HyCoRec: Hypergraph-Enhanced Multi-Preference Learning for Alleviating Matthew Effect in Conversational Recommendation

Yongsen Zheng[1], Ruilin Xu[1], Ziliang Chen[2,5], Guohua Wang[3*],
Mingjie Qian[1], Jinghui Qin[4], Liang Lin[1,2*]

[1]Sun Yat-sen University, [2]Peng Cheng Laboratory, [3]South China Agricultural University
[4]Guangdong University of Technology, [5]Jinan University
{z.yongsensmile, wangguohuagmial, scape1989}@gmail.com
{xurlin5, qianmj7}@mail2.sysu.edu.cn, c.ziliang@yahoo.com, linliang@ieee.org

Abstract

The Matthew effect is a notorious issue in Recommender Systems (RSs), *i.e.*, the rich get richer and the poor get poorer, wherein popular items are overexposed while less popular ones are regularly ignored. Most methods examine Matthew effect in static or nearly-static recommendation scenarios. However, the Matthew effect will be increasingly amplified when the user interacts with the system over time. To address these issues, we propose a novel paradigm, Hypergraph-Enhanced Multi-Preference Learning for Alleviating Matthew Effect in Conversational Recommendation (HyCoRec), which aims to alleviate the Matthew effect in conversational recommendation. Concretely, HyCoRec devotes to alleviate the Matthew effect by learning multi-aspect preferences, *i.e.*, item-, entity-, word-, review-, and knowledge-aspect preferences, to effectively generate responses in the conversational task and accurately predict items in the recommendation task when the user chats with the system over time. Extensive experiments conducted on two benchmarks validate that HyCoRec achieves new state-of-the-art performance and the superior of alleviating Matthew effect. Our code is available at https://github.com/zysensmile/HyCoRec.

1 Introduction

Conversational Recommender Systems (CRSs) engage in iterative conversations with users to provide personalized recommendations (Qin et al., 2023; Li et al., 2023; Mishra et al., 2023), which have been widely adopted in various domains such as music recommendation (Epure and Hennequin, 2023) and online e-commerce (Liu et al., 2023). Nevertheless, CRSs often face the prominent issue of Matthew effect (Liu and Huang, 2021), which can be described as "the rich get richer and the poor get

poorer". This phenomenon indicates that popular items/categories from past data receive more visibility in subsequent recommendations while less popular ones tend to be overlooked or ignored.

Recently, many research efforts have focused on examining the Matthew effect in static or relatively-static offline recommendation scenarios (Liu and Huang, 2021; Anderson et al., 2020; Hansen et al., 2021). These offline studies strive to explore the potential causes behind the manifestation of the Matthew effect, and two key causes have been identified. One cause (Anderson et al., 2020; Hansen et al., 2021; Liang et al., 2021; Zheng et al., 2021a) is that individuals with narrower and less diverse preferences exhibit a higher vulnerability to being trapped within the confines of the Matthew effect. Another cause (Zheng et al., 2021b) is that the severe popularity bias where popular items consistently receive amplified exposure while less popular ones are underexposed. Although these methods have undoubtedly contributed valuable insights into the phenomenon of the Matthew effect, they directly overlook the adverse impact stemming from the dynamic user-system feedback loop. More recently, Gao et al. (Gao et al., 2023) explore the Matthew effect in dynamic user-system interactions, but it lacks real-time user engagement through natural language conversations.

Despite their effectiveness, most methods still suffer from two major limitations. *1) Interactive Schema*. Many methods aim to mitigate Matthew effect in the static recommendation settings without considering the user-system feedback loop (Zhang et al., 2021). In reality, the Matthew effect will progressively amplify as users dynamically interact with the system over time. Worse still, such amplification will inevitably lead to a series of notorious issues such as filter bubbles (Steck, 2018) and echo chamber (Ge et al., 2020). Thus, it is crucial to consider the dynamic user-system interactions to alleviate Matthew effect. *2) Preference*

---

*Corresponding author.

2526

Learning. Prior studies (Anderson et al., 2020; Hansen et al., 2021; Liang et al., 2021; Zheng et al., 2021a) show that the key to mitigating Matthew effect is to learn diverse user preferences. Thus, many methods leverage multiplex external Knowledge Graphs (KGs) to model multi-aspect preferences. But traditional KG edges are limited to linking only two vertices (*i.e.*, factors), restricting preference learning to pairwise interactions. Instead, user relations exhibit intricate complexity, such as a user's preference for a garment involves multiple factors like color, brand, style, and texture simultaneously. Hence, extending the number of vertices for learning diverse preferences is rather important.

To address these issues, we propose a novel paradigm, **Hy**pergraph-Enhanced Multi-Preference Learning for Alleviating Matthew Effect in **Co**nversational **Rec**ommendation (**HyCoRec**), which consists of Hypergraph-Enhanced Multi-Preference Learning and Hypergraph-aware CRS. The former aims to model multi-aspect preferences, specifically targeting item-aspect, entity-aspect, word-aspect, review-aspect, and knowledge-aspect preferences. It addresses the Matthew effect in CRS by utilizing item-based hypergraph, entity-based hypergraph, word-based hypergraph, item reviews, and knowledge graphs to learn and derive these preferences. The latter focuses on leveraging these multi-aspect preferences as users interact with the system. Concretely, multi-aspect preferences are adopting to accurately predict the next utterances in the conversational task, and effectively make diverse item predictions in the recommendation task. By incorporating and utilizing these multi-aspect preferences, the system aims to provide precise and diverse recommendations that cater to the individual user's preferences and needs for alleviating the Matthew effect as they continue to engage with the system. Empirically, extensive experimental results on two benchmarks show that HyCoRec outperforms all the compared baselines, and the superior of mitigating Matthew effect.

Overall, our main contributions are included:

- To the best of our knowledge, this is the first work to model multi-aspect user preferences, *i.e.*, item-, entity-, word-, review-, knowledge-aspect preference, to alleviate Matthew effect in the CRS.
- We proposed a novel end-to-end framework, Hy-CoRec, which adopts the multi-aspect preferences to effectively generate responses in the conversational task and accurately predict items in the recommendation task.

- Quantitative and qualitative experimental results on two CRS-based datasets exhibit superior performance of HyCoRec and the effectiveness of mitigating Matthew effect in the CRS.

2 Related Work

2.1 Conversational Recommender System

Conversational Recommender System aims to capture user preferences through dialogues and provide high-quality recommendations. Previous research on CRS can be broadly categorized into two main types: attribute-based CRS (Deng et al., 2021a; Lei et al., 2020a,b; Ren et al., 2021; Xu et al., 2021) and generation-based CRS (Chen et al., 2019; Deng et al., 2023; Li et al., 2022; Zhou et al., 2020a, 2022; Shang et al., 2023). Attribute-based CRS involves capturing user preferences by asking questions about item attributes and generating responses using pre-defined templates (Lei et al., 2020a). But this strategy often neglects the importance of generating responses that resemble natural human language, which can negatively impact the user experience. On the other hand, generation-based CRS tackles this issue by utilizing the Seq2Seq architecture (Vaswani et al., 2017a) to integrate both conversation and recommendation tasks to produce fluent and coherent human-like responses. Despite their effectiveness, they fail to model users' diverse preferences since the user-item interactions data is rather sparse and limited. In contrast, our work aims to model multi-aspect preferences for exploring user diverse intricate relation patterns.

2.2 Matthew Effect in Recommendation

Matthew effect is a notorious issue in RSs. Recently, Liu et al. (Liu and Huang, 2021) have substantiated the occurrence of the Matthew effect in YouTube's recommendation system. Besides, Wang et al. (Wang et al., 2019) undertook a rigorous quantitative analysis, offering valuable insights into the quantitative characteristics of the Matthew effect in recommender systems based on collaborative filtering. To alleviate Matthew effect, one common method is to consider recommendation diversity strongly advocated by researchers (Anderson et al., 2020; Hansen et al., 2021; Liang et al., 2021; Zheng et al., 2021a), and another critical perspective is by removing popularity bias, a factor that has been identified as a catalyst for its amplification (Zheng et al., 2021b). But these methods

predominantly focus on investigating the Matthew effect in the static recommendation settings without considering the user-system feedback loop. Instead, our HyCoRec aims to alleviate Matthew effect considering dynamic user-system feedback loop.

3 HyCoRec

Matthew effect is a notorious issue in the CRS, and it inevitably becomes intensified over time due to the existence of the dynamic user-system feedback loop. To tackle these challenges, we propose a novel paradigm, HyCoRec, which consists of Hypergraph-Enhanced Multi-Preference Learning and Hypergraph-Aware CRS. The overall pipeline of our HyCoRec is depicted in Fig.1.

3.1 Preliminaries

3.1.1 Conversational Recommendation

Conversational recommendation is a personalized approach where the system engages in continuous dialogues with users to gain a deeper understanding of their preferences and deliver customized suggestions. This interactive method enables the system to collect additional insights into the user's preferences, context, and requirements, resulting in more precise and relevant recommendations. CRSs are widely applied across diverse domains like e-commerce, music streaming, movie recommendations, and others, aiming to enrich user experience and satisfaction.

3.1.2 Hypergraph

Hypergraphs demonstrate intricate configurations, capturing intricate relationships among numerous elements through hyperlinks. In our research, we depict user inclinations by constructing multi-grained hypergraphs, encompassing the item-based hypergraph $\mathcal{G}_{\text{item}}^{(t)}$, entity-based hypergraph $\mathcal{G}_{\text{entity}}^{(t)}$, and the word-based hypergraph $\mathcal{G}_{\text{word}}^{(t)}$. Each hypergraph can be delineated as $\mathcal{G}_{\text{item}}^{(t)} = (\mathcal{I}_*^{(t)}, \mathcal{H}_*^{(t)}, \mathbf{N}_*^{(t)})$, comprising: (1) a node collection $\mathcal{I}_*^{(t)}$; (2) a hyperege ensemble $\mathcal{H}_*^{(t)}$; (3) a $|\mathcal{I}_*^{(t)}| \times |\mathcal{H}_*^{(t)}|$ adjacent matrix $\mathbf{N}_*^{(t)}$ signifying the weighted link between each node and hyperedge.

3.2 Hypergraph-Enhanced Multi-Preference Learning

Extensive experiments by most existing methods (Hussein et al., 2020; Liu et al., 2021; Nguyen et al., 2014) have consistently shown that users with restricted preferences are highly influenced by the Matthew effect. Thus, the key to alleviating such bad effect is to model the diverse user preferences. Along this line, we formulate the Hypergraph-Enhanced Multi-Preference Learning, including Multi-Hypergraph Construction and Multi-Preference Learning.

3.2.1 Multi-Hypergraph Construction

Traditional KGs focus on pairwise interactions for preference learning, as edges connect only two vertices. However, user preferences often exhibit complex item relation patterns. To address this, we construct multiple hypergraphs (item-aspect, entity-aspect, and word-aspect), enabling connections between more than two vertices.

Item-based Hypergraph. Items directly reflect users' genuine preferences. Users might prefer related items, such as products from the same brand or with similar features. Thus, establishing connections among similar or functionally similar items is crucial for exploring a diverse preferences. To do this, we first extract items from a session and treat them as vertices, forming a hyperedge. Then, All hyperedges associated with a user are connected through shared items to create the item-based hypergraph $\mathcal{G}_{\text{item}}^{(t)}$ as:

$$\mathcal{G}_{\text{item}}^{(t)} = (\mathcal{I}_i^{(t)}, \mathcal{H}_i^{(t)}, \mathbf{N}_i^{(t)}). \tag{1}$$

where $\mathcal{I}_i^{(t)}$ means the item set extracted from the historical conversations, $\mathcal{H}_i^{(t)}$ is the hyperedge set, and $\mathbf{N}_i^{(t)} \in \{0,1\}^{|\mathcal{I}_i^{(t)}| \times |\mathcal{H}_i^{(t)}|}$ is the incidence matrix, which can be defined as:

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

The degree of a vertex $v \in \mathcal{I}_i^{(t)}$ is denoted as $d(v) = \sum_{h \in \mathcal{H}_i^{(t)}} \mathbf{N}_{v,h}^{(t)}$. Similarly, the degree of an edge $h \in \mathcal{H}_i^{(t)}$ is written as $\delta(h) = \sum_{v \in \mathcal{I}_i^{(t)}} \mathbf{N}_{v,h}^{(t)}$. Besides, we use $\mathbf{V}_i^{(t)} \in \mathbb{N}^{|\mathcal{I}_i^{(t)}| \times |\mathcal{I}_i^{(t)}|}$ and $\mathbf{E}_i^{(t)} \in \mathbb{N}^{|\mathcal{H}_i^{(t)}| \times |\mathcal{H}_i^{(t)}|}$ to be the diagonal matrices of the vertex degrees and edge degrees, respectively.

Entity-based Hypergraph. To address the sparsity and limitations of historical user-item interaction data, we utilize the extensive DBpedia KG (Auer et al., 2007) to construct the entity-based hypergraph. Specifically, we extract individual items mentioned in conversations as entities and their
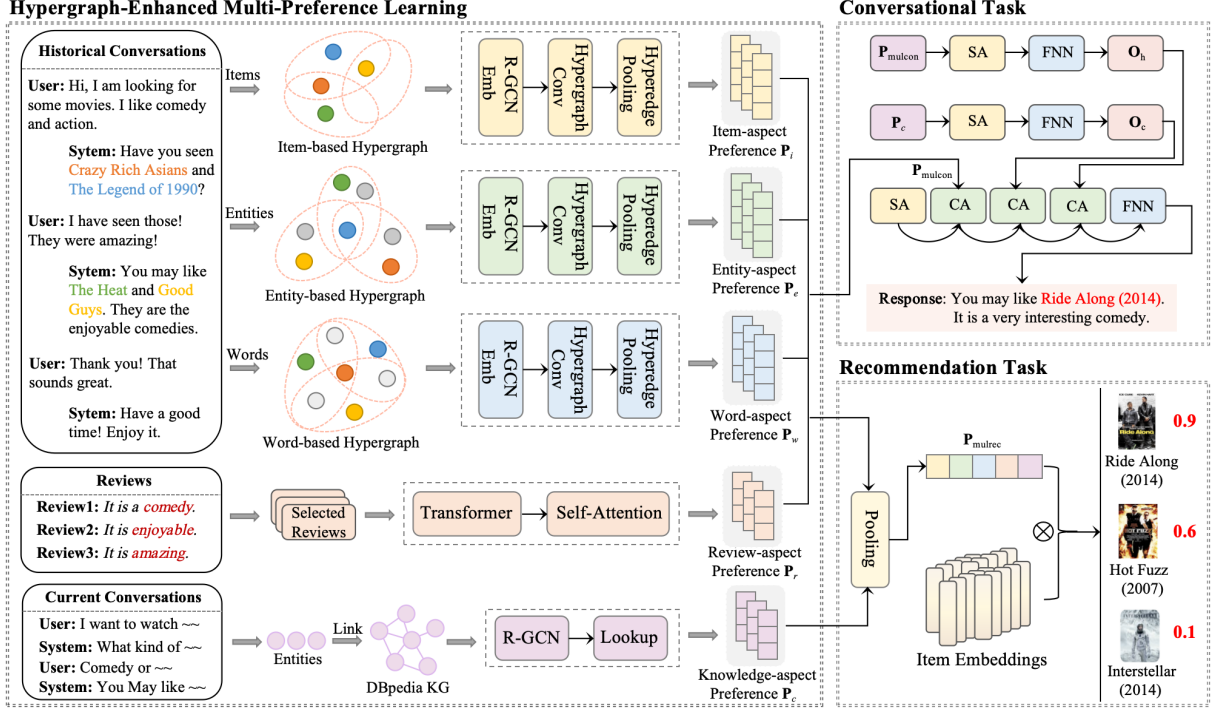
Figure 1: Overview of our HyCoRec framework, which consists of Hypergraph-Enhanced Multi-Preference Learning and Hypergraph-Aware CRS. The former aims to dynamically learn multi-aspect user preferences, while the latter contains the conversation task to generate diverse responses and the recommendation task to predict target items.

k-hop neighbors to form each hyperedge. This approach allows us to capture shared semantic connotations among the extended neighbors. The hyperedges are then connected based on the entities they have in common. Formally, the entity-based hypergraph $\mathcal{G}_{\text{entity}}^{(t)}$ can be represented as:

$$\mathcal{G}_{\text{entity}}^{(t)} = (\mathcal{I}_i^{(t)} \cup \mathcal{I}_e^{(t)}, \mathcal{H}_e^{(t)}, \mathbf{N}_e^{(t)}). \qquad (3)$$

where $\mathcal{I}_e^{(t)}$ represents the k-hop neighbors, $\mathcal{H}_e^{(t)}$ is the hyperedge set, and $\mathbf{N}_e^{(t)} \in \{0,1\}^{|\mathcal{I}_e^{(t)}| \times |\mathcal{H}_e^{(t)}|}$ is the incidence matrix defined by Eq.(2). Similarly, $\mathbf{V}_e^{(t)} \in \mathbb{N}^{|\mathcal{I}_e^{(t)}| \times |\mathcal{I}_e^{(t)}|}$ and $\mathbf{E}_e^{(t)} \in \mathbb{N}^{|\mathcal{H}_e^{(t)}| \times |\mathcal{H}_e^{(t)}|}$ denote the diagonal matrices of vertex degrees and edge degrees, respectively.

Word-based Hypergraph. Keywords in conversations are vital for understanding users' needs. By analyzing prominent words, we can identify specific preferences, which is crucial for modeling diverse user preferences. To achieve this, we build a word-based hypergraph using the word-oriented KG ConcetNet (Speer et al., 2017) to uncover semantic relations like synonymy, antonyms, and co-occurrence. We represent each historical conversation item as a keyword and extend it to include k-hop neighbors, forming a hyperedge. All the hyperedges connect through shared words. The

word-based hypergraph $\mathcal{G}_{\text{word}}^{(t)}$ can be defined as follows:

$$\mathcal{G}_{\text{word}}^{(t)} = (\mathcal{I}_i^{(t)} \cup \mathcal{I}_w^{(t)}, \mathcal{H}_w^{(t)}, \mathbf{N}_w^{(t)}). \qquad (4)$$

where $\mathcal{I}_w^{(t)}$ is k-hop neighbors, $\mathcal{H}_w^{(t)}$ means the hyperedge set, and $\mathbf{N}_w^{(t)} \in \{0,1\}^{|\mathcal{I}_w^{(t)}| \times |\mathcal{H}_w^{(t)}|}$ is the incidence matrix defined as Eq.(2). Similarly, let $\mathbf{V}_w^{(t)} \in \mathbb{N}^{|\mathcal{I}_w^{(t)}| \times |\mathcal{I}_w^{(t)}|}$ and $\mathbf{E}_w^{(t)} \in \mathbb{N}^{|\mathcal{H}_w^{(t)}| \times |\mathcal{H}_w^{(t)}|}$ denote the diagonal matrices of the vertex degrees and the edge degrees, respectively.

3.2.2 Multi-Preference Learning

Upon constructing multiple hypergraphs as described earlier, we will leverage these hypergraphs to effectively capture diverse user preferences for mitigating the Matthew effect. This includes preferences related to items, entities, words, reviews, and knowledge aspects, all of which play a role in modeling multi-aspect preferences.

Item-aspect Preference. Modeling item-aspect preferences holds significant importance in comprehending users' distinct tastes and preferences concerning the various items they interact with. In line with this objective, we derive the item-aspect preference $\mathbf{P}_i$ by leveraging the item-based hypergraph. To effectively capture high-order relations,

inspired by (Bai et al., 2021), we define our *Hypergraph Convolution* function $\text{HConv}(\cdot)$ as follows:

$$\mathbf{X}_m^{(l+1)} = \text{HConv}\left(\mathbf{X}^{(l)}, \mathbf{N}_i^{(t)}, \mathbf{V}_i^{(t)}, \mathbf{E}_i^{(t)}, \mathbf{I}_i^{(t)}\right),$$

$$\text{HConv}(\cdot) = (\mathbf{V}_i^{(t)})^{-1}\mathbf{N}_i^{(t)}(\mathbf{E}_i^{(t)})^{-1}(\mathbf{N}_i^{(t)})^T\mathbf{X}^{(l)}\mathbf{W}_i^{(l)},$$

$$\mathbf{X}^{(l+1)} = \text{Pooling}\left(\mathbf{X}_m^{(l+1)}\right)_{m=1}^{M}.$$

(5)

Here, $\mathbf{X}^l$ and $\mathbf{X}^{(l+1)}$ represent the input of the l-th and $(l+1)$-th layers, respectively, and $\mathbf{W}_i^{(l)}$ denotes the trainable parameter. The notations $\mathbf{N}_i^{(t)}$, $\mathbf{V}_i^{(t)}$, and $\mathbf{E}_i^{(t)}$ have been discussed in Section 3.2.1. Specifically, $\mathbf{I}_i^{(t)}$ signifies the item representations of $\mathcal{I}_i^{(t)}$ extracted from the encoded entity embeddings (Shang et al., 2023). Additionally, m denotes the number of heads in the multi-head architecture (Vaswani et al., 2017b). Finally, we apply an average pooling $\text{Pooling}(\cdot)$ on the representation $\mathbf{X}^{(L+1)}$ obtained from the last layer (*i.e.*, $(L+1)$ layer) to learn item-aspect preference $\mathbf{P}_i$:

$$\mathbf{P}_i = \mathbf{X}^{(L+1)} = \text{Pooling}\left(\mathbf{X}_m^{(L+1)}\right)_{m=1}^{M}, \quad (6)$$

Entity-aspect Preference. Capturing entity-aspect preferences is highly advantageous for unveiling complex relationship patterns underlying users' behaviors. To achieve this, we employ the entity-based hypergraph as a mechanism to learn entity-aspect preferences. In a similar vein to the item-aspect preference, the entity-aspect preference $\mathbf{P}_e$ can be represented as:

$$\mathbf{X}_j^{(l+1)} = \text{HConv}\left(\mathbf{X}^{(l)}, \mathbf{N}_e^{(t)}, \mathbf{V}_e^{(t)}, \mathbf{E}_e^{(t)}, \mathbf{I}_{i+e}^{(t)}\right),$$

$$\mathbf{P}_e = \mathbf{X}^{(L+1)} = \text{Pooling}\left(\mathbf{X}_j^{(L+1)}\right)_{j=1}^{J}.$$

(7)

The specifics of $\mathbf{N}_e^{(t)}$, $\mathbf{V}_e^{(t)}$, and $\mathbf{E}_e^{(t)}$ can be found in Section 3.2.1. Besides, $\mathbf{I}_{i+e}^{(t)}$ is the entity representations of the entity set $\mathcal{I}_i^{(t)} \cup \mathcal{I}_e^{(t)}$. Moreover, j denotes the number of heads in the multi-head architecture, and $\mathbf{W}_e^{(l)}$ is the trainable parameter.

Word-aspect Preference. Keywords occurring in conversations directly reflect users' specific or potential preferences. Derived from the word-based hypergraph, the word-aspect preference $\mathbf{P}_w$ can be formulated as:

$$\mathbf{X}_f^{(l+1)} = \text{HConv}\left(\mathbf{X}^{(l)}, \mathbf{N}_w^{(t)}, \mathbf{V}_w^{(t)}, \mathbf{E}_w^{(t)}, \mathbf{I}_{i+w}^{(t)}\right),$$

$$\mathbf{P}_w = \mathbf{X}^{(L+1)} = \text{Pooling}\left(\mathbf{X}_f^{(L+1)}\right)_{f=1}^{F}.$$

(8)

Here $\mathbf{N}_w^{(t)}$, $\mathbf{V}_w^{(t)}$, and $\mathbf{E}_w^{(t)}$ are explained in more detail in Section 3.2.1. Additionally, $\mathbf{I}_{i+w}^{(t)}$ represents the word representations obtained from the encoded entity embeddings. The variable f denotes the number of heads in the multi-head architecture, and $\mathbf{W}_f^{(l)}$ denotes the trainable parameter.

Review-aspect Preference. Item reviews provide valuable insights into users' experiences and reflections. Analyzing these reviews helps identify patterns, sentiment trends, and user attitudes, leading to a better understanding of user preferences. Taking inspiration from the merits of Transformer model, we utilize the Transformer framework to encode accessed reviews (Lu et al., 2021). Specifically, given a review R, the output embeddings from the previous transformer layer, denoted as $\mathcal{T}^l(R)$, define the subsequent layer $\mathcal{T}^{l+1}(R)$ using the *Multi-head Attention* function $\text{MHA}(\cdot)$ as:

$$\mathcal{T}^{l+1}(R) = \text{MHA}(\mathcal{T}^l(R), \mathcal{T}^l(R), \mathcal{T}^l(R)),$$

$$\text{MHA}(\boldsymbol{K}, \boldsymbol{Q}, \boldsymbol{V}) = [\text{head}_1^l; \cdots; \text{head}_g^l]\mathbf{W}^l,$$

$$\text{head}_g^l = \text{SA}(\mathcal{T}^l(R)\mathbf{W}_k, \mathcal{T}^l(R)\mathbf{W}_q, \mathcal{T}^l(R)\mathbf{W}_v),$$

$$\text{SA}(\boldsymbol{K}, \boldsymbol{Q}, \boldsymbol{V}) = \text{Softmax}(\frac{\boldsymbol{Q}\boldsymbol{K}^{\text{T}}}{\sqrt{d/g}})\boldsymbol{V},$$

(9)

where g is the number of heads, $\mathbf{W}^l$ denotes the trainable parameters, and each head head_g^l is computed using the *Scaled Dot-Product Attention* (Vaswani et al., 2017c) $\text{SA}(\cdot)$. $\boldsymbol{K}$, $\boldsymbol{Q}$ and $\boldsymbol{V}$ indicate the key, query and value matrices, respectively. $\mathbf{W}_k$, $\mathbf{W}_q$, and $\mathbf{W}_v$ are learnable parameters. For convenience, we consider the output embeddings of the final transformer layer as the review-aspect preferences $\mathbf{P}_r$:

$$\mathbf{P}_r = \text{MHA}(\mathcal{T}^{\mathcal{L}}(\mathcal{R}), \mathcal{T}^{\mathcal{L}}(\mathcal{R}), \mathcal{T}^{L}(\mathcal{R})). \quad (10)$$

Here $\mathcal{L}$ is the number of transformer layers.

Knowledge-aspect Preference. The information conveyed in the ongoing conversation reflects the dynamic preferences of the users, providing valuable insights into their current interests. Thus, our focus lies in modeling the preference for knowledge aspects by encoding the entities mentioned in the current conversation. Given the current conversation context $\mathcal{C}$, we leverage DBpedia and CN-DBpedia, to extract entities $\mathcal{E}_k = \{e_1, e_2, \cdots, e_k\}$

along the paths. To capture high-order entity representations, we use RGCN to explicitly capture relational semantics by adopt contrastive pre-training (Shang et al., 2023). The representation of entity e at the $(l+1)$-th layer can be computed as:

$$e^{l+1} = \sigma(\sum_{r \in \mathcal{R}} \sum_{\hat{e} \in \mathcal{N}_e^r} \frac{1}{Z_l} \mathbf{W}_1^l \hat{e}^l + \mathbf{W}_2^l e^l), \quad (11)$$

where e^l is the l-th layer's representation of entity e, σ means the sigmoid function, $\hat{e}$ refers to entities from the one-hop neighbor set $\mathcal{N}_e^r$ under relation r, and Z_l is the hyperparameter. $\mathbf{W}_1^l$ and $\mathbf{W}_2^l$ can be trained. We use the representation e^L from the last layer as knowledge-aspect preference $\mathbf{P}_c$:

$$\mathbf{P}_c = \text{RGCN}(\mathcal{E}_k) = \{e_1^T, e_2^T, \cdots, e_k^T\}, \quad (12)$$

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

3.3 Hypergraph-Aware CRS

To combat the Matthew effect in the CRS, we adopt multi-aspect preferences, *i.e.*, $\mathbf{P}_i$, $\mathbf{P}_e$, $\mathbf{P}_w$, $\mathbf{P}_r$, and $\mathbf{P}_c$, to accurately predict item in the recommendation task and effectively generate responses in the conversational task.

3.3.1 Recommendation Task

The recommendation task aims to accurately predict items for users through natural conversations in dynamic user-system interactions. To address the Matthew effect, we first integrate multiple preferences to induce the fused preference $\mathbf{P}_{\text{mulrec}}$ in the recommendation task as:

$$\mathbf{P}_h = [\mathbf{P}_i; \mathbf{P}_e; \mathbf{P}_w; \mathbf{P}_r],$$
$$\mathbf{P}_{\text{mulrec}} = \text{Pooling}([\text{Pooling}(\mathbf{P}_h); \mathbf{P}_c]). \quad (13)$$

where ; denotes the concatenation operation. Next, the vector $\mathbf{P}_{\text{mulrec}}$ is used to select the suitable items in all the candidate set from item set $\mathcal{I}$, and the recommendation prediction is calculated as:

$$\mathcal{P}_{\text{rec}} = \text{Softmax}(\mathbf{P}_{\text{mulrec}} \cdot \text{E}_I^T), \quad (14)$$

where E_I is embeddings of all candidate items from item set $\mathcal{I}$. We use cross-entropy loss (Shang et al., 2023) to learn the recommendation task:

$$\mathcal{L}_{\text{r}} = -\sum_{j=1}^{B} \sum_{i=1}^{|\mathcal{I}|} [-(1-y_{ij}) \cdot \log(1 - \mathcal{P}_{\text{rec}}^{(j)}(i))$$
$$+ y_{ij} \cdot \log(\mathcal{P}_{\text{rec}}^{(j)}(i))], \quad (15)$$

here the symbol B represents the size of the mini-batch, and $y_{ij} \in \{0, 1\}$ denotes the target label.

3.3.2 Conversational Task

The conversation task focuses on generating proper dialogue utterances to respond to user inputs. To generate diverse responses, we integrate multi-aspect preferences vectors to derive the fused preference in the conversation task $\mathbf{P}_{\text{mulcon}}$ as:

$$\mathbf{P}_{\text{mulcon}} = \text{MHA}([\mathbf{P}_c; \mathbf{P}_h; \mathbf{P}_h]), \quad (16)$$

here $\mathbf{P}_h$ is defined as Eq.(13). Then, this fused preference $\mathbf{P}_{\text{mulcon}}$ is fed into the Transformer-based encoder-decoder framework for generating diverse responses. Let $\mathbf{Y}^{n-1}$ be the output of the last time unit, then the current one $\mathbf{Y}^n$ is:

$$\begin{aligned}
\mathbf{A}_0^n &= \text{MHA}(\mathbf{Y}^{n-1}, \mathbf{Y}^{n-1}, \mathbf{Y}^{n-1}), \\
\mathbf{A}_1^n &= \text{MHA}(\mathbf{A}_0^n, \mathbf{P}_{\text{mulcon}}, \mathbf{P}_{\text{mulcon}}), \\
\mathbf{A}_2^n &= \text{MHA}(\mathbf{A}_1^n, \mathbf{P}_c, \mathbf{P}_c), \\
\mathbf{A}_3^n &= \text{MHA}(\mathbf{A}_1^n, \mathbf{P}_h, \mathbf{P}_h), \\
\mathbf{A}_4^n &= \beta \cdot \mathbf{A}_2^n + (1-\beta) \cdot \mathbf{A}_3^n, \\
\mathbf{Y}^n &= \text{FFN}(\mathbf{A}_4^n).
\end{aligned} \quad (17)$$

Here $\text{FFN}(\cdot)$ is the fully-connected feed-forward network, and β is hyper-parameter to balance two signals. To enhance the response diversity, we use preference-aware bias and item-related bias following (Shang et al., 2023). Given the predicted sequence $\{s_{t-1}\}$, the probability of the next token is calculated as:

$$\begin{aligned}
\mathcal{P}_{\text{conv}}(s_t | \{s_{t-1}\}) = P_1(s_t | Y_i) + P_2(s_t | \mathbf{P}_{\text{mulrec}}) \\
+ P_3(s_t | \mathbf{P}_{\text{mulrec}}),
\end{aligned}$$

$$(18)$$

where s_t is the t-th utterances, and $\{s_{t-1}\} = s_1, s_2, \cdots, s_{t-1}$. Inspired by (Shang et al., 2023), $P_1(\cdot)$, $P_2(\cdot)$, and $P_3(\cdot)$ are the vocabulary probability, vocabulary bias, and copy probability, respectively. Next, we use the cross-entropy loss:

$$\mathcal{L}_{\text{c}} = -\sum_{b=1}^{B} \sum_{t=1}^{T} \log(\mathcal{P}_{\text{conv}}(s_t | \{s_{t-1}\})). \quad (19)$$

Here T denotes the truncated length of utterances.

4 Experiments and Analyses

We conduct experiments to fully evaluate our Hy-CoRec and answer the following questions:

- **RQ1:** How does HyCoRec perform compared with all baselines in the recommendation task?
- **RQ2:** How does HyCoRec perform compared with all baselines in the conversation task?

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
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
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
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
HyCoRec*	**0.2231**	**0.4351**	**0.0797**	**0.0898**	**0.1123**	**0.1579**	**0.0377**	**0.0826**	**0.0154**	**0.0173**	**0.0162**	**0.0245**

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

- **RQ3:** How does HyCoRec alleviate Matthew effect in the CRS?
- **RQ4:** How do the item-based hypergraph $\mathcal{G}_{\text{item}}^{(t)}$, entity-based hypergraph $\mathcal{G}_{\text{enti}}^{(t)}$, word-based hypergraph $\mathcal{G}_{\text{word}}^{(t)}$, and item reviews R contribute to the performance?
- **RQ5:** How do parameters affect our HyCoRec?
- **RQ6:** It is better to provide the case studies to comprehensively understand about how HyCoRec handles Matthew effect in the CRS?

4.1 Experimental Protocol

Datasets. We evaluate our HyCoRec on two challenging CRS-based datasets REDIAL (Li et al., 2018b) and TG-REDIAL (Zhou et al., 2020b). The REDIAL consists of 11,348 dialogues involving 956 users and 6,924 items, while the TG-REDIAL contains 10,000 dialogues with 1,482 users and 33,834 items. The reviews in REDIAL are sourced from the IMDb, while the reviews in TG-REDIAL are collected from Douban.

Baselines. To fully evaluate our HyCoRec, we conduct a comprehensive evaluation by comparing our method with several state-of-the-art methods. The compared methods include **TextCNN** (Kim, 2014), **SASRec** (Kang and McAuley, 2018), **BERT4Rec** (Sun et al., 2019), **Transformer** (Vaswani et al., 2017b), **ReDial** (Li et al., 2018a), **KBRD** (Chen et al., 2019), **KGSF** (Zhou et al., 2020a), **KGConvRec** (Sarkar 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., 2023), **UniCRS** (Deng et al., 2021b), and **MHIM** (Shang et al., 2023).

4.2 Recommendation Performance (RQ1)

Following (Shang et al., 2023), we adopt Recall@K (R@K), MRR@K (M@K), NDCG@K (N@K) (K=10, 50) to evaluate the recommendation task. Experimental results in Table 1 validate that our HyCoRec outperforms all the compared methods.

The improvement of HyCoRec over these baselines can be attributed to three reasons: (1) Incorporating external knowledge sources like DBpedia and ConceptNet into the CRS proves beneficial in exploring users' intricate behaviors, considering the sparse and limited nature of user-item interaction data. (2) Dialogues serve as a treasure trove of valuable information beyond the explicit user inputs. By considering the ongoing conversation between the user and the system, HyCoRec can capture the user's current context and understand their immediate needs. (3) Modeling multi-aspect preferences, including item-, entity-, word-, review-, and knowledge-aspect preferences, to enhance recommendation diversity and alleviate the Matthew effect as users interact with the system over time.

4.3 Conversational Performance (RQ2)

In the conversational task, we adopt Distinct n-gram (Dist-n) (Shang et al., 2023) (n=2,3,4) to evaluate the diversity of generated responses. Table 2 summarizes the experimental results, it is observed that our HyCoRec is superior to all the compared baselines. We can observe that the performance rankings of the four baseline models remain consistent, with KBRD leading the way, followed by KGSF, Transformer, and finally ReDial. This can be attributed to the fact that KBRD leverages external knowledge sources to align the representations of items and words. Besides, KGSF enriches its decoder by incorporating cross-attention mecha-

Model	REDIAL			TG-REDIAL		
	Dist-2	Dist-3	Dist-4	Dist-2	Dist-3	Dist-4
ReDial	0.0214	0.0659	0.1333	0.2178	0.5136	0.7960
Trans.	0.0538	0.1574	0.2696	0.2362	0.7063	1.1800
KBRD	0.0765	0.3344	0.6100	0.8013	1.7840	2.5977
KGSF	0.0572	0.2483	0.4349	0.3891	0.8868	1.3337
C2-CRS	0.2623	0.3891	0.6202	0.5235	1.9961	2.9236
UniCRS	0.2464	0.4273	0.5290	0.6252	2.2352	2.5194
LOT-CRS	0.3312	0.6155	0.9248	0.9287	2.4880	3.4972
DialoGPT	0.3542	0.6209	0.9482	1.1881	2.4269	3.9824
GPT-3	0.3604	0.6399	0.9511	1.2255	2.5713	4.0713
MHIM	0.3278	0.6204	**0.9629**	1.1100	2.3520	3.8200
HyCoRec*	**0.3661**	**0.6434**	0.9523	**1.2590**	**2.6000**	**4.1210**

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

Datasets	REDIAL				
Models	C@5	C@10	C@15	C@20	Iso-Index
KBRD	0.0579	0.0810	0.0961	0.1072	0.1149
KGSF	0.0664	0.0831	0.1195	0.1366	0.1055
KGConvRec	0.0478	0.0735	0.1044	0.1235	0.1003
MHIM	0.1098	0.1492	0.1747	0.1977	0.0923
HyCoRec	**0.1168**	**0.1579**	**0.1848**	**0.2071**	**0.0617**
Datasets	TG-REDIAL				
Models	C@5	C@10	C@15	C@20	Iso-Index
KBRD	0.0757	0.1204	0.1468	0.1584	0.1222
KGSF	0.0847	0.1324	0.1606	0.1858	0.1198
KGConvRec	0.0720	0.0904	0.1228	0.1515	0.1091
MHIM	0.1749	0.2493	0.2939	0.3423	0.1042
HyCoRec	**0.1841**	**0.2743**	**0.3100**	**0.3608**	**0.0791**

Table 3: Results on *C@k* and *Iso-Index* metrics.

Model	REDIAL		TG-REDIAL	
	R@10	R@50	R@10	R@50
HyCoRec	**0.2231**	**0.4351**	**0.0377**	**0.0826**
w/o item hypergraph	0.2061	0.4177	0.0347	0.0733
w/o entity hypergraph	0.2049	0.4206	0.0267	0.0696
w/o word hypergraph	0.2058	0.4194	0.0257	0.0661
w/o item reviews	0.2102	0.4203	0.0267	0.0771

Table 4: Ablation studies on the recommendation task.

nisms along with embeddings from both entity- and word-level knowledge graphs (KGs). Nevertheless, Transformer and ReDial solely rely on token sequences, disregarding the user preferences that are concealed within the entities.

Compared with these baselines, the improvement of HyCoRec can be attributed to the fact that: (1) HyCoRec considers multi-aspect user preferences to generate diverse responses that effectively align with the user's multi-level preferences for alleviating Matthew effect. (2) To accurately forecast the next utterance, we integrate the fused preference obtained from various aspects into a Transformer-based encoder-decoder framework to generate high-quality responses to meet users' dynamic needs and interests.

4.4 Study on Matthew Effect (RQ3)

As our objective is to alleviate the Matthew effect in the CRS, we extensively examine the recommendation outcomes and compare them with the strongest baselines to assess whether HyCoRec can effectively mitigate Matthew effect. To mitigate the Matthew effect, the crucial factor is to enhance the diversity of the recommendation results. Thus, we adopt two commonly-used metrics *Coverage*@k (C@k) and Isolation-Index (Iso-Index) to assess the extent of recommendation diversification by taking into account the distinctions among recommended items. The higher coverage value demonstrates its superior capability to encompass a larger portion of the recommendation space, encompassing items from various categories. A lower isolation-index value indicates a higher diversity in the recommended results.

As shown in Table 3, it is evident that our HyCoRec consistently achieves the highest values of *Coverage* and the lowest isolation-index value

across all datasets compared with the strongest baselines. For instance, on the REDIAL, our HyCoRec achieves substantial improvements of 102.72%, 75.90%, 144.35%, and 63.75% in terms of *Cover@5* when compared to all the strong models, namely KBRD, KGSF, KGConvRec, and MHIM, respectively. The results show that HyCoRec effectively addresses isolation and ensures extensive coverage of recommended items by providing users with a broader choice range, validating the superiority in alleviating the Matthew effect as the user interacts with the system.

4.5 Ablation Studies (RQ4)

In this part, we conduct ablation experiments with different variants of HyCoRec to verify the contributions of each component, including: 1) w/o item hypergraph: we remove the item-based hypergraph; 2) w/o entity hypergraph: we remove the entity-based hypergraph; 3) w/o word hypergraph: we remove the word-based hypergraph; 4) w/o item reviews: we remove item reviews. As shown in Table 4, we can observe that a substantial decline in performance when removing any type of component. The main reason is that various knowledge data can effectively explore users' multi-aspect preferences. This observation highlights the effectiveness of HyCoRec in alleviating the Matthew effect by providing diverse recommendation results.
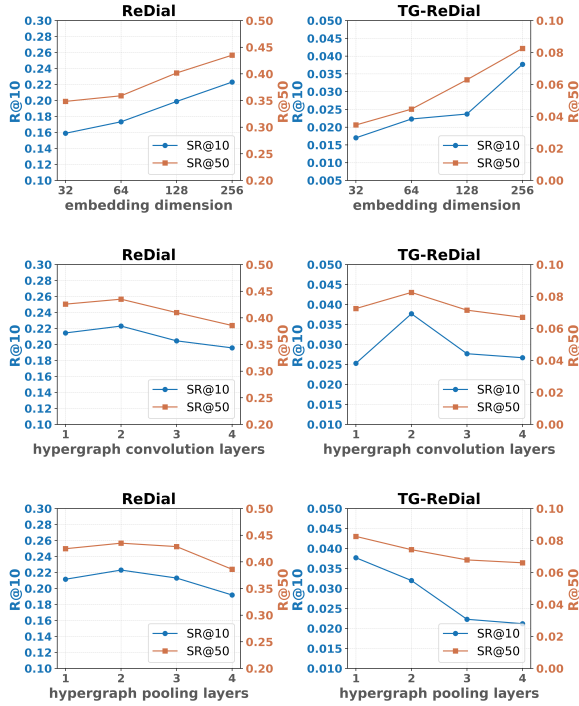
Figure 2: Impact of different hyperparameters.

(a) Our method effectively recommends a set of movies from different categories.

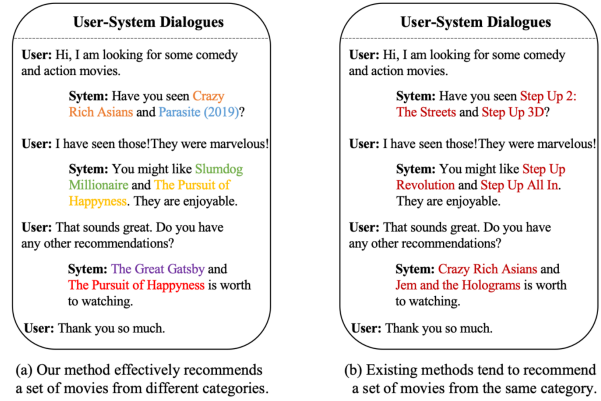(b) Existing methods tend to recommend a set of movies from the same category.

Figure 3: Case studies to comprehensively understand about how our proposed method HyCoRec handles Matthew effect in the CRS compared with most existing methods. Different colors denote different categories (see (a)) while the same color means the same category (see (b)).

4.6 Hyperparameters Analysis (RQ5)

Next, we investigate the impact of several important hyperparameters on the recommendation performance. As depicted in Fig.2, we can observe:

Firstly, with the increase of embedding dimension, the recommendation performance continually improves. This is because the large dimension could encode sufficient high-level feature representations. Secondly, the model performance is optimal when the layer number is set to 2 on both datasets. The main reason is that larger hypergraph convolution layer numbers easily lead to model overfitting while the smaller one fail to capture enough feature representations. Lastly, a suitable hypergraph pooling layer number can enhance the model performance but the larger one might damage recommendation performance. The reason is that too large hypergraph pooling layer numbers might lose the important feature representations.

4.7 Case Studies (RQ6)

For a more in-depth understanding of how our proposed method, HyCoRec, tackles the Matthew effect during user-system interactions, we present comparative case studies between our approach and existing methods, visually illustrating the dialogue recommendation outcomes in human-computer interaction. As illustrated in Fig.3, our method ef-

fectively recommends a diverse range of movies from different categories (see (a)), setting itself apart from existing methods that typically recommend movies from the same category (see (b)). The results highlight that our method achieves higher recommendation diversity, while most existing methods demonstrate lower diversity in recommendations. Generally, an effective strategy to alleviate the Matthew effect involves enhancing recommendation diversification (Anderson et al., 2020; Hansen et al., 2021; Liang et al., 2021; Zheng et al., 2021a). Thus, these results validate the effectiveness of our proposed method in mitigating the Matthew effect as users interact with the system over time in CRS.

5 Conclusion

The Matthew effect is a notorious issue in the CRS, and it will be increasingly amplified due to the dynamic user-system feedback loop. To address these issues, we propose a novel paradigm, HyCoRec, which aims to learn multi-aspect user preferences, *i.e.*, item-, entity-, word-, review-, and knowledge-aspect preferences, to effectively generate diverse responses in the conversation task and accurately predict items in the recommendation task for alleviating Matthew effect. Extensive experiments validate that our HyCoRec outperforms all the compared baselines and the superior of HyCoRec in alleviating Matthew effect in the CRS.

6 Limitations

While our HyCoRec has attained a remarkable state-of-the-art performance, it does have certain limitations. Firstly, the complexity and extensive nature of item reviews make the construction of the review-based hypergraph challenging and difficult. Consequently, the current version does not include the review-based hypergraph to capture a wider range of multiplex user relation patterns. Secondly, our proposed method necessitates the design of individual hypergraphs for learning multi-aspect preferences. This limitation could be addressed by developing a general framework that integrate any types of hypergraphs, thereby automatically unifying various knowledge sources.

7 Ethics Statement

The data utilized in our study 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

Ashton Anderson, Lucas Maystre, Ian Anderson, Rishabh Mehrotra, and Mounia Lalmas. 2020. Algorithmic effects on the diversity of consumption on spotify. In *The Web Conference*, pages 2155–2165.

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.

Song Bai, Feihu Zhang, and Philip H. S. Torr. 2021. Hypergraph convolution and hypergraph attention. *Pattern Recognit.*, 110:107637.

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*.

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.

Yang Deng, Wenxuan Zhang, Weiwen Xu, Wenqiang Lei, Tat-Seng Chua, and Wai Lam. 2023. A unified multi-task learning framework for multi-goal conversational recommender systems. *ACM Transactions on Information Systems*, 41(3):1–25.

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.

Elena V. Epure and Romain Hennequin. 2023. A human subject study of named entity recognition in conversational music recommendation queries. In *European Chapter of the Association for Computational Linguistics*, pages 1273–1288.

Chongming Gao, Kexin Huang, Jiawei Chen, Yuan Zhang, Biao Li, Peng Jiang, Shiqi Wang, Zhong Zhang, and Xiangnan He. 2023. Alleviating matthew effect of offline reinforcement learning in interactive recommendation. In *Conference on Research and Development in Information Retrieval*, pages 238–248. 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.

Christian Hansen, Rishabh Mehrotra, Casper Hansen, Brian Brost, Lucas Maystre, and Mounia Lalmas. 2021. Shifting consumption towards diverse content on music streaming platforms. In *Conference on Web Search and Data Mining*, pages 238–246. ACM.

Eslam Hussein, Prerna Juneja, and Tanushree Mitra. 2020. Measuring misinformation in video search platforms: An audit study on youtube. *ACM on Human-Computer Interaction*, 4(CSCW):048:1–048:27.

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.

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

Wenqiang Lei, Gangyi Zhang, Xiangnan He, Yisong Miao, Xiang Wang, Liang Chen, and Tat-Seng Chua. 2020b. Interactive path reasoning on graph for conversational recommendation. In *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.

Shimin Li, Qinyuan Cheng, Linyang Li, and Xipeng Qiu. 2023. Mitigating negative style transfer in hybrid

dialogue system. In *Association for the Advancement of Artificial Intelligence*, pages 13103–13111.

Shuokai Li, Ruobing Xie, Yongchun Zhu, Xiang Ao, Fuzhen Zhuang, and Qing He. 2022. User-centric conversational recommendation with multi-aspect user modeling. In *Conference on Research and Development in Information Retrieval*, pages 223–233.

Yile Liang, Tieyun Qian, Qing Li, and Hongzhi Yin. 2021. Enhancing domain-level and user-level adaptivity in diversified recommendation. In *Conference on Research and Development in Information Retrieval*, pages 747–756. ACM.

Ping Liu, Karthik Shivaram, Aron Culotta, Matthew A. Shapiro, and Mustafa Bilgic. 2021. The interaction between political typology and filter bubbles in news recommendation algorithms. In *The Web Conference*, pages 3791–3801.

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.

Yuanxing Liu, Weinan Zhang, Baohua Dong, Yan Fan, Hang Wang, Fan Feng, Yifan Chen, Ziyu Zhuang, Hengbin Cui, Yongbin Li, and Wanxiang Che. 2023. U-NEED: A fine-grained dataset for user needs-centric e-commerce conversational recommendation. In *Conference on Research and Development in Information Retrieval*, pages 2723–2732. ACM.

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.

Kshitij Mishra, Priyanshu Priya, and Asif Ekbal. 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*, pages 14408–14416.

Tien T. Nguyen, Pik-Mai Hui, F. Maxwell Harper, Loren G. Terveen, and Joseph A. Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *The Web Conference*, pages 677–686. ACM.

Libo Qin, Zhouyang Li, Qiying Yu, Lehan Wang, and Wanxiang Che. 2023. Towards complex scenarios: Building end-to-end task-oriented dialogue system across multiple knowledge bases. In *Association for the Advancement of Artificial Intelligence*, pages 13483–13491.

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, Ł ukasz Kaiser, and Illia Polosukhin. 2017a. Attention is all you need. *Advances in neural information processing systems*, 30.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017b. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017c. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008.

Hao Wang, Zonghu Wang, and Weishi Zhang. 2019. Quantitative analysis of matthew effect and sparsity problem of recommender systems. *CoRR*.

Kerui Xu, Jingxuan Yang, Jun Xu, Sheng Gao, Jun Guo, and Ji-Rong Wen. 2021. Adapting user preference to online feedback in conversational recommendation. In *Web Search and Data Mining*, pages 364–372.

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.

Yang Zhang, Fuli Feng, Xiangnan He, Tianxin Wei, Chonggang Song, Guohui Ling, and Yongdong Zhang. 2021. Causal intervention for leveraging popularity bias in recommendation. In *Conference on Research and Development in Information Retrieval*, pages 11–20. ACM.

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.

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

Yu Zheng, Chen Gao, Liang Chen, Depeng Jin, and Yong Li. 2021a. DGCN: diversified recommendation with graph convolutional networks. In *The Web Conference*, pages 401–412.

Yu Zheng, Chen Gao, Xiang Li, Xiangnan He, Yong Li, and Depeng Jin. 2021b. Disentangling user interest and conformity for recommendation with causal embedding. In *The Web Conference*, pages 2980–2991.

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.