0166	0.8291	0.8472	0.8529	0.8518	0.0188	0.0246	0.0291	0.0321	0.8504	0.8494	0.8413	0.8370
KGConvRec	0.0063	0.0099	0.0132	0.0162	0.8366	0.8534	0.8297	0.8539	0.0161	0.0227	0.0265	0.0303	0.8681	0.8328	0.8188	0.8063
MHIM	0.0038	0.0056	0.0071	0.0084	0.7962	0.8198	0.8569	0.8323	0.0024	0.0034	0.0043	0.0051	0.6858	0.7078	0.7129	0.7185
HyCoRec	0.0030	0.0045	0.0057	0.0068	0.7653	0.7961	0.8063	0.8117	0.0012	0.0017	0.0021	0.0025	0.5325	0.5789	0.5993	0.6105
HiCore	0.0031	0.0045	0.0056	0.0067	0.7836	0.8129	0.8254	0.8302	0.0011	0.0017	0.0021	0.0026	0.4892	0.5380	0.5617	0.5734
HyFairCRS	**0.0028**	**0.0040**	**0.0051**	**0.0061**	**0.7163**	**0.7504**	**0.7645**	**0.7711**	**0.0010**	**0.0015**	**0.0018**	**0.0022**	**0.4825**	**0.5313**	**0.5545**	**0.5666**
	L@5↓	L@10↓	L@15↓	L@20↓	D@5↑	D@10↑	D@15↑	D@20↑	L@5↓	L@10↓	L@15↓	L@20↓	D@5↑	D@10↑	D@15↑	D@20↑
KBRD	5.1706	5.6387	5.8862	6.0446	0.3967	0.3890	0.3908	0.5557	5.5726	5.7429	5.8589	5.9245	0.2262	0.1759	0.2102	0.2594
KGSF	5.1796	5.6642	5.9170	6.0768	0.3540	0.3509	0.3819	0.4010	5.7729	6.0155	6.1143	6.1734	0.1794	0.1796	0.2092	0.2558
KGConvRec	5.2627	5.6834	5.9335	6.0751	0.3535	0.3649	0.3692	0.5320	5.8173	5.8702	5.9034	5.9428	0.1259	0.1501	0.1932	0.2366
MHIM	4.5794	5.0229	5.2573	5.4124	0.6067	0.6499	0.6700	0.6624	3.0119	3.3942	3.6156	3.7780	0.5490	0.6106	0.6385	0.6632
HyCoRec	4.1947	4.6710	4.9217	5.0967	0.6198	0.6634	0.6754	0.6802	1.8647	2.2999	2.5614	2.7640	0.6814	0.7099	0.7394	0.7548
HiCore	4.3901	4.8461	5.0915	5.2607	0.6472	0.6804	0.6870	0.7002	1.6797	2.1458	2.4333	2.6492	0.6871	0.7137	0.7359	0.7484
HyFairCRS	**3.5946**	**4.0402**	**4.2999**	**4.4765**	**0.6359**	**0.6813**	**0.7142**	**0.7249**	**1.5995**	**2.0094**	**2.2756**	**2.4718**	**0.7055**	**0.7341**	**0.7606**	**0.7751**

Table 3: Fair recommendation comparisons via four fairness-aware metrics. Lower A@K, G@K, L@K values indicate greater fairness, while a higher D@K value signifies a fairer recommendation.

Model	REDIAL				REDIAL			
	A@5↓	A@10↓	G@5↓	G@10↓	L@5↓	L@10↓	D@5↑	D@10↑
Ours*	**0.0028**	**0.0040**	**0.7163**	**0.7504**	**3.5946**	**4.0402**	**0.6359**	**0.6813**
w/o $\mathcal{G}_e^{(t)}$	0.0034	0.0050	0.7441	0.7808	3.9215	4.375	0.5540	0.6348
w/o $\mathcal{G}_i^{(t)}$	0.0034	0.0049	0.7457	0.7802	3.9138	4.3647	0.5573	0.6327
w/o $\mathcal{G}_w^{(t)}$	0.0034	0.0049	0.7455	0.7806	3.9133	4.3691	0.5602	0.6305
w/o $\mathcal{G}_r^{(t)}$	0.0036	0.0052	0.7490	0.7839	4.0019	4.4320	0.5671	0.6185
w/o $\mathcal{L}_e^{(t)}$	0.0034	0.0049	0.7447	0.7808	3.9216	4.3701	0.5673	0.6426
w/o $\mathcal{L}_i^{(t)}$	0.0034	0.0050	0.7460	0.7807	3.9138	4.3702	0.5608	0.6267
w/o $\mathcal{L}_w^{(t)}$	0.0034	0.0050	0.7447	0.7802	3.9124	4.3686	0.5617	0.6291
w/o $\mathcal{L}_r^{(t)}$	0.0035	0.0050	0.7469	0.7830	3.9673	4.3905	0.5721	0.6395
Model	TG-REDIAL				TG-REDIAL			
	A@5↓	A@10↓	G@5↓	G@10↓	L@5↓	L@10↓	D@5↑	D@10↑
Ours*	**0.0010**	**0.0015**	**0.4825**	**0.5313**	**1.5995**	**2.0094**	**0.7055**	**0.7341**
w/o $\mathcal{G}_e^{(t)}$	0.0013	0.0018	0.5253	0.5671	1.8841	2.2918	0.6726	0.7132
w/o $\mathcal{G}_i^{(t)}$	0.0013	0.0019	0.5375	0.5787	1.9514	2.3702	0.6570	0.6917
w/o $\mathcal{G}_w^{(t)}$	0.0013	0.0018	0.5266	0.5702	1.8790	2.3144	0.6725	0.7028
w/o $\mathcal{G}_r^{(t)}$	0.0013	0.0018	0.5190	0.5622	1.8549	2.2863	0.6756	0.7090
w/o $\mathcal{L}_e^{(t)}$	0.0012	0.0018	0.5225	0.5646	1.8598	2.2795	0.6689	0.7089
w/o $\mathcal{L}_i^{(t)}$	0.0013	0.0018	0.5285	0.5706	1.9003	2.3253	0.6609	0.6973
w/o $\mathcal{L}_w^{(t)}$	0.0013	0.0018	0.5213	0.5655	1.8613	2.2885	0.6678	0.6981
w/o $\mathcal{L}_r^{(t)}$	0.0013	0.0018	0.5164	0.5572	1.8452	2.2682	0.6643	0.7054

Table 4: Ablation studies on the recommendation task.

Figure 2: Hyperparameter Analysis.

Popularity (A@K), Gini Coefficient (G@K), KL-Divergence (L@K), and Difference (D@K). Lower values for A@K, G@K, and L@K indicate better fairness, while a higher D@K suggests fairer recommendations. Table 3 shows our method achieves lower A@K, G@K, and L@K values, with a higher D@K compared to other baselines, indicating improved fairness. This success can be attributed to several key factors: 1) HyFairCRS constructs eight distinct user interest representations, which offer deeper insights into users' preferences and effectively mitigate preference bias. This multifaceted approach allows for a more nuanced understanding of user behavior. 2) By decoupling the hypergraph into hypergraphs and line graphs, we are able to learn both node-level and hyperedge-level features. This dual perspective enhances preference diversification, allowing the model to capture a broader
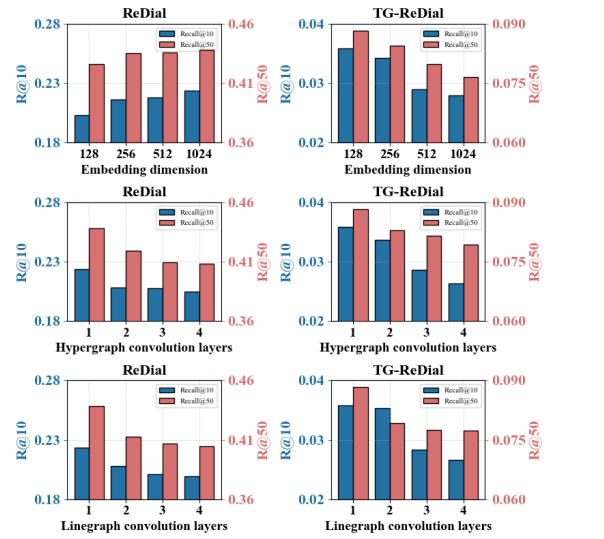
range of user interests and interactions. 3) We employ multi-set graph contrastive learning to refine user interest features. This technique helps to distinguish between similar user preferences, improving the model's ability to make accurate and relevant recommendations. These factors contribute to the superior performance of HyFairCRS, enabling it to adapt effectively across various domains and user contexts.

4.5 Ablation Studies (RQ4)

To verify the contributions of each component in our HyFairCRS, we conduct ablation experiments, including: 1) remove different hypergraphs: w/o $\mathcal{G}_e^{(t)}$, w/o $\mathcal{G}_i^{(t)}$, w/o $\mathcal{G}_w^{(t)}$, w/o $\mathcal{G}_r^{(t)}$; 2) remove different line graphs: w/o $\mathcal{L}_e^{(t)}$, w/o $\mathcal{L}_i^{(t)}$, w/o $\mathcal{L}_w^{(t)}$, w/o $\mathcal{L}_r^{(t)}$. The results of these ablation studies, summarized in Table 4, reveal a consistent trend:

removing any type of hypergraph leads to a significant decrease in performance metrics. This outcome confirms the critical role that each hypergraph plays in enhancing the model's overall effectiveness. For instance, the removal of the entity hypergraph may hinder the model's ability to accurately represent and relate different entities within the dataset, which is essential for generating contextually relevant recommendations. Similarly, excluding the user hypergraph could diminish the model's understanding of user preferences and behaviors, thereby impacting the personalization aspect of recommendations. Overall, these ablation experiments not only validate the individual contributions of each component in HyFairCRS but also highlight the synergistic effects that arise from their combined use, ultimately leading to a more robust and effective recommendation system.

4.6 Study on Cross-domain Datasets (RQ5)

To thoroughly evaluate our proposed HyFairCRS, we incorporate two additional datasets that span diverse domains, including movies, music, books, sports, news, and restaurants. Table 5 summarizes the experimental results. Notably, our method consistently outperforms the strongest baselines in both conversational and recommendation tasks across these cross-domain datasets. This consistent performance across varied domains underscores the robustness and scalability of HyFairCRS. The model's ability to adapt effectively to different contexts and user preferences is a significant advantage, as it allows for more personalized and relevant recommendations. For instance, whether users are seeking movie suggestions or looking for the latest news articles, HyFairCRS demonstrates a remarkable capacity to tailor its outputs to meet individual needs. Furthermore, the statistically significant improvements highlighted by an asterisk (*) reinforce the effectiveness of our approach in enhancing both recommendation accuracy and conversational quality. Overall, the results from these diverse datasets confirm that HyFairCRS is not only effective in its primary tasks but also possesses the versatility to handle a wide range of user interactions across different domains.

4.7 Hyperparameters Analysis (RQ6)

We analyze the impact of key hyperparameters on overall performance (see Fig. 2). We can see that: 1) As we increase the number of hypergraph and line graph convolution layers in both datasets, we

Model	OpenDialKG		DuRecDial	
	R@1	R@10	R@1	R@10
KBRD	0.1448	0.3162	0.0618	0.3971
KGSF	0.0626	0.1757	0.1395	0.4367
ReDial	0.0008	0.0134	0.0005	0.0336
TGReDial	0.2149	0.4035	0.0956	0.4882
HyCoRec	0.2742	0.4490	0.1279	0.4750
HiCore	0.2628	0.4526	0.1735	0.5471
HyFairCRS*	**0.2895**	**0.4630**	**0.1952**	**0.5688**

Model	OpenDialKG		DuRecDial	
	Dist-2	Dist-3	Dist-2	Dist-3
KBRD	0.3192	1.7660	0.5180	1.5500
KGSF	0.1687	0.5387	0.1389	0.3862
ReDial	0.1579	0.5808	0.1095	0.3981
TGReDial	0.4836	2.1430	0.5453	2.0030
HyCoRec	2.8190	4.7710	1.0820	2.4440
HiCore	2.8430	4.8120	1.0940	2.4280
HyFairCRS*	**2.9162**	**5.0194**	**1.1210**	**2.4688**

Table 5: Results on both recommendation and conversation tasks on OpenDialKG and DuRecDial datasets involving various domains. * indicates statistically significant improvement ($p < 0.05$) over all baselines.

observe a decline in overall performance. This suggests that a single convolution layer is sufficient for effectively capturing the underlying patterns in the data. Additional layers may lead to overfitting, where the model becomes too tailored to the training data and struggles to generalize to unseen examples. 2) In the ReDial dataset, higher feature dimensions correlate with improved performance, indicating that more complex representations can enhance the model's ability to capture intricate user behaviors and preferences. In contrast, in the TGReDial dataset, we see a decline in performance with increased feature dimensions. This suggests that smaller datasets may benefit from simpler representations, as more complex features can introduce noise and reduce the model's effectiveness. More complex datasets, like ReDial, require higher dimensions to adequately represent the diversity of user interests and interactions.

5 Conclusion

To enhance recommendation fairness in the CRS, we propose HyFairCRS, a novel paradigm that learns diverse interest representations through hypergraphs and line graphs during user interactions via natural language dialogues. These representations are used for generating responses in the conversation task and ensuring fair recommendations in the recommendation task. Extensive experiments show that HyFairCRS outperforms strong baselines in reducing unfair recommendations.

6 Limitations

Despite the promising results of HyFairCRS, two limitations warrant consideration. First, while HyFairCRS aims to promote fairness, it may not fully eliminate all biases inherent in the data. The model's effectiveness is contingent on the quality and representativeness of the training data; if the data reflects existing biases, these may persist in the recommendations. Second, user satisfaction and engagement are influenced by factors beyond fairness, such as relevance and personalization. Balancing these aspects while maintaining fairness remains a critical challenge for future research.

7 Ethics Statement

The data used in this paper are sourced from open-access repositories, and do not pose any privacy concerns. We are confident that our research adheres to the ethical standards set forth by ACL.

8 Acknowledgements

References

Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary G. Ives. 2007. Dbpedia: A nucleus for a web of open data. In *International Semantic Web Conference/Asian Semantic Web Conference*, volume 4825, pages 722–735.

Alex Beutel, Jilin Chen, Tulsee Doshi, Hai Qian, Li Wei, Yi Wu, Lukasz Heldt, Zhe Zhao, Lichan Hong, Ed H. Chi, and Cristos Goodrow. 2019. Fairness in recommendation ranking through pairwise comparisons. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019*, pages 2212–2220. ACM.

Jesús Bobadilla, Raúl Lara-Cabrera, Ángel González-Prieto, and Fernando Ortega. 2021. Deepfair: Deep learning for improving fairness in recommender systems. *Int. J. Interact. Multim. Artif. Intell.*, 6(6):86–94.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Conference on Neural Information Processing Systems*.

Robin Burke, Nasim Sonboli, and Aldo Ordonez-Gauger. 2018. Balanced neighborhoods for multi-sided fairness in recommendation. In *Conference on Fairness, Accountability and Transparency, FAT 2018*, volume 81 of *Proceedings of Machine Learning Research*, pages 202–214. PMLR.

Qibin Chen, Junyang Lin, Yichang Zhang, Ming Ding, Yukuo Cen, Hongxia Yang, and Jie Tang. 2019. Towards knowledge-based recommender dialog system. *arXiv preprint arXiv:1908.05391*.

Yang Deng, Yaliang Li, Fei Sun, Bolin Ding, and Wai Lam. 2021a. Unified conversational recommendation policy learning via graph-based reinforcement learning. In *Conference on Research and Development in Information Retrieval*, pages 1431–1441.

Yang Deng, Yaliang Li, Fei Sun, Bolin Ding, and Wai Lam. 2021b. Unified conversational recommendation policy learning via graph-based reinforcement learning. In *Conference on Research and Development in Information Retrieval*, pages 1431–1441.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4171–4186. Association for Computational Linguistics.

Yingqiang Ge, Shuchang Liu, Ruoyuan Gao, Yikun Xian, Yunqi Li, Xiangyu Zhao, Changhua Pei, Fei Sun, Junfeng Ge, Wenwu Ou, and Yongfeng Zhang. 2021. Towards long-term fairness in recommendation. In *WACM International Conference on Web Search and Data Mining, WSDM 2021*, pages 445–453. ACM.

Yingqiang Ge, Shuya Zhao, Honglu Zhou, Changhua Pei, Fei Sun, Wenwu Ou, and Yongfeng Zhang. 2020. Understanding echo chambers in e-commerce recommender systems. In *Conference on Research and Development in Information Retrieval*, pages 2261–2270. ACM.

Di Jin, Luzhi Wang, He Zhang, Yizhen Zheng, Weiping Ding, Feng Xia, and Shirui Pan. 2023a. A survey on fairness-aware recommender systems. *Information Fusion*, 100:101906.

Xin Jin, Wu Zhou, Jinyu Wang, Duo Xu, and Yongsen Zheng. 2023b. An order-complexity aesthetic assessment model for aesthetic-aware music recommendation. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 6938–6947.

Wang-Cheng Kang and Julian J. McAuley. 2018. Self-attentive sequential recommendation. In *IEEE International Conference on Data Mining*, pages 197–206.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Empirical Methods in Natural Language Processing (Demonstrations)*, pages 1746–1751.

Hang Lei, Yin Zhao, and Longjun Cai. 2020a. Multi-objective optimization for guaranteed delivery in video service platform. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2020*, pages 3017–3025. ACM.

Wenqiang Lei, Xiangnan He, Yisong Miao, Qingyun Wu, Richang Hong, Min-Yen Kan, and Tat-Seng Chua. 2020b. Estimation-action-reflection: Towards deep interaction between conversational and recommender systems. In *Web Search and Data Mining*, pages 304–312.

Wenqiang Lei, Gangyi Zhang, Xiangnan He, Yisong Miao, Xiang Wang, Liang Chen, and Tat-Seng Chua. 2020c. Interactive path reasoning on graph for conversational recommendation. In *International Conference on Knowledge Discovery and Data Mining*, pages 2073–2083.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *the Association for Computational Linguistics*, pages 7871–7880. Association for Computational Linguistics.

Raymond Li, Samira Ebrahimi Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Chris Pal. 2018a. Towards deep conversational recommendations. *Advances in Neural Information Processing Systems*, 31.

Raymond Li, Samira Ebrahimi Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Chris Pal. 2018b. Towards deep conversational recommendations. In *Advances in Neural Information Processing Systems*, pages 9748–9758.

Qin Liu, Xuan Feng, Tianlong Gu, and Xiaoli Liu. 2024. Faircrs: Towards user-oriented fairness in conversational recommendation systems. In *Proceedings of ACM Conference on Recommender Systems, RecSys 2024*, pages 126–136. ACM.

Ying Chieh Liu and Min Qi Huang. 2021. Examining the matthew effect on youtube recommendation system. In *Conference on Technologies and Applications of Artificial Intelligence*, pages 146–148.

Zeming Liu, Haifeng Wang, Zhengyu Niu, Hua Wu, and Wanxiang Che. 2021. Durecdial 2.0: A bilingual parallel corpus for conversational recommendation. In *Conference on Empirical Methods in Natural Language Processing EMNLP*, pages 4335–4347. Association for Computational Linguistics.

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

Seungwhan Moon, Pararth Shah, Anuj Kumar, and Rajen Subba. 2019. Opendialkg: Explainable conversational reasoning with attention-based walks over knowledge graphs. In *Conference of the Association for Computational Linguistics ACL*, pages 845–854. Association for Computational Linguistics.

Gourab K. Patro, Arpita Biswas, Niloy Ganguly, Krishna P. Gummadi, and Abhijnan Chakraborty. 2020. Fairrec: Two-sided fairness for personalized recommendations in two-sided platforms. In *WWW The Web Conference 2020*, pages 1194–1204. ACM / IW3C2.

Mingjie Qian, Yongsen Zheng, Jinghui Qin, and Liang Lin. 2023. Hutcrs: Hierarchical user-interest tracking for conversational recommender system. In *Conference on Empirical Methods in Natural Language Processing*, pages 10281–10290.

Xuhui Ren, Hongzhi Yin, Tong Chen, Hao Wang, Zi Huang, and Kai Zheng. 2021. Learning to ask appropriate questions in conversational recommendation. In *Conference on Research and Development in Information Retrieval*, pages 808–817.

Rajdeep Sarkar, Koustava Goswami, Mihael Arcan, and John Philip McCrae. 2020. Suggest me a movie for tonight: Leveraging knowledge graphs for conversational recommendation. In *Conference on Computational Linguistics*, pages 4179–4189.

Chenzhan Shang, Yupeng Hou, Wayne Xin Zhao, Yaliang Li, and Jing Zhang. 2023. Multi-grained hypergraph interest modeling for conversational recommendation. *AI Open*, 4:154–164.

Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *Association for the Advancement of Artificial Intelligence*, pages 4444–4451.

Harald Steck. 2018. Calibrated recommendations. In *Conference on Recommender Systems*, pages 154–162.

Fei Sun, Jun Liu, Jian Wu, Changhua Pei, Xiao Lin, Wenwu Ou, and Peng Jiang. 2019. Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer. In *International Conference on Information and Knowledge Management*, pages 1441–1450. 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*, pages 5998–6008.

Yifan Wang, Peijie Sun, Weizhi Ma, Min Zhang, Yuan Zhang, Peng Jiang, and Shaoping Ma. 2024. Intersectional two-sided fairness in recommendation. In *WWW on Web Conference 2024*, pages 3609–3620. ACM.

Haolun Wu, Yansen Zhang, Chen Ma, Fuyuan Lyu, Fernando Diaz, and Xue Liu. 2022. A survey of diversification techniques in search and recommendation. *CoRR*, abs/2212.14464.

Le Wu, Lei Chen, Pengyang Shao, Richang Hong, Xiting Wang, and Meng Wang. 2021a. Learning fair representations for recommendation: A graph-based perspective. In *WWW '21: The Web Conference 2021*, pages 2198–2208. ACM / IW3C2.

Yao Wu, Jian Cao, Guandong Xu, and Yudong Tan. 2021b. TFROM: A two-sided fairness-aware recommendation model for both customers and providers. In *ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1013–1022. ACM.

Lanling Xu, Zihan Lin, Jinpeng Wang, Sheng Chen, Wayne Xin Zhao, and Ji-Rong Wen. 2024. Promoting two-sided fairness with adaptive weights for providers and customers in recommendation. In *ACM RecSys Conference on Recommender Systems*, pages 918–923. ACM.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime G. Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in Neural Information Processing Systems*, pages 5754–5764.

Meike Zehlike and Carlos Castillo. 2020. Reducing disparate exposure in ranking: A learning to rank approach. In *The Web Conference 2020, WWW 2020*, pages 2849–2855. ACM / IW3C2.

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. 2020. DIALOGPT : Large-scale generative pre-training for conversational response generation. In *the Association for Computational Linguistics*, pages 270–278. Association for Computational Linguistics.

Yuying Zhao, Yu Wang, Yunchao Liu, Xueqi Cheng, Charu C. Aggarwal, and Tyler Derr. 2023a. Fairness and diversity in recommender systems: A survey. *CoRR*, abs/2307.04644.

Zhipeng Zhao, Kun Zhou, Xiaolei Wang, Wayne Xin Zhao, Fan Pan, Zhao Cao, and Ji-Rong Wen. 2023b. Alleviating the long-tail problem in conversational recommender systems. In *ACM Conference on Recommender Systems*, pages 374–385. ACM.

Yongsen Zheng, Ziliang Chen, Jinghui Qin, and Liang Lin. 2024a. Facetcrs: multi-faceted preference learning for pricking filter bubbles in conversational recommender system. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 9405–9413.

Yongsen Zheng, Jinghui Qin, Pengxu Wei, Ziliang Chen, and Liang Lin. 2023a. Cipl: counterfactual interactive policy learning to eliminate popularity bias for online recommendation. *IEEE Transactions on Neural Networks and Learning Systems*.

Yongsen Zheng, Guohua Wang, Yang Liu, and Liang Lin. 2024b. Diversity matters: User-centric multi-interest learning for conversational movie recommendation. MM '24. Association for Computing Machinery.

Yongsen Zheng, Pengxu Wei, Ziliang Chen, Yang Cao, and Liang Lin. 2021. Graph-convolved factorization machines for personalized recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 35(2):1567–1580.

Yongsen Zheng, Pengxu Wei, Ziliang Chen, Chengpei Tang, and Liang Lin. 2023b. Routing user-interest markov tree for scalable personalized knowledge-aware recommendation. *IEEE Transactions on Neural Networks and Learning Systems*.

Yongsen Zheng, Ruilin Xu, Ziliang Chen, Guohua Wang, Mingjie Qian, Jinghui Qin, and Liang Lin. 2024c. Hycorec: Hypergraph-enhanced multi-preference learning for alleviating matthew effect in conversational recommendation. In *Proceedings of Annual Meeting of the Association for Computational Linguistics ACL 2024*, pages 2526–2537. Association for Computational Linguistics.

Yongsen Zheng, Ruilin Xu, Guohua Wang, Liang Lin, and Kwok-Yan Lam. 2024d. Mitigating matthew effect: Multi-hypergraph boosted multi-interest self-supervised learning for conversational recommendation. In *Proceedings of Conference on Empirical Methods in Natural Language Processing, EMNLP 2024*, pages 1455–1466. 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 *International Conference on Knowledge Discovery and Data Mining*, pages 1006–1014.

Kun Zhou, Yuanhang Zhou, Wayne Xin Zhao, Xiaoke Wang, and Ji-Rong Wen. 2020b. Towards topic-guided conversational recommender system. In *International Conference on Computational Linguistics*, pages 4128–4139.

Yuanhang Zhou, Kun Zhou, Wayne Xin Zhao, Cheng Wang, Peng Jiang, and He Hu. 2022. C^2-crs: Coarse-to-fine contrastive learning for conversational recommender system. In *Web Search and Data Mining*, pages 1488–1496. ACM.