Diversity Matters: User-Centric Multi-Interest Learning for Conversational Movie Recommendation

Yongsen Zheng
Nanyang Technological University
Jurong West, Singapore
z.yongsensmile@gmail.com

Guohua Wang*
South China Agricultural University
Guangzhou, China
wangguohuagmial@gmail.com

Yang Liu
Sun Yat-sen University
Guangzhou, China
liuy856@mail.sysu.edu.cn

Liang Lin
Sun Yat-sen University
Guangzhou, China
linliang@ieee.org

Abstract

Diversity plays a crucial role in Recommender Systems (RSs) as it ensures a wide range of recommended items, providing users with access to new and varied options. Without diversity, users often encounter repetitive content, limiting their exposure to novel choices. While significant efforts begin to enhance recommendation diversification in static offline scenarios, relatively less attention has been given to online Conversational Recommender Systems (CRSs). However, the lack of recommendation diversity in CRSs will increasingly exacerbate over time due to the dynamic user-system feedback loop, resulting in challenges such as the Matthew effect, filter bubbles, and echo chambers. To address these issues, we propose a novel paradigm, User-Centric Multi-Interest Learning for Conversational Movie Recommendation (CoMoRec), aiming to learn multiple user interests to improve result diversity for movie recommendations. Firstly, CoMoRec automatically models various facets of user interests, including context-, graph-, and review-based interests, to explore a wide range of user potential intentions. Then, it leverages these multi-aspect user interests to accurately predict personalized and diverse movie recommendations and generate fluent and informative responses during conversations. Extensive experiments on two publicly CRS-based movie datasets show that our CoMoRec achieves a new state-of-the-art performance and the superiority of improving recommendation diversity in the CRS.

CCS Concepts

• **Information systems** → **Recommender systems**.

Keywords

Conversational Recommendation, Diversified Movie Recommendation, Multiple User Interests , Natural Language Dialogues

ACM Reference Format:
Yongsen Zheng, Guohua Wang, Yang Liu, and Liang Lin. 2024. Diversity Matters: User-Centric Multi-Interest Learning for Conversational Movie Recommendation. In *Proceedings of the 32nd ACM International Conference on Multimedia (MM '24), October 28-November 1, 2024, Melbourne, VIC, Australia.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3664647.3680909

1 Introduction

With the rapid development of intelligent agents, Conversational Recommender Systems (CRSs) [17, 20, 24, 26, 40, 51, 53] has emerged as a prominent research topic, aiming to provide effective recommendations by engaging in natural language conversations between the user and the system. CRSs have been broadly adopted in various domains like music recommendation [11], electric commerce [23], and health counseling [36], *etc.* Despite these advancements, CRSs still face the challenge of inadequate recommendation diversity, and this problem becomes more pronounced as user interactions prolong. The lack of diverse recommendations can give rise to significant issues like exposure bias, filter bubbles, and echo chambers. Thus, enhancing diversification is crucial for the success of CRSs.

Recently, many research efforts have been devoted to facilitate the diversified recommendation. These endeavors can be classified into three primary directions: 1) Post-Processing (PP) methods [1, 3–5, 29, 54] involve the addition of a re-ranking or post-processing module to recommended items to strike a balance between relevance and diversity. 2) Determinantal Point Process (DPP) methods [6, 12, 14, 15, 42, 43] aim to select a diverse subset of items from a larger pool of retrieved items, utilizing heuristics different from those used in PP-based methods. 3) Learning To Rank (LTR) methods [8, 21] enhance recommendation diversity by optimizing the ranking strategy to generate an ordered list of items instead of a candidate set. While these methods have made significant strides in improving recommendation diversity, both PP-based methods and DDP-based ones rely heavily on the quality of user and item representations, whereas LTR-based methods face challenges in acquiring appropriate datasets. To this end, numerous graph-based algorithms [22, 37, 38, 44] have emerged to expand the spectrum of diverse items. By constructing the user-item bipartite graphs, these algorithms facilitate greater access to a wide array of diverse items.

Despite their effectiveness, most existing methods still encounter two major issues: *1) Diversity Exploration.* While many current methods [1, 6, 8, 22] primarily focus on exploring recommendation diversity in offline settings with relatively static conditions, there

[*] Corresponding author.

MM'24, October 28-November 1, 2024, Melbourne, VIC, Australia

ACM ISBN 979-8-4007-0686-8/24/10
https://doi.org/10.1145/3664647.3680909

exists a noticeable research gap when it comes to enhancing result diversity in interactive CRS contexts. In practical scenarios, insufficient recommendation diversity can lead to users being exposed to repetitive content, limiting their access to novel and varied options. Moreover, the issue of recommendation diversity becomes more prominent as users interact with the system over time [46], giving rise to notorious problems such as exposure bias [41], filter bubbles [31], and echo chambers [13]. Therefore, improving recommendation diversification in CRSs is of utmost importance. *2) Graph Structure.* Many graph-based algorithms [22, 37, 38, 44] strive to enhance the coverage of diverse items by constructing user-item bipartite graphs. However, these graphs often encounter sparsity issues, as a significant proportion of users engage with or express preferences for only a limited subset of items. This sparse graph structure presents challenges in capturing meaningful relationships or patterns between users and items. Moreover, bipartite graphs are not ideally suited for capturing higher-order relationships or interactions that involve more than two sets of entities. Representing complex relationships that extend beyond pairwise interactions becomes challenging within the confines of bipartite graphs. Consequently, this limitation can impact the quality of recommendations and the overall user experience.

To address these issues, we propose a novel end-to-end paradigm, User-Centric Multi-Interest Learning for **Co**nversational **Mo**vie **Rec**ommendation (**CoMoRec**), which is comprised of User-Centric Multi-Interest Learning and Interest-Enhanced CRS, aiming to modeling multi-aspect user interests for improving recommendation diversity as users interact with the system over time in the CRS. Considering the sparsity issue of the traditional user-item bipartite graphs, User-Centric Multi-Interest Learning paradigm first devises a high-order and densely connected Temporal Knowledge Graph (TKG) by leveraging the large-scale DBpedia Knowledge Graph (KG) as a valuable repository of structured entity data. Specifically, we extract entities from the conversation context and item reviews as the seed set at each time step. Based on these seed entities, we traverse the DBpedia KG to collect one-hop triples that where these triples consist of head-relation-tail associations that provide meaningful connections between the entities. After constructing the TKG, User-Centric Multi-Interest Learning paradigm further models multi-aspect user interests including context-based, graph-based, and review-based interest to capture the wide array of user intentions and preferences by adopting historical conversations, TKG, and item reviews. Moreover, the Interest-Enhanced CRS module focuses on leveraging these multiple user interests to make conversational movie recommendation. Concretely, it excels in making precise predictions for personalized and diverse movie recommendations that align with users' intentions and interests in the recommendation task, and generating fluent and informative responses in the conversation task. By leveraging the information obtained from context-based, graph-based, and review-based interests, the system can produce more tailored and engaging dialogue, fostering a positive user experience. Extensive experiments conducted on two publicly CRS-based movie datasets have provided compelling evidence of the superiority of our proposed CoMoRec for conversational movie recommendation, and the effectiveness of improving recommendation diversity in the CRS.

Overall, our main contributions are included as follows:

- To the best of our knowledge, this is the first work to model multi-aspect user interests, including context-, graph-, and review-based interests, to improve recommendation diversity in the CRS.
- We propose a novel end-to-end paradigm, CoMoRec, including User-Centric Multi-Interest Learning and Interest-Enhanced CRS. The former aims to model users' multi-interest while the latter devotes to generate responses and predict movies effectively.
- Extensive experiments on two CRS-based movie datasets show the superiority of our CoMoRec and its effectiveness improving recommendation diversification for movie recommendation as users dynamically interact with the system over time in the CRS.

2 Related Work

2.1 Conversational Recommender System

With the rapid development of intelligent agents in various domains, Conversational Recommender Systems (CRSs) have attracted a lot of attention from researchers [17, 20, 24, 26, 40, 51, 53], which aim to provide accurate recommendation through natural language conversations between users and systems [24, 53]. These CRS-based methods can be divided into two groups: attributed-based CRS and human-like CRS. The former [9, 33, 47, 50] aims to ask users which attributes they like or dislike for efficiently explore user preference by leveraging pre-defined actions (*e.g.*, item attributes and intent slots). These methods usually utilize the multi-armed bandit models [9] or reinforcement learning [33] to optimize the interaction strategy. Due to the heavy dependencies on the pre-defined actions and templates, they cannot be flexibly applied in various domains. The latter [17, 20, 24, 51, 53] is more realistic because they tend to provide recommendation via human-like responses. Human-like CRS usually designs a conversation module to provide proper response and a recommendation module to make recommendations. However, these approaches suffer from the limited and inadequate contextual information in the initial conversational utterances. To address these problems, most existing methods either introduce structured external data (*e.g.*, knowledge graph) [49, 53], or unstructured external data (*e.g*, item reviews) [24], to complement the conversation utterance. These approaches still fall short in terms of providing diversified recommendations. Instead, we follow the latter category and model multi-aspect user interests by leveraging various knowledge to improve recommendation diversity as user interacts with the system over time in the CRS.

2.2 Diversified Recommendation

The concept of recommendation diversification was initially introduced by Ziegler et al. [54], who employed a greedy algorithm [5] inspired by the field of information retrieval. Since then, a series of Post-Processing (PP)-based methods [1, 3–5, 29, 54] have been proposed to achieve a balance between recommendation relevance and diversity. For instance, Sha et al. [29] propose an advanced framework that incorporates the notions of relevance, user preferences, and variety. Similarly, Qin et al. [25] address this issue by employing a linear combination of the rating function and an entropy regularizer. Later, Determinantal Point Processes (DPP)-based methods [6, 12, 14, 15, 42, 43] focus on selecting a diverse subset of items from a larger pool of retrieved items, replacing the heuristics employed in PP-based methods. For instance, Gartrell et al. [12] introduce a novel approach to learning the DPP kernel from observed data by employing a low-rank factorization of the kernel. Recently, a new line of methods based on Learning to Rank (LTR) [8, 21] has emerged, aiming to enhance recommendation diversity through the adoption of ranking strategies. For instance, Cheng et al. [8] put forward a machine learning-based diversification approach by integrating the recommendation model with a structured Support Vector Machine (SVM) [34]. Despite their effectiveness, these existing methods focus on the recommendation diversity in the offline recommendation settings, instead, our proposed work aims to enhance recommendation diversification as users chat with the system over time in the CRS.

3 CoMoRec

In the CRS, as users continually with the online system over time, if the lack of recommendation diversification persists, it can lead to a series of notorious issues such as filter bubbles and echo chambers. To address these issues, we propose a novel paradigm CoMoRec, which is comprised of User-Centric Multi-Interest Learning and Interest-Enhanced CRS. The former focuses on modeling multi-aspect user interests, while the latter aims to adopt these multiple interests to accurately predict items in the recommendation task and effectively generate responses in the conversation task. The pipeline of our CoMoRec is depicted in Fig.1.

3.1 User-Centric Multi-Interest Learning

3.1.1 ***Temporal Knowledge Graph.*** Most conventional methods strive to explore the diverse range of user interests by constructing user-item bipartite graphs, which establish connections between users and items based on their historical interaction logs. However, these bipartite graphs often face the challenge of sparsity since the majority of users tend to interact with or express preferences for only a small subset of items. Therefore, capturing meaningful relationships or patterns between users and items becomes difficult in the face of such sparsity. To address these issues, we propose the Temporal Knowledge Graph (TKG) to build higher-order structural connectivity by leveraging large-scale knowledge graph.

Context-based Entities Extraction. To construct the TKG, we first extract entities from the conversations by retrieving the entity names over the large-scale DBpedia [2] KG $\mathcal{G}$ due to its fruitful facts and relations. It consists of a large number of triples $(\boldsymbol{e}_1, \boldsymbol{r}, \boldsymbol{e}_2)$, where $\boldsymbol{e}_1$ and $\boldsymbol{e}_2 \in \mathcal{E}$ refer to the head and tail entities, and $\boldsymbol{r}$ denotes the relation between them. Let $C = \{s_t\}_{t=1}^{n}$ denote the conversation context, comprising all utterances s_t that form the dialogue history provided by the user and the system in alternating turns. Firstly, we establish a mapping between each item in the item set $\mathcal{V}$ and the corresponding entity in the entity set $\mathcal{E}$ using their names, inspired by [49]. For example, the movie item "The Heat" mentioned in the C would be linked to "http://dbpedia.org/resource/The_Heat_(film)" in the DBpedia KG $\mathcal{G}$. Besides, we utilize a similar approach to associate informative non-item entities that appear in C with entities within $\mathcal{E}$. This step assists in identifying relevant entities that are connected to the items and conversation responses. Moreover, we perform entity linking on the conversation history, which involves identifying and extracting entities mentioned in the conversation. Formally, context-based entity set $\mathcal{E}_c^{(t)}$ at t can be described as:

$$\mathcal{E}_c^{(t)} = \mathcal{F}_{\text{extract}}(C, \mathcal{V}, \mathcal{E}, \mathcal{G}). \tag{1}$$

Review-based Entities Extraction. Next, our attention turns to the crucial task of extracting entities from the relevant reviews. It is important to note that our primary objective is to identify and retrieve reviews that provide valuable insights, as not all reviews contain meaningful information. In fact, irrelevant reviews can hinder the exploration of diverse user interests. Additionally, reviews expressing inconsistent attitudes can introduce noise into the discussion, making it challenging to generate coherent responses. To this end, we aim to source important and useful reviews that are coherent with the ongoing conversation [24].

Concretely, when considering the entire set of reviews $\mathcal{R}$, the key is to select those reviews that exhibit a similar sentiment polarity to the conversational history. Suppose that $\mathcal{R}_v^{(c)} = \{r_1, r_2, \cdots, r_n\} \in \mathcal{R}$ represents the reviews associated with the item mentioned in the conversation history C. For each review $r_j = \{w_1, w_2, \cdots, w_m\} \in \mathcal{R}_v^{(c)}$, we employ a transformer-based sentiment predictor [24] to predict its sentiment polarity. Here, $\mathcal{H}^{(l-1)}(\boldsymbol{r}_j)$ represents the output embeddings of the previous transformer layer, and the output of the current layer $\mathcal{H}^{(l)}(\boldsymbol{r}_j)$ can be defined using the *Multi-head Attention Function* $\mathsf{MHA}(\cdot,\cdot,\cdot)$ as follows:

$$\begin{aligned}
\mathcal{H}^{(l)}(\boldsymbol{r}_j) &= \mathsf{MHA}(\mathcal{H}^{(l-1)}(\boldsymbol{r}_j), \mathcal{H}^{(l-1)}(\boldsymbol{r}_j), \mathcal{H}^{(l-1)}(\boldsymbol{r}_j),\\
\mathsf{MHA}(\boldsymbol{K}, \boldsymbol{Q}, \boldsymbol{V}) &= [\boldsymbol{h}_1^l, \boldsymbol{h}_2^l, \cdots, \boldsymbol{h}_h^l]\boldsymbol{W}_j^l,\\
\boldsymbol{h}_j^l &= \mathsf{SA}(\mathcal{H}^{(l)}(\boldsymbol{r}_j)\boldsymbol{W}_j^k, \mathcal{H}^{(l)}(\boldsymbol{r}_j)\boldsymbol{W}_j^q, \mathcal{H}^{(l)}(\boldsymbol{r}_j)\boldsymbol{W}_j^v),\\
\mathsf{SA}(\boldsymbol{K}, \boldsymbol{Q}, \boldsymbol{V}) &= \mathrm{Softmax}(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d/h}})\boldsymbol{V}.
\end{aligned} \tag{2}$$

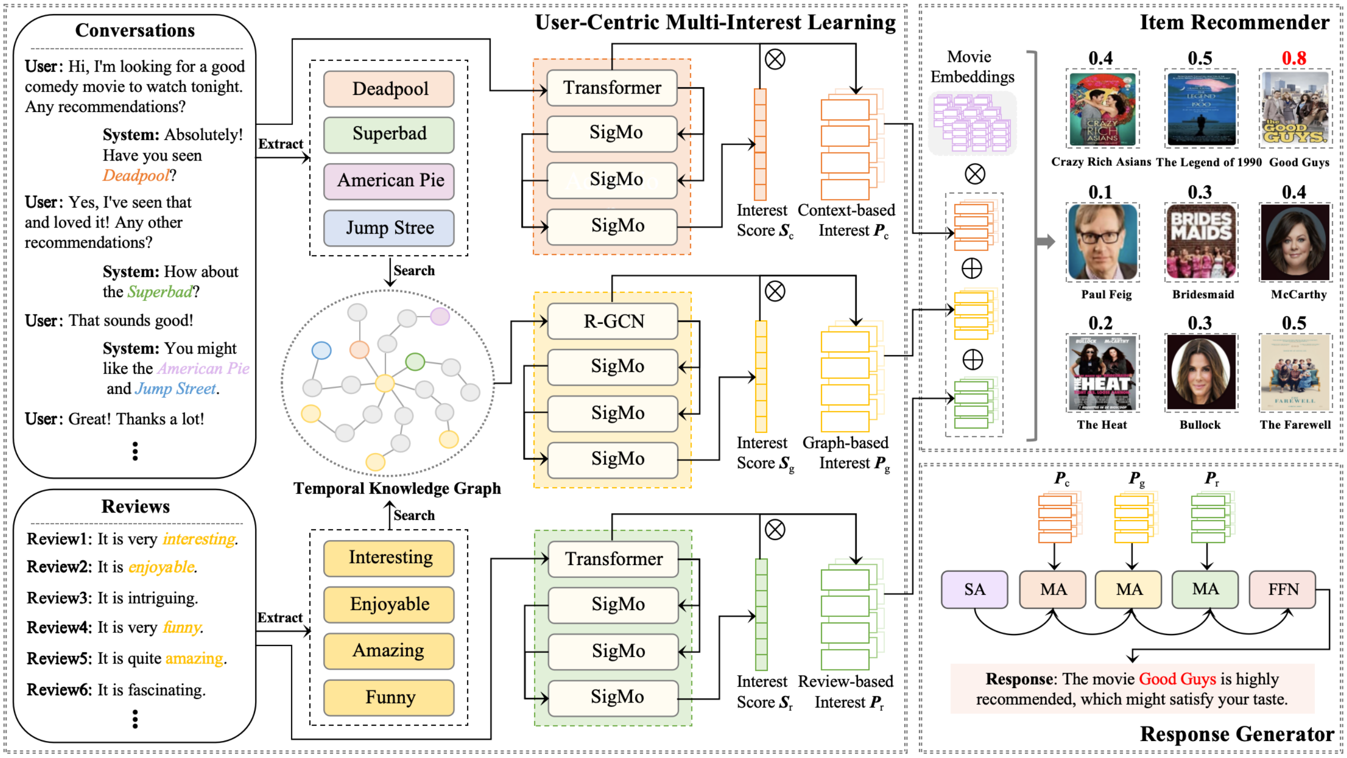
Here $\boldsymbol{r}_j = \{\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_m\}$ is the embedding of the review r_j, and $\boldsymbol{w}_i$ denotes the embedding of each word w, h represents the number of heads, $\boldsymbol{W}_j^l$ is a learned parameter during model training, and each head $\boldsymbol{h}_j^l$ is calculated using the attention mechanism $\mathsf{SA}(\cdot,\cdot,\cdot)$. In this attention mechanism, $\boldsymbol{K}, \boldsymbol{Q}$, and $\boldsymbol{V}$ denote the key, query, and value matrices, respectively, while $\boldsymbol{W}_j^k$, $\boldsymbol{W}_j^q$, and $\boldsymbol{W}_j^v$ are trainable parameters. For simplicity, we consider the output embeddings of the top transformer layer as the final review representations $\boldsymbol{H}$. Formally, this process can be described as follows:

$$\begin{aligned}
\boldsymbol{H} &= \mathsf{MHA}(\mathcal{H}^{(L-1)}(\boldsymbol{r}_j), \mathcal{H}^{(L-1)}(\boldsymbol{r}_j), \mathcal{H}^{(L-1)}(\boldsymbol{r}_j),\\
\boldsymbol{P}_v &= \mathrm{Softmax}(\boldsymbol{W}_1 \tanh(\boldsymbol{W}_2 \boldsymbol{H}^T)).
\end{aligned} \tag{3}$$

L represents the number of transformer layers, and $\boldsymbol{P}_v$ denotes the predicted sentiment towards the movie v in review r_j. Similarly, we use this transformer-based sentiment prediction to evaluate the sentiment polarity $\boldsymbol{P}_v^*$ for the conversation sentence that mentions the movie v. Ultimately, we select reviews that share a similar sentiment polarity with the conversation sentence to establish the retrieved reviews $\widetilde{\mathcal{R}}_v^{(c)}$, which can be written as:

$$\begin{aligned}
\widetilde{\mathcal{R}}_v^{(c)} &= \{\tilde{r}_1, \tilde{r}_2, \cdots, \tilde{r}_{\tilde{n}}\},\\
\widetilde{\mathcal{R}}_v^{(c)} &\in R_v^{(c)}, \tilde{n} << n.
\end{aligned} \tag{4}$$

Note that there are multiple reviews related to the item v that are currently being discussed in the conversation history C. To simplify



Conversations

User: Hi, I'm looking for a good comedy movie to watch tonight. Any recommendations?

System: Absolutely! Have you seen Deadpool?

User: Yes, I've seen that and loved it! Any other recommendations?

System: How about the Superbad?

User: That sounds good!

System: You might like the American Pie and Jump Street.

User: Great! Thanks a lot!

Reviews

Review1: It is very interesting.

Review2: It is enjoyable.

Review3: It is intriguing.

Review4: It is very funny.

Review5: It is quite amazing.

Review6: It is fascinating.

Response: The movie Good Guys is highly recommended, which might satisfy your taste.

Figure 1: Overview of the proposed framework, CoMoRec, which is comprised of User-Centric Multi-Interest Learning and Interest-Enhanced CRS. The former devotes to adaptively model multi-aspect user interests including context-based, graph-based, and review-based interests by incorporating conversations, temporal knowledge graph, and item reviews; the latter aims to make diverse movie predictions in the recommendation task (*i.e.*, item recommender) and generate informative responses in the conversation task (*i.e.*, response generator) by employing these learned multi-aspect user interests.

the process, we choose to select one sentence for each mentioned item. To achieve this, we employ a word-wise method to randomly select a set of words or phrases to construct each "sentence". While this strategy may affect the fluency of the sentences, it brings forth a substantial level of diversity given the extensive pool of words and phrases accessible. The definitive review, represented as r_v (*i.e.*, one sentence), for each item v mentioned in the conversation responses C can be written as follows:

$$r_v = \mathcal{F}_{\text{retrieve}}(\widetilde{\mathcal{R}}_v^{(c)}, \mathcal{V}, \mathcal{W}, C), \tag{5}$$

where $\mathcal{F}_{\text{retrieve}}(\cdot)$ signifies the review retrieval function, $\mathcal{W}$ represents all the words in $\widetilde{\mathcal{R}}_v^{(c)}$. Upon acquiring the found review, r_v, it is incorporated into the review set $\mathcal{R}$, which houses reviews of items that have appeared in conversation C. Once we've garnered the pivotal reviews, we go on to extract the entities $\mathcal{E}_r$ from the selected set of reviews, $\mathcal{R}$. At the time step t, this can be articulated:

$$\mathcal{E}_r^{(t)} = \mathcal{F}_{\text{extract}}(\mathcal{R}, \mathcal{V}, \mathcal{E}, \mathcal{G}). \tag{6}$$

Graph Relations Extraction. Finally, we combine the conversation entities $\mathcal{E}_c^{(t)}$ and the review entities $\mathcal{E}_r^{(t)}$ by concatenating them, allowing us to model diverse user-item interactions and accurately capture structural information. This process can be represented as:

$$\mathcal{E}_g^{(t)} = (\mathcal{E}_c^{(t)} \oplus \mathcal{E}_r^{(t)}), \tag{7}$$

where $\oplus$ represents the concatenation operation. The resulting set of entities $\mathcal{E}_g^{(t)}$ is regarded as the seed set. Subsequently, we extract one-hop triples from the DBpedia graph using these seed entities, thereby constructing our target TKG $\mathcal{G}_t$, as shown below:

$$\mathcal{G}_t = \text{One-hop}(\mathcal{E}_g^{(t)}) = \{(e, r, e') \mid \text{Triple}(e, r, e') \in \mathcal{G},\ e \in \mathcal{E}_g^{(t)}\}. \tag{8}$$

Here e and e' are the head and tail entities while r is the relation between them.

3.1.2 *Multi-Interest Modeling*.

After building the TKG, we model multi-aspect user interests (i.e., context-based interest, graph-based interest, and review-based interest) by utilizing conversation contexts, TKG, and item reviews, aiming to enhance recommendation diversification as users progressively interact with the system over time in the CRS.

Context-based Interest. In contrast to fixed user profiles or direct input from users, conversations offer a more detailed perspective on user preferences by capturing the ever-changing nature of their interests. The continuous exchange of dialogues presents an opportunity to comprehend the evolving interests of users. Taking into account the conversational context, sentiment, and the variety of discussed topics, the CRS can dynamically adjust its recommendations to cater to the user's current interests. This approach enables a more personalized and customized recommendation experience that aligns with the user's current preferences. To do this, we employ the Transformer as the encoder to efficiently encode the conversations and derive their corresponding representations inspired by the valuable attributes of the Transformer model [35]. Given a conversation context C, let $\mathcal{H}^{(l-1)}(C)$ be the output embeddings of the previous transformer layer, and the output of the current layer $\mathcal{H}^{(l)}(C)$ can be defined by the $\mathsf{MHA}(\cdot,\cdot,\cdot)$ as:

$$\mathcal{H}^{(l)}(C) = \mathsf{MHA}(\mathcal{H}^{(l-1)}(C), \mathcal{H}^{(l-1)}(C), \mathcal{H}^{(l-1)}(C)). \quad (9)$$

The specifics of the function $\mathsf{MHA}(\cdot,\cdot,\cdot)$ can be found in Eq. (2). To streamline the procedure, we opt for the output embedding of the final transformer layer as the ultimate representation, labeled as $\boldsymbol{H}_c$, and can be formally defined as:

$$\boldsymbol{H}_c = \mathsf{MHA}(\mathcal{H}^{(\mathcal{L}_c-1)}(C), \mathcal{H}^{(\mathcal{L}_c-1)}(C), \mathcal{H}^{(\mathcal{L}_c-1)}(C)). \quad (10)$$

Here $\mathcal{L}_c$ represents the maximum number of transformer layers. Once we acquire the ultimate output representations $\mathcal{L}_c$, we employ them in three non-linear functions as outlined below to generate the context-based interest score:

$$\boldsymbol{S}_c = \mathrm{Softmax}(\mathrm{S_{op}}(\frac{\mathrm{SigMo}[(\mathrm{SigMo}(\mathbf{W}_q^{(c)}\boldsymbol{H}_c))(\mathrm{SigMo}(\mathbf{W}_k^{(c)}\boldsymbol{H}_c))^T]}{\sqrt{d}})), \quad (11)$$

where the function Softmax represents the softmax function used to normalize the embeddings. The function $\mathrm{S_{op}}(\cdot)$ involves the process of summing each row and subsequently averaging the rows within the embedding matrix, while masking the diagonal elements. $\mathrm{SigMo}(\cdot)$ denotes the sigmoid function, and $\mathbf{W}_q^{(c)}$ and $\mathbf{W}_k^{(c)}$ are trainable parameters. Finally, we perform the interactions between the score $\boldsymbol{S}_c$ and $\boldsymbol{H}_c$ for generating the context-based interest as:

$$\boldsymbol{P}_c = \boldsymbol{S}_c \otimes \boldsymbol{H}_c. \quad (12)$$

Graph-based Interest. Due to the sparsity of the user-item bipartite graph, we merge conversations and reviews to create the TKG by integrating the comprehensive DBpedia KG. This enables us to harness the wealth of information present in both the conversations/reviews and the DBpedia KG. As relations within the TKG play a pivotal role in uncovering users' interests, we employ Relational Graph Convolutional Networks (R-GCNs) [24] to encode structural and relational details from the extracted subgraph $\mathcal{G}_t$ into entity representations. Let e represent a node at the $(l+1)$-th layer in $\mathcal{G}_t$, and its feature representation can be calculated as:

$$\boldsymbol{h}_e^{l+1} = \sigma\left(\sum_{r\in\mathcal{R}}\sum_{\hat{e}\in\mathcal{N}_e^r}\frac{1}{z_{e,r}}\boldsymbol{W}_r^l\boldsymbol{h}_{\hat{e}}{}^l + \boldsymbol{W}^l\boldsymbol{h}_e^l\right), \quad (13)$$

Here $\boldsymbol{h}_e^l \in \mathbb{R}^d$ represents the hidden representation of entity e at the l-th layer of the graph neural network, where d denotes the feature dimensionality. The set $\mathcal{N}_e^r$ refers to the one-hop neighbor set of entity e under the relation r. The hyperparameter $z_{e,r}$ corresponds to the normalization factor. The matrices $\boldsymbol{W}_r^l$ and $\boldsymbol{W}^l$ are trainable parameters that are updated during the model training process. These matrices are used to transform and update the hidden representations of entities within the graph neural network. Let $\boldsymbol{H}_g$ be the final output embedding of R-GCN, then we can obtain the graph-based interest score:

$$\boldsymbol{S}_g = \mathrm{Softmax}(\mathrm{S_{op}}(\frac{\mathrm{SigMo}[(\mathrm{SigMo}(\mathbf{W}_q^{(g)}\boldsymbol{H}_g))(\mathrm{SigMo}(\mathbf{W}_k^{(g)}\boldsymbol{H}_g))^T]}{\sqrt{d}})). \quad (14)$$

Similarly, we can induce the graph-based interest:

$$\boldsymbol{P}_g = \boldsymbol{S}_g \otimes \boldsymbol{H}_g. \quad (15)$$

Review-based Interest. Item reviews serve as a crucial resource for discerning users' genuine intentions and preferences. Therefore, we construct the review-based interest by leveraging the valuable insights conveyed through item reviews. To achieve this, we employ the Transformer to encode the retrieved reviews and learn their representations. Given a review $\mathcal{R}$, let $\mathcal{H}^{l-1}(\mathcal{R})$ represent the output of embeddings from the previous transformer layer, and $\mathcal{H}^l(\mathcal{R})$ denote the output of the current layer. By leveraging the $\mathsf{MHA}(\cdot,\cdot,\cdot)$ function, the review-based interest $\boldsymbol{P}_r$ can be described:

$$\mathcal{H}^{(l)}(\mathcal{R}) = \mathsf{MHA}(\mathcal{H}^{(l-1)}(\mathcal{R}), \mathcal{H}^{(l-1)}(\mathcal{R}), \mathcal{H}^{(l-1)}(\mathcal{R})). \quad (16)$$

$$\boldsymbol{H}_r = \mathsf{MHA}(\mathcal{H}^{(\mathcal{L}_r-1)}(C), \mathcal{H}^{(\mathcal{L}_r-1)}(C), \mathcal{H}^{(\mathcal{L}_r-1)}(C)). \quad (17)$$

$$\boldsymbol{S}_r = \mathrm{Softmax}(\mathrm{S_{op}}(\frac{\mathrm{SigMo}[(\mathrm{SigMo}(\mathbf{W}_q^{(r)}\boldsymbol{H}_r))(\mathrm{SigMo}(\mathbf{W}_k^{(r)}\boldsymbol{H}_r))^T]}{\sqrt{d}})). \quad (18)$$

$$\boldsymbol{P}_r = \boldsymbol{S}_r \otimes \boldsymbol{H}_r. \quad (19)$$

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

3.2 Interest-Enhanced CRS

In this section, we integrate these multi-aspect user interests, namely the context-based interest $\boldsymbol{P}_c$, graph-based interest $\boldsymbol{P}_g$, and review-based interest $\boldsymbol{P}_r$, into the Interest-Enhanced CRS. By doing so, we are able to accurately predict items in the recommendation task and generate dialogue responses effectively in the conversation task.

3.2.1 *Item Recommender*.

In order to enhance the diversity of recommendation results, we leverage these multi-faceted user interests to delve into users' authentic preferences. To achieve this, we concatenate the different user interests, creating the ultimate recommendation-based user references denoted as $\boldsymbol{P}_{\mathrm{rec}}$. Next, we facilitate interactions between $\boldsymbol{P}_{\mathrm{rec}}$ and the feature embeddings of the candidate movie set to compute the rating scores. This approach enables us to effectively predict users' preferences and provide well-informed item recommendations. Formally, this process can be outlined as follows:

$$\begin{aligned} \boldsymbol{P}_{\mathrm{con}} &= [\boldsymbol{P}_c \oplus \boldsymbol{P}_g \oplus \boldsymbol{P}_r]; \\ \boldsymbol{P}_{\mathrm{rec}} &= \mathrm{Softmax}(\mathrm{MLP}(\boldsymbol{P}_{\mathrm{con}})); \\ V_{\mathrm{sco}} &= \boldsymbol{P}_{\mathrm{rec}} \otimes \boldsymbol{v}. \end{aligned} \quad (20)$$

Here $\oplus$ means the concatenation operation, MLP is the Multilayer Perceptron Layer, $\boldsymbol{v}$ represents the feature embedding of the movie item v, while V_{sco} corresponds to the user's rating score for the item v. Then, we adopt the cross-entropy to train the recommendation parameters. Formally, the cross-entropy loss $\mathcal{L}_r$ between the prediction $\boldsymbol{P}_{\mathrm{rec}}$ and the target item category can be computed as:

$$\mathcal{L}_r = -\frac{1}{N}\sum_{j=1}^{N}\log P_{\mathrm{rec}}^j, \quad (21)$$

here N is the number of total recommendations and P_{rec}^{j} denotes the target category in the j-th recommendation.

3.2.2 ***Response Generator.*** To effectively generate diverse responses, we incorporate the multi-aspect user interests $\boldsymbol{P}_c$, $\boldsymbol{P}_g$, and $\boldsymbol{P}_r$ into a multi-head attention network to predict the next utterances. The main reason for adopting these attention layers is to seamlessly integrate the entities from the knowledge graph (KG) and reviews into the context information, following the approach of previous work [24]. Furthermore, we augment the attention mechanism to enhance data representations and filter out noise by leveraging this multi-aspect knowledge, as illustrated below:

$$\begin{aligned}
\mathbf{A}_0^i &= \text{MHA}(\mathbf{Y}^{i-1}, \mathbf{Y}^{i-1}, \mathbf{Y}^{i-1}),\\
\mathbf{A}_1^i &= \text{MHA}(\mathbf{A}_0^i, \boldsymbol{P}_c, \boldsymbol{P}_c),\\
\mathbf{A}_2^i &= \text{MHA}(\mathbf{A}_1^i, \boldsymbol{P}_g, \boldsymbol{P}_g),\\
\mathbf{A}_3^i &= \text{MHA}(\mathbf{A}_2^i, \boldsymbol{P}_r, \boldsymbol{P}_r),\\
\mathbf{Y}^i &= \text{FFN}(\mathbf{A}_3^i).
\end{aligned} \tag{22}$$

Here, $\mathbf{Y}^{i-1}$ denotes the output from the previous time step, $\mathbf{Y}^i$ represents the current output, and $\text{MHA}(\mathbf{Q}, \mathbf{K}, \mathbf{V})$ signifies the multi-head attention module, which can be referred to as Eq. (2). Additionally, $\text{FFN}(\cdot)$ corresponds to the fully-connected feed-forward network comprising two linear layers with a ReLU activation [24]. For the conversation task, we employ the cross-entropy loss [24] as the learning objective for response generation. Formally, the loss can be described as:

$$\mathcal{L}_c = -\frac{1}{M}\sum_{t=1}^{M}\log(\text{Prob}(s_t|s_1, \cdots, s_{t-1})), \tag{23}$$

where M is the number of turns, s_t denotes the t-th sentence in the conversation, and the function $\text{Prob}(\cdot)$ means the generation probability s_t of the next token, which can be expressed as:

$$\begin{aligned}
\text{Prob}(s_t|s_1, \cdots, s_{t-1}) = &\ \text{Prob}_{\text{v}}(s_t|\mathbf{Y}_i)\\
&+ \text{Prob}_{\text{g}}(s_t|\mathbf{Y}_i, \mathcal{G})\\
&+ \text{Prob}_{\text{r}}(s_t|\mathbf{Y}_i, \mathcal{R}),
\end{aligned} \tag{24}$$

where $\text{Prob}_{\text{v}}(\cdot)$, $\text{Prob}_{\text{g}}(\cdot)$, and $\text{Prob}_{\text{r}}(\cdot)$ are the probability functions over the vocabulary, entities from the knowledge graph $\mathcal{G}$, and reviews $\mathcal{R}$, respectively, following the previous work [16, 24].

4 Experiments and Analyses

In this section, we conduct experiments to evaluate the performance of CoMoRec on movie datasets and answer the following questions:

- **RQ1:** How does CoMoRec perform compared to state-of-the-art methods in the conversation task?
- **RQ2:** How does CoMoRec perform compared to state-of-the-art methods in the recommendation task?
- **RQ3:** How does CoMoRec enhance the recommendation diversification in the CRS?
- **RQ4:** How do the context-based interest $\boldsymbol{P}_c$, graph-based interest $\boldsymbol{P}_g$, and review-based interest $\boldsymbol{P}_r$ contribute to the performance?

4.1 Experimental Protocol

4.1.1 Datasets. We evaluate CoMoRec on two widely-adopted movie datasets: **REDIAL** [20] and **TG-REDIAL** [51]. REDIAL is an English dataset for real-world dialogues on movie recommendations, featuring 10,006 conversations about 51,699 movies. It also includes a review database with 30 reviews per movie from IMDb[1] on the previous work [24]. TG-REDIAL is a Chinese conversational recommendation dataset with 10,000 dialogues and 129,392 utterances about 33,834 movies. Each conversation starts with the first sentence and progresses to generate responses or recommendations. The review data for TG-REDIAL is sourced from Douban[2].

4.1.2 Baselines. To fully evaluate our CoMoRec, we compare our CoMoRec with a series of state-of-the-art methods in both conversational task and recommendation task. The compared methods include **Trans** [35], **Redial** [20], **KBRD** [7], **KGSF** [49], **KECRS** [45], **RevCore** [24], **KGCR** [28], **C^2-CRS** [53], **TextCNN** [18], **SASRec** [27], **BERT4Rec** [32], **TG-ReDial** [52], **BERT** [10], **BART** [19], and **MHIM** [30]. By conducting a thorough comparison with these baselines, we can effectively evaluate the performance of the proposed CoMoRec in both tasks.

4.1.3 Evaluation Metrics. Our CoMoRec comprises the conversation and recommendation tasks. For the conversation task, we use automatic evaluation and human evaluation to evaluate the performance of the response generation. For automatic evaluation, we adopt Distinct n-gram (D-n, n=2, 3, 4) [24, 53] and Bleu-m (B-m, m=2, 3) [39] to evaluate the diversity of generated response contexts at sentence level. Besides, we provide annotators to manually estimate the generated candidates in *Fluency* and *Informativeness* [53]. For the recommendation task, we adopt MRR@k (M@k, k =10, 50) and NDCG@k (N@k, k =10, 50) [48] as the evaluation metrics.

5 Performance Comparison

5.1 Evaluation on Conversation Task (RQ1)

5.1.1 Automatic Evaluation. Table 1 showcases the experimental results, underscoring the superior performance of our model in comparison to other competitive methods. Specifically, in the REDIAL and TG-REDIAL datasets, ReDial surpasses Trans on D-2 by 22.4% and 3.8%, respectively. This success can be attributed to ReDial's utilization of a pre-trained RNN model to enhance representations of past conversations. However, it's noteworthy that both KBRD and KGSF exhibit even greater performance than ReDial on D-2, with a relative improvement of 4.9% and 39.0% in the REDIAL dataset, respectively. The integration of additional information, such as DBpedia, in KBRD and KGSF enhances feature representation learning, leading to enhanced performance. Moreover, RevCore outperforms KBRD by 7.0% on D-2 in the REDIAL dataset. This advancement is a result of RevCore's capability to retrieve pertinent reviews and integrate them into the dialogue context, thereby enriching the overall comprehension of the conversation. Furthermore, C^2-CRS surpasses various competitive baselines, including KBRD, KGSF, and RevCore, with improvements of 89.5%, 43.0%, and 77.2% on D-2 in the REDIAL dataset, respectively. The exceptional performance of C^2-CRS can be attributed to its innovative contrastive learning-based coarse-to-fine strategy, which effectively merges diverse data representations and enhances dialogue understanding.

[1] https://www.dbpedia.org/

[2] https://movie.douban.com/

Datasets	REDIAL							TG-REDIAL						
Models	D-2	D-3	D-4	B-2	B-3	Flu.	Inf.	D-2	D-3	D-4	B-2	B-3	Flu.	Inf.
Trans	0.067	0.139	0.227	0.0164	0.0027	0.97	0.92	0.053	0.121	0.204	0.0335	0.0075	0.81	0.83
ReDial	0.082	0.143	0.245	0.0198	0.0054	1.35	1.04	0.055	0.123	0.215	0.0387	0.0094	0.98	0.101
KBRD	0.086	0.153	0.265	0.0203	0.0061	1.23	1.15	0.045	0.096	0.233	0.0411	0.0107	0.112	0.115
KGSF	0.114	0.204	0.282	0.0211	0.0067	1.48	1.37	0.086	0.186	0.297	0.0442	0.0128	1.21	1.30
KECRS	0.040	0.090	0.149	0.0124	0.0042	1.39	1.19	0.047	0.114	0.193	0.0319	0.0053	0.86	0.90
RevCore	0.092	0.163	0.221	0.0219	0.0083	1.52	1.34	0.043	0.105	0.175	0.0431	0.0118	1.21	1.28
C^2-CRS	0.163	0.291	0.417	0.0223	0.0088	1.55	1.47	0.189	0.334	0.424	0.0434	0.0120	1.25	1.37
MHIM	0.164	0.293	0.415	0.0226	0.0089	**1.60**	1.44	0.186	0.333	0.426	0.0435	0.0118	1.28	1.35
CoMoRec	**0.167***	**0.298***	**0.421***	**0.0230***	**0.0089***	1.57	**1.51***	**0.192***	**0.342***	**0.428***	**0.0437***	**0.0125***	**1.32***	**1.39***

Table 1: Conversation results. Flu. and Inf. stand for Fluency and Informativeness, respectively. Numbers marked with * denote that there is a statistically significant improvement compared with the best baseline (t-test with p-value $<$ 0.05).

From Table 1, it is apparent that our CoMoRec outshines all competing methods on both datasets. For example, on the REDIAL dataset, CoMoRec surpasses Trans, ReDial, KBRD, KGSF, KECRS, RevCore, and C^2-CRS by 149.3%, 103.7%, 94.2%, 46.5%, 317.5%, 81.5%, and 2.5% in D-2, respectively. On the TG-REDIAL dataset, CoMoRec outperforms these methods by 262.3%, 249.0%, 326.7%, 123.3%, 308.5%, 346.5%, and 1.6% in D-2, respectively. Notably, CoMoRec also consistently exceeds all state-of-the-art methods in both B-2 and B-3 metrics, showcasing the effectiveness of our approach for response generation. The improvement of CoMoRec can be attributed to its focus on enhancing response diversity by modeling multi-aspect user interests by user-system dialogues. By comprehensively considering and capturing various facets of user preferences and interests, CoMoRec enhances its capacity to generate diverse and personalized responses, thereby achieving superior performance compared to other baselines.

5.1.2 Human Evaluation. Table 1 encapsulates the results of the human evaluation in the conversation task in terms of *Fluency* and *Informativeness* metrics. There are several key observations: 1) ReDial demonstrates superior performance over Transformer, owing to the implementation of a pre-trained RNN encoder. This encoder significantly enhances the quality and fluency of the responses generated by the system. 2) KGSF excels in terms of *Informativeness* compared to various other baselines. This achievement can be attributed to its integration of an external information knowledge graph, which effectively aligns the semantics of the conversation context with the items discussed. 3) RevCore achieves the highest level of performance in terms of *Fluency* when compared to several other baselines. This success can be attributed to its utilization of additional reviews to enhance the decoder, resulting in the generation of more coherent and fluent responses. 4) C^2-CRS emerges as the top performer in terms of *Informativeness* among all the baselines. The efficacy of C^2-CRS can be credited to its emphasis on integrating diverse data types to generate informative words and entities. Notably, CoMoRec consistently outperforms all the compared methods in terms of both metrics. This success can be attributed to its utilization of multi-aspect knowledge to model various levels of user interests, enabling the generation of fluent and diverse responses.

Datasets	REDIAL				TG-REDIAL			
Models	M@10	M@50	N@10	N@50	M@10	M@50	N@10	N@50
TextCNN	0.0235	0.0285	0.0328	0.0580	0.0040	0.0045	0.0053	0.0077
SASRec	0.0540	0.0593	0.0674	0.0936	0.0011	0.0017	0.0019	0.0047
BERT4Rec	0.0475	0.0555	0.0663	0.1045	0.0013	0.0020	0.0020	0.0058
ReDial	0.0677	0.0738	0.0925	0.1222	0.0012	0.0017	0.0018	0.0045
TG-ReDial	0.0694	0.0771	0.0924	0.1286	0.0048	0.0050	0.0062	0.0076
KBRD	0.0722	0.0800	0.0972	0.1333	0.0077	0.0090	0.0106	0.0171
KGSF	0.0705	0.0796	0.0956	0.1379	0.0069	0.0087	0.0103	0.0194
BERT	0.0597	0.0688	0.0831	0.1255	0.0011	0.0017	0.0018	0.0050
BART	0.0646	0.0744	0.0888	0.1350	0.0012	0.0017	0.0020	0.0048
MHIM	0.0742	0.0830	0.1027	0.1440	0.0108	0.0129	0.0152	0.0256
CoMoRec*	**0.0785**	**0.0871**	**0.1086**	**0.1532**	**0.0135**	**0.0144**	**0.0153**	**0.0263**

Table 2: Recommendation results. Numbers marked with * denote that there is a statistically significant improvement compared with the best baseline (t-test with p-value < 0.05).

5.2 Evaluation on Recommendation Task (RQ2)

Table 2 summarizes the experimental results on the recommendation task for movie prediction. The results clearly demonstrate the superiority of our model over all the baselines. Firstly, the results show that both KBRD and KGSF outperform ReDial. For instance, KGSF surpasses ReDial on M@10 by 3.97% and 82.61% on the REDIAL and TG-REDIAL datasets, respectively. This improvement can be attributed to the integration of external information, such as DBpedia, which enriches the representations of items and words in both KBRD and KGSF. Additionally, both KBRD and KGSF outperform SASRec on M@10 by 25.21% and 23.40% on the REDIAL datasets, respectively. Furthermore, MHIM demonstrates better performance than KBRD, KGSF, BART, achieving a gain of approximately 2.70%, 4.99%, and 12.94% on M@10 in the REDIAL dataset, respectively. The primary reason behind this improvement is that MHIM incorporates large-scale knowledge graph to enhance the user vector representation.

It is worth noting that our CoMoRec achieves the best performance among the state-of-the-art methods. Concretely, CoMoRec outperforms BART on M@10 by 17.71%, and 91.11% on REDIAL and TG-REDIAL datasets, respectively. CoMoRec is also superior to MHIM on M@10 by 5.48%, and 20.00% on REDIAL and TG-REDIAL datasets, respectively. This improvement can be attributed to CoMoRec's innovative approach, which integrates a dynamic temporal knowledge graph to effectively tackle the sparsity issues of traditional user-item bipartite graphs. Additionally, CoMoRec models multi-aspect user interests, encompassing context-, graph-, and review-based factors. This strategy enhances recommendation diversification as users engage with the system over time through natural language conversations in the CRS.

5.3 Study on Recommendation Diversity (RQ3)

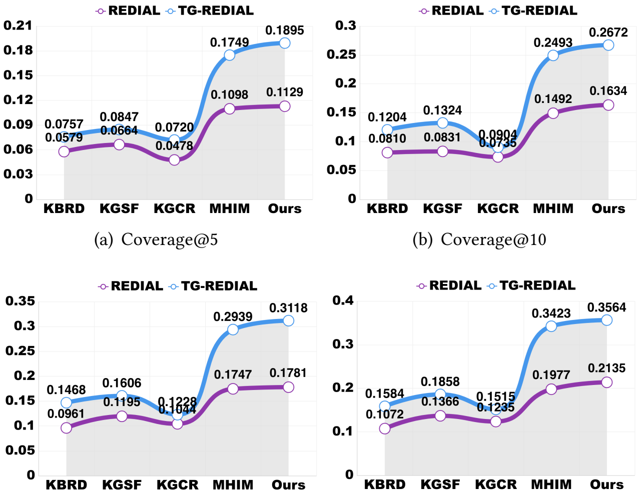
Given our primary objective of enhancing recommendation diversity as users engage with the system in the CRS, we meticulously analyze the recommendation outcomes and conduct a comprehensive comparison with the strongest baselines to assess the effectiveness of CoMoRec in achieving this goal. Along this line, we utilized the widely recognized metric *Coverage@k* (*k*=5, 10, 15, 20) to quantify the level of recommendation diversification and account for variations among the recommended items. This well-established metric allowed us to measure the extent to which our recommendations spanned a broad spectrum of the recommendation space. A higher coverage value indicates a greater capacity to encompass items from diverse categories. It signifies the ability of our system to provide recommendations that cover a wide range of item types, catering to varying user preferences and interests.

Figure 2 shows that it consistently achieves the highest *Coverage* values across all datasets when compared to the competitive baselines, validating the superiority of our CoMoRec in diversified recommendation. On the TG-REDIAL dataset, our CoMoRec exhibits significant improvements of 121.93%, 101.81%, 195.58%, and 7.18% in terms of *Coverage@10* when compared to the robust models KBRD, KGSF, KGCR, and MHIM, respectively. These compelling results underscore the effectiveness of CoMoRec in effectively mitigating isolation concerns by ensuring the comprehensive coverage of recommended items. This, in turn, provides users with a broader spectrum of choices, enhancing their overall experience. Consequently, this reinforces CoMoRec's pivotal role in enhancing recommendation diversification in the dynamic user-system feedback loop as users interact with the system over time.

5.4 Ablation Studies (RQ4)

Finally, we conduct ablation experiments with different variants of CoMoRec to verify the contributions of each component, including: 1) CoMoRec w/o $\boldsymbol{P}_c$: remove the context-based interest $\boldsymbol{P}_c$; 2) CoMoRec w/o $\boldsymbol{P}_g$: remove the graph-based interest $\boldsymbol{P}_g$; 3) CoMoRec w/o $\boldsymbol{P}_r$: remove the review-based interest $\boldsymbol{P}_r$.

The ablation results, encompassing two metrics Recall@*k* (R@*k*, *k*=1, 10, 50) and D-*n* on the Redial dataset, are summarized in Table 3 and Table 4. These results offer several noteworthy insights: 1) In the recommendation task, CoMoRec consistently outperforms other models. By integrating components ($\boldsymbol{P}_c$, $\boldsymbol{P}_g$, and $\boldsymbol{P}_r$), it effectively captures diverse user interests, leading to more precise and personalized recommendations. 2) In the conversation task, CoMoRec delivers the best performance. By synergistically combining $\boldsymbol{P}_c$, $\boldsymbol{P}_g$, and $\boldsymbol{P}_r$, it produces high-quality and coherent dialog responses, surpassing models that neglect certain components. 3) Ablation studies highlight the significance of each component in CoMoRec. $\boldsymbol{P}_c$ greatly enhances performance in both recommendation and conversation tasks, while $\boldsymbol{P}_g$ has a moderate effect on recommendations. Additionally, $\boldsymbol{P}_r$ also contributes to this task. These observations validate the effectiveness of each component.



(a) Coverage@5

(b) Coverage@10

(c) Coverage@15

(d) Coverage@20

Figure 2: Results on *Coverage* metrics.

Models	R@1	R@10	R@50
CoMoRec	**0.056**	**0.231**	**0.411**
CoMoRec w/o $\boldsymbol{P}_c$	0.052	0.225	0.408
CoMoRec w/o $\boldsymbol{P}_g$	0.054	0.228	0.406
CoMoRec w/o $\boldsymbol{P}_r$	0.055	0.230	0.404

Table 3: Ablation studies on the recommendation task.

Models	D-2	D-3	D-4
CoMoRec	**0.167**	**0.298**	**0.421**
CoMoRec w/o $\boldsymbol{P}_c$	0.157	0.265	0.413
CoMoRec w/o $\boldsymbol{P}_g$	0.161	0.278	0.419
CoMoRec w/o $\boldsymbol{P}_r$	0.164	0.282	0.420

Table 4: Ablation studies on the conversation task.

6 Conclusion

To improve recommendation diversification for movie predictions, we propose a novel paradigm, CoMoRec, comprising User-Centric Multi-Interest Learning and Interest-Enhanced CRS. The former aims to explore the wide array of user interests, including context-, graph-, and review-based interests, to enrich the result diversity for conversational movie recommendations, while the latter devotes to employ these multiple user interests to predict items and generate responses effectively. Extensive experiments on two publicly CRS-based movie datasets show that our CoMoRec achieves a new state-of-the-art performance, and the superior of improving recommendation diversification in the CRS.

7 Acknowledgments

This work was supported by National Science and Technology Major Project (No.2021ZD0111601), National Natural Science Foundation of China (No.62325605), and Guangdong Basic and Applied Basic Research Foundation (No.2023A1515011374, No.2023A1515011530), and Guangzhou Science and Technology Program (No.2024A04J6365, No. 2023A04J2030), and Guangzhou Basic Research Project for Basic and Applied Research (Project No. 202201010334), and Guangdong Province Key Laboratory of Information Security Technology, Sun Yat-sen University.

References

[1] Azin Ashkan, Branislav Kveton, Shlomo Berkovsky, and Zheng Wen. 2015. Optimal Greedy Diversity for Recommendation. In *International Joint Conference on Artificial Intelligence IJCAI*. 1742–1748.

[2] 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*, Vol. 4825. 722–735.

[3] Rubi Boim, Tova Milo, and Slava Novgorodov. 2011. Diversification and refinement in collaborative filtering recommender. In *ACM Conference on Information and Knowledge Management CIKM*. 739–744.

[4] Allan Borodin, Aadhar Jain, Hyun Chul Lee, and Yuli Ye. 2017. Max-Sum Diversification, Monotone Submodular Functions, and Dynamic Updates. *ACM Trans. Algorithms* 13, 3 (2017), 41:1–41:25.

[5] Jaime G. Carbonell and Jade Goldstein. 1998. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. In *ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 335–336.

[6] Laming Chen, Guoxin Zhang, and Eric Zhou. 2018. Fast Greedy MAP Inference for Determinantal Point Process to Improve Recommendation Diversity. In *Advances in Neural Information Processing Systems NIPS*. 5627–5638.

[7] 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* (2019).

[8] Peizhe Cheng, Shuaiqiang Wang, Jun Ma, Jiankai Sun, and Hui Xiong. 2017. Learning to Recommend Accurate and Diverse Items. In *International Conference on World Wide Web WWW*. ACM, 183–192.

[9] Konstantina Christakopoulou, Filip Radlinski, and Katja Hofmann. 2016. Towards Conversational Recommender Systems. In *International Conference on Knowledge Discovery and Data Mining*. 815–824.

[10] 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*. Association for Computational Linguistics, 4171–4186.

[11] 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*. 1273–1288.

[12] Mike Gartrell, Ulrich Paquet, and Noam Koenigstein. 2017. Low-Rank Factorization of Determinantal Point Processes. In *AAAI Conference on Artificial Intelligence*. 1912–1918.

[13] Yingqiang Ge, Shuya Zhao, Honglu Zhou, Changhua Pei, Fei Sun, Wenwu Ou, and Yongfeng Zhang. 2020. *Understanding Echo Chambers in E-Commerce Recommender Systems*. 2261–2270.

[14] Jennifer Gillenwater, Alex Kulesza, Emily B. Fox, and Benjamin Taskar. 2014. Expectation-Maximization for Learning Determinantal Point Processes. In *Neural Information Processing Systems 2014*. 3149–3157.

[15] Jennifer Gillenwater, Alex Kulesza, Zelda Mariet, and Sergei Vassilvitskii. 2019. A Tree-Based Method for Fast Repeated Sampling of Determinantal Point Processes. In *International Conference on Machine Learning ICML (Proceedings of Machine Learning Research, Vol. 97)*. 2260–2268.

[16] Çaglar Gülçehre, Sungjin Ahn, Ramesh Nallapati, Bowen Zhou, and Yoshua Bengio. 2016. Pointing the Unknown Words. In *ACL the Association for Computational Linguistics*. The Association for Computer Linguistics.

[17] Shirley Anugrah Hayati, Dongyeop Kang, Qingxiaoyang Zhu, Weiyan Shi, and Zhou Yu. 2020. INSPIRED: Toward Sociable Recommendation Dialog Systems. In *Empirical Methods in Natural Language Processing*. 8142–8152.

[18] Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *EMNLP*. 1746–1751.

[19] 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*. Association for Computational Linguistics, 7871–7880.

[20] Raymond Li, Samira Ebrahimi Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Chris Pal. 2018. Towards Deep Conversational Recommendations. In *Neural Information Processing Systems*. 9748–9758.

[21] Shuang Li, Yuezhi Zhou, Di Zhang, Yaoxue Zhang, and Xiang Lan. 2017. Learning to Diversify Recommendations Based on Matrix Factorization. In *Autonomic and Secure Computing*. IEEE Computer Society, 68–74.

[22] Han Liu, Yinwei Wei, Jianhua Yin, and Liqiang Nie. 2023. HS-GCN: Hamming Spatial Graph Convolutional Networks for Recommendation. *IEEE Trans. Knowl. Data Eng.* 35, 6 (2023), 5977–5990.

[23] 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*. ACM, 2723–2732.

[24] 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*. 1161–1173.

[25] Lijing Qin and Xiaoyan Zhu. 2013. Promoting Diversity in Recommendation by Entropy Regularizer. In *IJCAI International Joint Conference on Artificial Intelligence*. IJCAI/AAAI, 2698–2704.

[26] Filip Radlinski, Craig Boutilier, Deepak Ramachandran, and Ivan Vendrov. 2022. Subjective Attributes in Conversational Recommendation Systems: Challenges and Opportunities. In *AAAI Conference on Artificial Intelligence*. 12287–12293.

[27] Ruiyang Ren, Zhaoyang Liu, Yaliang Li, Wayne Xin Zhao, Hui Wang, Bolin Ding, and Ji-Rong Wen. 2020. Sequential Recommendation with Self-Attentive Multi-Adversarial Network. In *International ACM SIGIR conference on research and development in Information Retrieval*. ACM, 89–98.

[28] Rajdeep Sarkar, Koustava Goswami, Mihael Arcan, and John P. Mccrae. 2020. Suggest me a movie for tonight: Leveraging Knowledge Graphs for Conversational Recommendation.. In *International Conference on Computational Linguistics*.

[29] Chaofeng Sha, Xiaowei Wu, and Junyu Niu. 2016. A Framework for Recommending Relevant and Diverse Items. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI*. IJCAI/AAAI Press, 3868–3874.

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

[31] Harald Steck. 2018. Calibrated recommendations. In *ACM Conference on Recommender Systems*. 154–162.

[32] 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*. ACM, 1441–1450.

[33] Yueming Sun and Yi Zhang. 2018. Conversational Recommender System. In *Conference on Research & Development in Information Retrieval*. 235–244.

[34] Ioannis Tsochantaridis, Thorsten Joachims, Thomas Hofmann, and Yasemin Altun. 2005. Large Margin Methods for Structured and Interdependent Output Variables. *Journal of machine learning research* 6 (2005), 1453–1484.

[35] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. [n. d.]. Attention is All you Need. In *Neural Information Processing Systems*.

[36] Abdullah Wahbeh, Mohammad Al-Ramahi, Omar F. El-Gayar, Ahmed Elnoshokaty, and Tareq Nasralah. 2023. Conversational Agents for Mental Health and Well-being: Discovering Design Recommendations Using Text Mining. In *Hawaii International Conference on System Sciences*. 3184–3193.

[37] Hongwei Wang, Miao Zhao, Xing Xie, Wenjie Li, and Minyi Guo. 2019. Knowledge Graph Convolutional Networks for Recommender Systems. In *WWW The World Wide Web Conference*. ACM, 3307–3313.

[38] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. KGAT: Knowledge Graph Attention Network for Recommendation. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD*. ACM, 950–958.

[39] Xiaolei Wang, Xinyu Tang, Xin Zhao, Jingyuan Wang, and Ji-Rong Wen. 2023. Rethinking the Evaluation for Conversational Recommendation in the Era of Large Language Models. In *Conference on Empirical Methods in Natural Language Processing EMNLP*. Association for Computational Linguistics, 10052–10065.

[40] Yingxu Wang, Xiaoru Chen, Jinyuan Fang, Zaiqiao Meng, and Shangsong Liang. 2023. Enhancing Conversational Recommendation Systems with Representation Fusion. *ACM Transactions on the Web* 17, 1 (2023), 6:1–6:34.

[41] Zihao Wang, Kejun Zhang, Yuxing Wang, Chen Zhang, Qihao Liang, Pengfei Yu, Yongsheng Feng, Wenbo Liu, Yikai Wang, Yuntai Bao, and Yiheng Yang. 2022. SongDriver: Real-time Music Accompaniment Generation without Logical Latency nor Exposure Bias. In *ACM International Conference on Multimedia*. ACM, 1057–1067.

[42] Romain Warlop, Jérémie Mary, and Mike Gartrell. 2019. Tensorized Determinantal Point Processes for Recommendation. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 1605–1615.

[43] Mark Wilhelm, Ajith Ramanathan, Alexander Bonomo, Sagar Jain, Ed H. Chi, and Jennifer Gillenwater. 2018. Practical Diversified Recommendations on YouTube

with Determinantal Point Processes. In *ACM International Conference on Information and Knowledge Management*. ACM, 2165–2173.

[44] Ke Xu, Yuanjie Zhu, Weizhi Zhang, and Philip S. Yu. 2023. Graph Neural Ordinary Differential Equations-based method for Collaborative Filtering. In *IEEE International Conference on Data Mining, ICDM*. 1445–1450.

[45] Tong Zhang, Yong Liu, Peixiang Zhong, Chen Zhang, Hao Wang, and Chunyan Miao. 2021. KECRS: Towards Knowledge-Enriched Conversational Recommendation System. *CoRR* abs/2105.08261 (2021).

[46] 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 *International ACM SIGIR Conference on Research and Development in Information Retrieval*. 11–20.

[47] Yiming Zhang, Lingfei Wu, Qi Shen, Yitong Pang, Zhihua Wei, Fangli Xu, Bo Long, and Jian Pei. 2022. Multiple Choice Questions based Multi-Interest Policy Learning for Conversational Recommendation. In *World Wide Web*. 2153–2162.

[48] Kun Zhou, Xiaolei Wang, Yuanhang Zhou, Chenzhan Shang, Yuan Cheng, Wayne Xin Zhao, Yaliang Li, and Ji-Rong Wen. 2021. CRSLab: An Open-Source Toolkit for Building Conversational Recommender System. In *Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing ACL*. Association for Computational Linguistics, 185–193.

[49] Kun Zhou, Wayne Xin Zhao, Shuqing Bian, Yuanhang Zhou, Ji-Rong Wen, and Jingsong Yu. 2020. Improving conversational recommender systems via knowledge graph based semantic fusion. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 1006–1014.

[50] Kun Zhou, Wayne Xin Zhao, Hui Wang, Sirui Wang, Fuzheng Zhang, Zhongyuan Wang, and Ji-Rong Wen. 2020. Leveraging Historical Interaction Data for Improving Conversational Recommender System. In *International Conference on Information and Knowledge Management*. 2349–2352.

[51] Kun Zhou, Yuanhang Zhou, Wayne Xin Zhao, Xiaoke Wang, and Ji-Rong Wen. 2020. Towards Topic-Guided Conversational Recommender System. In *International Conference on Computational Linguistics*. 4128–4139.

[52] Kun Zhou, Yuanhang Zhou, Wayne Xin Zhao, Xiaoke Wang, and Ji-Rong Wen. 2020. Towards topic-guided conversational recommender system. *arXiv preprint arXiv:2010.04125* (2020).

[53] 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*. ACM, 1488–1496.

[54] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *International Conference on World Wide Web WWW*. ACM, 22–32.