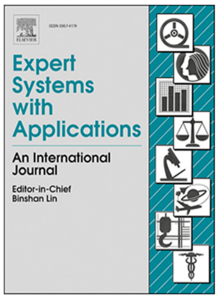
Journal Pre-proof



MACR: Multi-information Augmented Conversational Recommender

Chengyang Zhang, Xianying Huang, Jiahao An

PII: S0957-4174(22)01999-6
DOI: https://doi.org/10.1016/j.eswa.2022.118981
Reference: ESWA 118981

To appear in: *Expert Systems With Applications*

Received date : 7 July 2022
Revised date : 29 September 2022
Accepted date : 3 October 2022

Please cite this article as: C. Zhang, X. Huang and J. An, MACR: Multi-information Augmented Conversational Recommender. *Expert Systems With Applications* (2022), doi: https://doi.org/10.1016/j.eswa.2022.118981.

Chengyang Zhang

https://orcid.org/0000-0002-2443-030X

Email

zcy_cqut@163.com

Education and qualifications (2)

Chongqing University of Technology: Chongqing, Chongqing, CN

2021-09-01 to 2023-07-01 | M.E. (College of Computer Science and Engineering)

Education

Source: Chengyang Zhang

Chongqing University of Technology: Chongqing, Chongqing, CN

2017-09-01 to 2021-07-01 | B.E. (Internet of things Engineering)

Education

Source: Chengyang Zhang

Record last modified Jul 4, 2022, 9:22:50 AM

Xianying Huang

https://orcid.org/0000-0002-3667-6198

Employment (1)

Chongqing University of Technology: Chongqing, Chongqing, CN

| (College of Computer Science and Engineering)

Employment

Source: Xianying Huang

Record last modified Mar 27, 2022, 12:35:40 PM

Jiahao An

https://orcid.org/0000-0002-5887-2827

Education and qualifications (2)

Chongqing University of Technology: Chongqing, Chongqing, CN

2021-09-01 to 2023-07-01 | M.E. (College of Computer Science and Engineering)

Education

Source: Jiahao An

Weinan Normal University: Weinan, Shaanxi, CN

2017-09-01 to 2021-07-01 | B.E. (College of Computer Science)

Education

Source: Jiahao An

Record last modified Jul 4, 2022, 9:33:45 AM

MACR: Multi-information Augmented Conversational Recommender

Chengyang Zhang, Xianying Huang* and Jiahao An

College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China

ARTICLE INFO

Keywords:
Conversational Recommender System
Knowledge Graph
Category Information
Item Introduction

ABSTRACT

Conversational recommender systems (CRS) aim to provide high-quality recommendations through fewer multi-turn conversations. However, because short conversation histories lack sufficient item information, CRSs not only struggle to make accurate recommendations but also lack diversity in the generated responses. Existing CRSs mainly alleviate these problems by introducing external information (*e.g.,* reviews) while ignoring information inside the conversations (*e.g.,* potential category preferences in user utterances). Besides, item introduction is a kind of external information that is more objective and contains more entities than reviews. Therefore, we propose a Multi-information Augmented Conversational Recommender (**MACR**), which improves the performance of recommendation and response generation by mining the underlying category preferences in users' utterances and incorporating item introductions. Specifically, we enhance the category associations among entities by constructing a knowledge graph DBMG with category nodes, extracting and encoding the item categories that match the user preferences into the user representation. For item introductions, we extract the entities in them and fuse them into the conversation using an introduction-attentive encoder-decoder. Extensive experiments on the dataset REDIAL show that our MACR significantly outperforms previous state-of-the-art approaches. The source code will be available at https://github.com/zcy-cqut/MACR.

1. Introduction

Over the years, various intelligent systems have emerged, such as Q&A systems that can accurately answer users' questions (Gomes et al., 2022) and expert systems that can leverage the knowledge and experience of experts to solve domain-specific problems (Nayeri et al., 2022; Ng et al., 2022). Among them, recommendation systems have shown more significant commercial potential as they are widely used in various fields such as e-commerce (Shen et al., 2022), news (Alam et al., 2022).

Traditional recommendation systems usually rely on historical data of user-item interactions, such as e-commerce platforms recommending products to users based on their purchase history. However, interaction data is generally sparse, which makes it difficult for recommendation systems to explore users' actual preferences through such implicit feedback. To address this problem, Conversational Recommendation Systems (CRS) conduct multiple rounds of conversation with users through natural language and capture their preferences gradually during the conversation. Thus, CRS not only provides higher quality recommendations but also provides a better way of interaction.

CRS usually consists of a recommendation module and a dialogue module. The recommendation module learns user preferences based on conversation content. It provides a set of recommendations that match the user's preferences, while the dialogue module interacts with the user by generating natural language. Figure 1 shows an example of user interaction with CRS in a movie recommendation scenario. The system focuses on the user's favorite movie genre and two movies the user has seen before, after which the system provides recommendations to the user through natural language, and the whole process ends when the user accepts the recommendations.

Figure 1: An illustrative example of a conversation on movie recommendation between a user and the system. The movies and movie genre involved are marked in blue and red.

Since CRS aims to complete recommendations in fewer conversation rounds, the conversation context is usually shorter, and there is less information available to capture user preferences, making the recommendations less accurate. Existing studies enrich the context and enhance the user representation by introducing external information such as knowledge graphs (KG) (Zhou et al., 2022), relevant reviews (Lu et al., 2021). Although existing studies have improved CRS performance, there are still limitations: 1) The accuracy of the recommendations is yet relatively low, and the interpretability of the results needs to be improved; 2) The responses generated by the system lacked diversity and informativeness, and the system did hardly describe or introduce the items when recommending them. For the first problem,

*Corresponding author.
E-mail addresses: zcy_cqut@163.com (C. Zhang), wldsj_cqut@163.com (X. Huang), ajh_cqut@163.com (J. An)
ORCID(s): 0000-0002-2443-030X (C. Zhang)

traditional recommendation systems (Su et al., 2021; Seo et al., 2021; Kawai et al., 2022) often use item category information to improve recommendations' performance and interpretability. However, there is no relevant research in the field of CRS that utilizes category information. We know that users who like an item are likely to like other items of the same type, and this can be exploited in CRS. As shown in Figure 1, the categories of "*The Shawshank Redemption*" and "*Schindler's List*" mentioned by the user both contain "*Drama*", while "*The Godfather*" recommended by the system also includes this genre. As a result, this recommendation may be more appropriate to the user's preferences and help explain why the system recommended the movie, thus improving the model's recommendation effectiveness and interpretability. For the second problem, we find that item introduction are a more desirable type of potential external information that can effectively help CRS generate descriptive and informative responses. Compared with subjective reviews, item introductions are a more objective text and contain more external entities that reflect users' preferences.

To address the above issues, we propose a novel model, Multi-information Augmented Conversational Recommender (**MACR**). Multi-information augmentation refers to enhancing the overall performance of the model by category information and introduction information. For the category information, we enhance the category learning ability of the model from implicit and explicit perspectives, respectively. First, we construct a new KG, namely DBpedia with Movie Genres (**DBMG**), which allows the model to implicitly fuse category information when using the KG to obtain entity embeddings. Second, we extract the categories involving movies in the context and explicitly combine the movie category embedding and user preference embedding using a gating mechanism. For the introduction information, we construct a database of movie information from which MACR can retrieve the movie introductions mentioned in the context and then integrate them into the conversation through an introduction-attentive encoder-decoder. The enhanced context not only brings in more relevant external entities to compensate for the lack of contextual information but also helps the dialogue module generate more diverse and descriptive utterances. To the best of our knowledge, this is the first time category information and introduction information have been used to address the above CRS issues. Extensive experiments based on benchmark datasets have shown that MACR is significantly more effective in item recommendation and dialogue generation.

The contributions of this paper are summarized as follows: 1) We construct a movie knowledge graph DBMG with category information. 2) We extract the genres of movies in context and incorporated their embedding into user preferences. 3) We use movie introductions to compensate for the lack of contextual information, enhance the diversity of system responses and make it capable of generating descriptive statements.

2. Related Work

CRS is divided into two main categories, attribute-centric CRS and open-ended CRS, and MACR belongs to the latter. In this section, we will separately introduce the work related to these two types of CRSs.

2.1. Attribute-centric CRS

Attribute-centric CRS asks users classification questions about their preferences based on item attributes, such as "*What type of movies do you like?*". This type of CRS determines user preferences within a specific range by gradually reducing the attribute space of items. The difficulty lies in grasping the timing of questions and recommendations, so it usually has a strategy module to help its judgment. Researchers have proposed various questioning strategies to improve the strategy module, such as memory network based approach (Zhang et al., 2018), generalized binary search based approaches (Zou and Kanoulas, 2019; Zou et al., 2020), reinforcement learning based approaches (Chen et al., 2018; Lei et al., 2020; Ren et al., 2021; Deng et al., 2021; Li et al., 2021). These work using pre-defined dialogue templates with slots that only need to be populated with the appropriate item attributes or recommended items when asked or recommended by the system. This type of system is popular in the industry because it is easy to implement, and the conversational content is more stable, *e.g.,* intelligent customer service. However, because of their lack of flexibility and interactivity, the user experience in a real-world environment is often not very good.

2.2. Open-ended CRS

Instead of pre-defined templates and strategy modules, the open-ended CRS uses a dialogue module to generate natural language and uses a switching mechanism (Gulcehre et al., 2016) or CopyNet (Gu et al., 2016) to incorporate the recommendation results into the responses. Thus, the open-ended CRS makes up for the shortcomings of the attribute-centric CRS with a freer format and greater interactivity. Li et al. (2018) proposed the first HRED-based (Serban et al., 2017) CRS model and released the dataset of CRS in a movie recommendation scenario. Chen et al. (2019) uses KG for the first time to enhance the representation of the items. Zhou et al. (2020a) fuses entity-level KG and word-level KG and eliminates the semantic gap between the two KGs by Mutual Information Maximization (MIM). Zhou et al. (2020b) proposes the use of topics to guide conversations. Liang et al. (2021) designs dialogue template generators to further combine the recommendation results into the system responses. Lu et al. (2021) enriches the information of the context by introducing reviews while enhancing the diversity of response generation. Zhou et al. (2022) proposes a coarse-to-fine contrastive learning approach to better integrate external information.

Existing studies (Chen et al., 2019; Zhou et al., 2020a) are more concerned with introducing external information and do not fully utilize the information inside the conversation (*e.g.,* category preferences in user utterances). In fact,

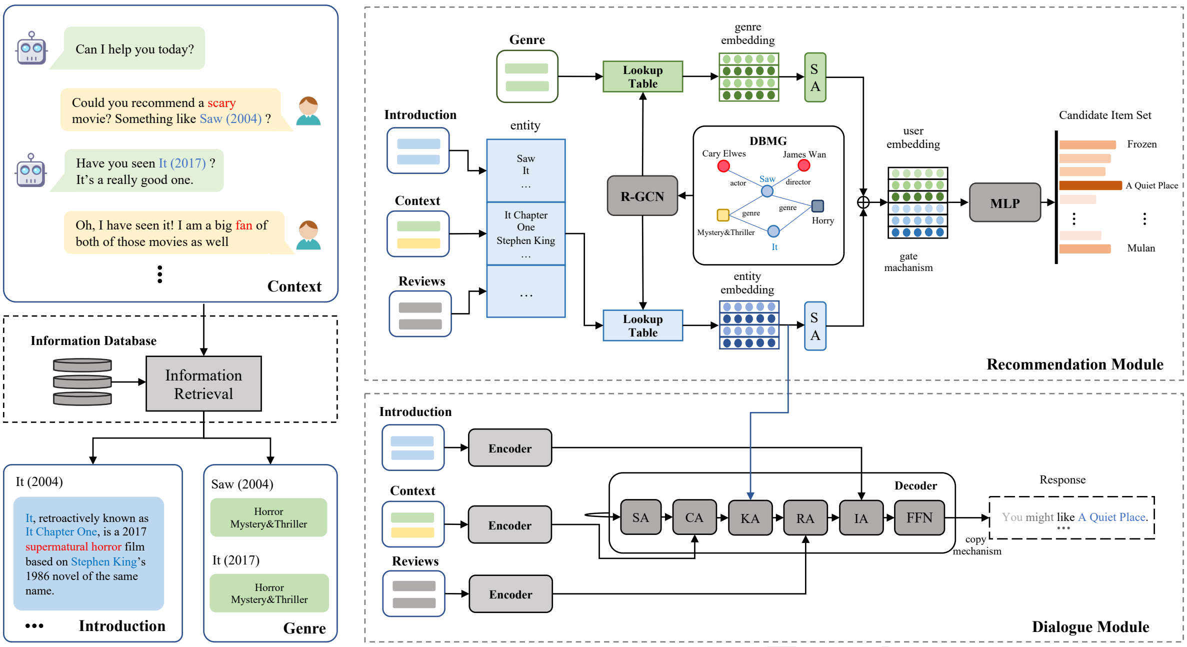


Figure 2: The overview of our model. The blue font in the context indicates entities, and the red font indicates keywords. The "SA", "CA", "KA", "RA", and "IA" in the dialogue module indicate self-attention, context-based attention, KG-based attention, review-based attention, and introduction-based attention, respectively.

the potential information in user utterances can effectively enhance the representation of user preferences. In addition, the reviews commonly used in existing studies (Lu et al., 2021; Zhou et al., 2022) do not sufficiently improve recommendation performance and interpretation because they contain fewer entities and consist of users' subjective opinions about the items. In contrast, item introduction is a more desirable type of external information, which has more entities and describes the item objectively. Therefore, we propose MACR to leverage category and introduction information for enhancement.

3. Preliminary

The open-ended CRS consists of a recommendation module, which uses entities mentioned by the user in the context of a conversation to predict their preferences and recommend items, and a dialogue module, which uses conversational statements to train its response generation capabilities. Hence there are three main objects in CRS: the entities mentioned by the user, the items recommended by the system (in this case, movies), and the sentences in the context.

Formally, we use $e \in \mathcal{E}$ and $m \in \mathcal{M}$ to denote entities and movies, respectively, where the entities are extracted from the dialogues by the recommendation module based on a specific KG. Usually, a dialogue context C consists of n rounds of dialogue utterances, which can be expressed as $C = \{s_t\}_{t=1}^{n}$, where s_t is the sentence of the t-th round of conversation. At round t, the recommendation module will select a subset $\mathcal{M}_t$ of candidate movies from the movie set $\mathcal{M}$, and the dialogue module will generate a reply s_t containing the movie $m \in \mathcal{M}_t$ to recommend to the user. It is worth noting that not every conversation will be recommended, meaning that the candidate movie subset $\mathcal{M}_t$ may be the empty set $\varnothing$. In this case, the dialogue module generates chit-chat utterances to explore the user's interest preferences.

4. Proposed MACR

In this section, we present the proposed Multi-information Augmented Conversational Recommender, called **MACR**. As shown in Figure 2, MACR consists of three main modules: the information retrieval module, the recommendation module and the dialogue module, where the recommendation module contains the knowledge graph DBMG we have constructed. First, we enhance the category representation of the movie by constructing a knowledge graph DBMG (Section 4.1). Then, we use the information retrieval module to retrieve the genres and relevant introductions of the movies mentioned in the dialogues (Section 4.2). Next, we will use the movie genre and introduction information to enhance the performance of the recommendation module (Section 4.3) and the dialogue module (Section 4.4). Finally, we present the training algorithm of MACR (Section 4.5).

4.1. Constructing Knowledge Graph DBMG

The KG usually consists of the triple (h, r, t), where h and t denote the head entity and the tail entity, and r denotes the relationship between the two. The entity-level KG used in existing research is mainly DBpedia. However, the relationship between movie entities and genres is not established in DBpedia. Therefore, we propose DBpedia

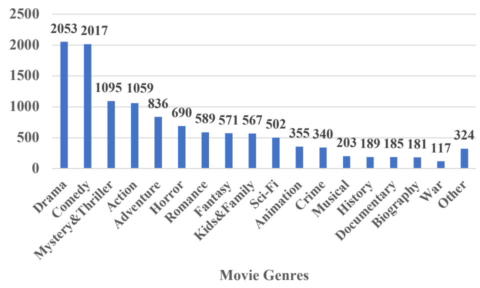


Figure 3: Statistical results of movie entities in the dataset

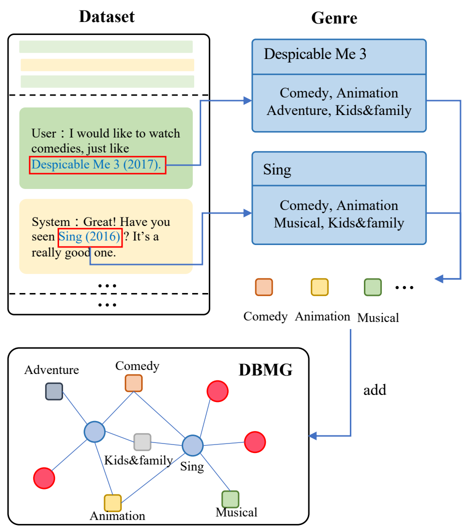


Figure 4: The construction process of DBMG. Square nodes represent category entities, blue round nodes represent DBpedia's movie entities, and red round nodes represent DBpedia's non-movie entities (*e.g.*, directors and actors).

with Movie Genres (**DBMG**) to incorporate movie genre information into the KG. Based on data from the movie website Rotten Tomatoes, we collected and organized the movie categories in the dataset REDIAL. As shown in Figure 3, there are roughly 18 genres of movies. Then, we add these 18 movie genres to DBpedia as new tail entities and a new relationship "*genre_is*" to the corresponding movie entities. It is important to note that a film may belong to more than one genre. The construction process of DBMG is shown in Figure 4. In addition, to efficiently obtain information about entities in the conversation history, we take a similar approach as KBRD (Chen et al., 2019). We first match each movie in the movie set $\mathcal{M}$ by name with the entities in the DBMG. Then we perform entity linking for the conversation content, thus obtaining all the non-movie entities in the conversation. These movie entities and non-movie entities mentioned by the user can represent the user's preferences.

4.2. Information Retrieval

We collected introductions to the movies in the dataset from Wikipedia and organized them into the information database I_{db} together with the movie genres contained in Section 4.1. The information retrieval module can identify movies in dialogue context C and then retrieve the corresponding introductions and genres from the information database I_{db}.

For the introductions, considering that movie introductions are often a long text, we only use a part of the introductions. Different from Lu et al. (2021), which randomly selects a portion of words or phrases in the review, we randomly select a whole sentence in the introduction I_{intro}. Although the random selection of words or phrases allows for more flexibility in system responses, it may also result in the selection of more non-entity words, making it difficult to take advantage of the rich entity information in the introduction. In addition, the latter generates responses with better fluency because it retains the integrity of the sentence. For movie genres, the information retrieval module will retrieve the genre to which the movie belongs based on its name. The calculation of the information retrieval module is conducted as follows:

$$I_{intro}, I_{genre} = \text{Retrieve}\left(C, I_{db}\right) \tag{1}$$

where Retrieve $(\cdot)$ denotes the retrieval operation, I_{intro} and I_{genre} represent the introduction set and genre set obtained after retrieval, respectively.

4.3. Multi-information Augmented Recommendation

The recommendation module is built on the RevCore framework (Lu et al., 2021), while we keep the movie reviews since a small number of entities also exist in the reviews. In past studies, recommendation modules used the entities involved in a conversation to represent user preferences. However, the number of entities involved in some specific conversations may be small, resulting in predicted user preferences that may not be accurate enough. In our method, the user's movie genre preference can complement the entity preference, thus improving the recommendation effect. In addition, we introduce movie introductions to add more external entities that can represent user preferences, thus making them more comprehensive.

Specifically, MACR first extracts the entities $E^{(C)}$, $E^{(R)}$, and $E^{(I_{intro})}$ contained in each conversation context C, the related reviews R, and the related introductions I_{intro}, respectively, and then merges all the entities afterward. The process of extracting and merging entities is shown as follows:

$$\begin{aligned} E^{(C)} &= \text{Extract}\,(C) \\ E^{(R)} &= \text{Extract}\,(R) \\ E^{(I_{intro})} &= \text{Extract}\left(I_{intro}\right) \\ E^{(CRI_{intro})} &= \{E^{(C)}, E^{(R)}, E^{(I_{intro})}\} \end{aligned} \tag{2}$$

where Extract $(\cdot)$ denotes the entity extraction operation, and $E^{(CRI_{intro})}$ denotes the merged set of entities.

For the entity-level knowledge graph DBMG, we must consider the semantic relationships between entities. Therefore, we use R-GCN (Schlichtkrull et al., 2018; Lu et al., 2021; Zhou et al., 2022) to obtain the entity embedding dictionary $\mathcal{D}$ for DBMG so that the relationships between entities and the relationships between movies and genres are also represented together. Formally, the representation of entity e at $(l+1)$-th layer is calculated as:

$$h_e^{(l+1)} = \sigma(\sum_{r\in\mathcal{A}} \sum_{e'\in\mathcal{E}_e^r} \frac{1}{\mathrm{Z}_{e,r}} \mathrm{W}_r^{(l)} h_{e'}^{(l)} + \mathrm{W}^{(l)} h_e^{(l)}) \tag{3}$$

where $h_e^{(l)} \in \mathbb{R}^d$ is the node representations of n at the l-th layer, d denotes the embedding dimension, r is entity relation from the relation set $\mathcal{A}$, $\mathcal{E}_e^r$ denotes the set of neighboring nodes for n under the relation r, $\mathrm{W}_r^{(l)}$ and $\mathrm{W}^{(l)}$ are learnable matrics and $\mathrm{Z}_{e,r}$ is a normalization factor.

Next, the entity set $E^{(CRI_{intro})}$ and the movie genre set I_{genre} are queried into the embedding dictionary $\mathcal{D}$, respectively, to obtain the embedding vectors of entities and genres. Then they are concatenated into the entity embedding matrix $\mathbf{E}_e \in \mathbb{R}^{a\times d}$ and the genre embedding matrix $\mathbf{E}_g \in \mathbb{R}^{b\times d}$, where a and b denote the number of entities and genres, respectively. Finally, after aggregating $\mathbf{E}_e$ and $\mathbf{E}_g$ through the self-attention layer, they are then fused into the user preference embedding $u^{(eg)}$ through the gating mechanism. The specific process is shown as follows:

$$u^{(e)} = \mathrm{SA}\left(\mathbf{E}_e\right), u^{(g)} = \mathrm{SA}\left(\mathbf{E}_g\right) \tag{4}$$

$$u^{(eg)} = \beta \cdot u^{(e)} + (1-\beta)\cdot u^{(g)} \tag{5}$$

$$\beta = \sigma\left(\mathrm{W}_{gate}\left[u^{(e)}; u^{(g)}\right]\right) \tag{6}$$

where $\mathrm{SA}(\cdot)$ denotes the self-attention aggregation operation, $u^{(e)}$ and $u^{(g)}$ represent the user's entity preference embedding and genre preference embedding, respectively, and β denotes the gating probability. Based on the final user preference embedding $u^{(eg)}$, we can calculate the probability $\mathrm{P}_{rec}(m)$ of recommending movie m from movie set $\mathcal{M}$ to user u:

$$\mathrm{P}_{rec}(m) = \mathrm{softmax}\left(\mathrm{MLP}\left(u^{(eg)}\right)\right) \tag{7}$$

where $\mathrm{MLP}(\cdot)$ denotes a multi-layer perceptron. To learn the parameters in the recommendation module, we set a cross-entropy loss as the optimization objective:

$$L_{rec} = -\sum_{j=1}^{N}\sum_{i=1}^{K} y_{ij}\cdot\log\left(\mathrm{P}_{rec}^{j}(i)\right) \tag{8}$$

where N is the total number of dialogues, j is the index of current dialogues, K is the total number of movies, and i is the index of movies.

4.4. Multi-information Augmented Response Generation

The multi-information augmented dialogue module can generate either chit-chat utterances to explore the user's interests or recommendation utterances to recommend movies to the user. To improve the diversity and informativeness of the system's responses, as well as to generate both subjective opinions and objective introductions to the recommended movies when recommending them, we keep reviews and added introductions of the relevant movies.

We use an encoder-decoder architecture for the conversation task. The transformer-based approaches (Vaswani et al., 2017) have achieved excellent performance on many NLP tasks (Hu et al., 2021; Karimi Mahabadi et al., 2021; Ji and Huang, 2021; Li et al., 2022). Therefore, we use three standard transformer encoders to encode context, reviews, and introductions in the encoding stage to obtain the corresponding context embedding $\mathrm{X}^{(c)}$, reviews embedding $\mathrm{X}^{(r)}$, and introductions embedding $\mathrm{X}^{(i)}$. In the decoding stage, we use a method similar to RevCore (Lu et al., 2021) to gradually fuse the context information, entity embedding matrix, reviews information, and introductions information. The decoding process is shown as follows:

$$\begin{aligned}
\mathrm{A}_0^n &= \mathrm{MHA}\left(\mathrm{Y}^{n-1}, \mathrm{Y}^{n-1}, \mathrm{Y}^{n-1}\right)\\
\mathrm{A}_1^n &= \mathrm{MHA}\left(\mathrm{A}_0^n, \mathrm{X}^{(c)}, \mathrm{X}^{(c)}\right)\\
\mathrm{A}_2^n &= \mathrm{MHA}\left(\mathrm{A}_1^n, \mathbf{E}_e, \mathbf{E}_e\right)\\
\mathrm{A}_3^n &= \mathrm{MHA}\left(\mathrm{A}_3^n, \mathrm{X}^{(r)}, \mathrm{X}^{(r)}\right)\\
\mathrm{A}_4^n &= \mathrm{MHA}\left(\mathrm{A}_4^n, \mathrm{X}^{(i)}, \mathrm{X}^{(i)}\right)\\
\mathrm{Y}^n &= \mathrm{FFN}\left(\mathrm{A}_4^n\right)
\end{aligned} \tag{9}$$

where Y^{n-1} denotes the output of the decoder at time step n-1, $\mathbf{E}_e$ is the entity embedding matrix from the recommendation module, $\mathrm{A}_0^n, \mathrm{A}_1^n, \mathrm{A}_2^n, \mathrm{A}_3^n, \mathrm{A}_4^n$ represent the embedding of the output after the self-attention layer and the cross-attention layer with multiple informations, respectively. $\mathrm{MHA}(\mathrm{Q},\mathrm{K},\mathrm{V})$ in Eq. 9 denotes the multi-headed attention function (Vaswani et al., 2017) that takes a query matrix Q, a key matrix K, and a value matrix V as input:

$$\begin{aligned}
\mathrm{MHA}(\mathrm{Q},\mathrm{K},\mathrm{V}) &= [\mathrm{h}_1; ...; \mathrm{h}_h]\mathrm{W}^o\\
h_i &= \mathrm{Attention}\left(\mathrm{QW}_i^q, \mathrm{KW}_i^k, \mathrm{VW}_i^v\right)
\end{aligned} \tag{10}$$

where h is the number of heads, and W_i is the parameter matrix. $\mathrm{FFN}(\cdot)$ in Eq. 9 is a fully connected feed-forward neural network consisting of two linear layers with ReLu activation functions:

$$\mathrm{FFN}(x) = \mathrm{ReLU}\left(x\mathrm{W}_1 + b_1\right)\mathrm{W}_2 + b_2 \tag{11}$$

where $\mathrm{W}_1, \mathrm{W}_2$ are learnable parameters and b_1, b_2 are two bias terms.

In contrast to ordinary conversation models, CRS often requires switching mechanisms (Gulcehre et al., 2016) or CopyNet (Gu et al., 2016) to generate responses with relevant recommended items. Formally, MACR generates the next token y_i based on the last generated sequence $\{y_{i-1}\} = y_1, y_2, \dots y_{i-1}$, which is computed as shown as follows:

$$\begin{aligned}
\Pr\left(y_i|\{y_{i-1}\}\right) =& \Pr_1\left(y_i|\mathrm{Y}_i\right) + \Pr_2\left(y_i|\mathrm{Y}_i, \mathrm{KG}\right) +\\
& \Pr_3\left(y_i|\mathrm{Y}_i, R\right) + \Pr_4\left(y_i|\mathrm{Y}_i, I_{intro}\right)
\end{aligned} \tag{12}$$

where Y_i is the output of the decoder, $\mathrm{Pr}_1(\cdot)$ is the probability function for generating ordinary words from the vocabulary, $\mathrm{Pr}_2(\cdot), \mathrm{Pr}_3(\cdot), \mathrm{Pr}_4(\cdot)$ are the probability functions for words from the KG, related reviews and related introductions, respectively, by the standard switching mechanism (Gulcehre et al., 2016). To optimize the response generation of the dialogue module, we set the cross-entropy loss:

$$L_{gen} = -\frac{1}{N} \sum_{t=1}^{N} \log \mathrm{Pr}\left(s_t | s_1, \ldots, s_{t-1}\right) \tag{13}$$

where N is the number of rounds of a conversation C and s_t is the t-th sentence in the conversation.

4.5. Parameter Learning

The parameters of MACR mainly come from the recommendation module and the dialogue module, denoted by Θ^r and Θ^g, respectively. These two modules share some parameters and influence each other. Algorithm 1 presents the training algorithm of MACR.

Before starting the training, we first need to retrieve the introduction and genre in the information database based on the context in the dataset. Then, we optimize the parameters Θ^r and Θ^g. At each iteration, we need to obtain the set of entities from the current context, reviews, and introductions and utilize the knowledge graph DMBG to obtain the corresponding embeddings. Then, we need to perform self-attention to get entity preferences $u^{(e)}$ and category preferences $u^{(g)}$ and use the gating mechanism to fuse them into a user representation. After calculating the recommendation probability for each movie, we compute the cross-entropy loss by Eq. 8 and perform gradient descent to update the parameter Θ^r.

When the loss of the recommendation module converges, we start to optimize the parameter Θ^g. At each iteration, we still have to obtain the entity embedding through DBMG first. Then, we encode the relevant context, reviews, and introductions using Transformer's encoder and decode them with the enhanced decoder. After calculating the generation probabilities using Eq. 12, we perform gradient descent using Eq. 13 and update the parameters Θ^g.

Algorithm 1: The training algorithm of MACR.

Input: CRS dataset $\mathcal{R}$, information database I_{db}, reviews R and knowledge graph $\mathcal{G}$.
Output: Model parameters Θ^r, Θ^g.

Randomly initialize Θ^r, Θ^g.
Acquire I_{intro} and I_{genre} from $\mathcal{R}$ and I_{db} by Eq. 1.
for $i = 1 \rightarrow |\mathcal{R}|$ **do**
    Acquire a set of entities $E^{(CRI_{intro})}$ from $\mathcal{R}$, R, and I_{intro} by Eq. 3.
    Acquire entity embeddings from $\mathcal{G}$ and use Eq. 4 to obtain $u^{(e)}$ and $u^{(g)}$ by self-attention.
    Acquire $u^{(eg)}$ by gate mechanism using Eq. 5.
    Compute $\mathrm{P}_{rec}(m)$ using Eq. 7.
    Perform GD on Eq. 8 w.r.t. Θ^r.
end
for $j = 1 \rightarrow |\mathcal{R}|$ **do**
    Acquire entity embeddings from $\mathcal{G}$.
    Acquire embedding X^c, X^r, and X^i by encoder.
    Acquire Y^n by Transformer decoder using Eq. 9.
    Compute $\mathrm{Pr}\left(y_i | y_{i-1}\right\}$ using Eq. 12.
    Perform GD on Eq. 13 w.r.t. Θ^g.
end
return Θ^r, Θ^g.

5. Experiment Settings

5.1. Dataset

REDIAL is the most commonly used dataset in the CRS domain. It contains 10,006 conversations consisting of 504 users involving 51,699 movies. The conversations are conducted by Amazon Mechanical Turk (AMT) workers in the roles of seekers and recommenders according to a comprehensive set of instructions. In addition, this dataset gives three levels of preference for each user-mentioned movie: like, dislike, and don't know. Because of the small number of movies the user mentions in the conversation, existing studies will default to the user liking all the entities. We take the same operation to ensure that the number of entities is sufficient. However, when obtaining the movie genres, we only target the genres of movies that the user actually likes.

5.2. Baselines

In CRS, we need to measure the model's performance in two ways, one for the recommendation task and one for the conversation task. Therefore, we introduce multiple competitive baselines from previous studies:

- **Popularity**: It ranks items based on the frequency of recommendations from the training set in the corpus.
- **TextCNN** (Kim, 2014): It uses CNN models to extract users' features from conversation text to rank items.
- **Transformer** (Vaswani et al., 2017): It utilizes the Transformer-based encoder-decoder approach to generate dialog responses.
- **ReDial** (Li et al., 2018): This model consists of an auto-encoder (Li et al., 2016a) recommender, a dialogue module based on HRED (Serban et al., 2017) and a sentiment prediction models.
- **KBRD** (Chen et al., 2019): This model utilizes a KG enhance the entity representation in context. The transformer-based (Vaswani et al., 2017) dialogue module uses information from the KG as the word vocabulary for utterances.
- **KGSF** (Zhou et al., 2020a): This model incorporates two knowledge graphs, *i.e.,* a word-level KG and an entity-level KG, to further enhance the semantic

representations of words and entities, and uses MIM to align the semantic spaces of different KG.

- **RevCore** (Lu et al., 2021): This model uses reviews to augment the entity and word representations in context, enabling more diverse response generation. For a fair comparison, the same reviews are used in our model.
- **C²-CRS** (Zhou et al., 2022): This model extracts and represents multi-granularity semantic units from different data signals and aligns the semantic space by pre-training from coarse to fine to produce a more coherent fusion representation.

In these baselines, Popularity and TextCNN (Kim, 2014) are recommendation methods, and Transformer (Vaswani et al., 2017) is a text generation method. Since there are no historical user-item interaction records except dialogue utterances, we did not include other recommendation models. Besides, Redial (Li et al., 2018), KBRD (Chen et al., 2019), KGSF (Zhou et al., 2020a), RevCore (Lu et al., 2021), C^2-CRS (Zhou et al., 2022) and are conversational recommendation methods.

5.3. Evaluation Metrics

We use different evaluation metrics for the recommendation task and the conversation task. The evaluation metric for the recommendation task is Recall@k (k=1,10,50), which indicates whether the top-k items recommended by the model contain true labels, where the dataset provides the true labels. Conversation evaluation comprises automatic and human evaluation. The automatic evaluation includes: 1) Perplexity (PPL) (Jelinek et al., 1977): this metric is used to evaluate the fluency of the generated responses, where a smaller value indicates a more fluent sentence, and 2) Distinct-n(Dist-n, n=2,3,4) (Li et al., 2016b): this metric is used to evaluate the diversity of the generated responses, which is defined as the number of different n-grams words divided by the total number of words. For the human evaluation, annotators score the system's responses based on fluency and informativeness. The range of score is 0 to 2.

5.4. Implementation Details

The MACR model was implemented by PyTorch and trained on an NVIDIA GeForce RTX 3090 card. For a fair comparison, we kept the training parameters consistent with the baseline RevCore and did not change the review information. The maximum length of the context and reviews are set to 256 and 30, respectively, and the embedding size is 300. Similarly, we perform three epochs of pre-training to maximize the mutual information. After that, we perform the training of the recommendation module and dialogue module, respectively.

Table 1
Results on the recommendation task. The best results are indicated in bold. "*-intro*" means remove introduction information, "*-DBMG*" means use DBpedia instead of DBMG, and "*-genre*" means no fusion of categories information to user preferences.

Models	Recall@1	Recall@10	Recall@50
Popularity	0.012	0.061	0.179
TextCNN	0.013	0.068	0.191
ReDial	0.024	0.140	0.320
KBRD	0.031	0.150	0.336
KGSF	0.039	0.183	0.378
RevCore	0.046	0.220	0.396
C^2-CRS	0.053	0.233	0.407
MACR	**0.064**	**0.251**	**0.443**
-intro	0.055	0.240	0.414
-DBMG	0.061	0.246	0.438
-genre	0.056	0.236	0.442

6. Results and Analysis

6.1. Evaluation on Recommendation Task

We use Recall@k to evaluate the recommendation task, and the experimental results are shown in Table 1. First, as we have seen, conversational recommendation methods are generally better than recommendation methods (*e.g.,* TextCNN and Popularity), which is facilitated by the fact that CRS is better at obtaining information from sparse conversational contexts. Second, KBRD outperforms ReDial among conversational recommendation methods because it introduces KG to enhance entity representation. Next, KGSF outperforms KBRD due to introducing both entity-level KG and word-level KG. Based on KGSF, RevCore further enhances user preferences by integrating external reviews. Finally, C^2-CRS better integrates the KG and external reviews into the context using a contrastive learning approach.

As shown in Table 1, our MACR outperforms all baselines because we fully leverage the potential category preferences of users in conversations and introduce external introductions to enrich short conversation histories further. Compared with C^2-CRS, MACR improves R@1 scores by 20.7%, R@10 scores by 7.7%, and R@50 scores by 8.8%.

To validate the effectiveness of our proposed approach, we conducted three ablation experiments, as shown in the last three rows of Table 1. The performance of MACR decreases after removing the introduction information, DBMG, and movie genre information, respectively. Among them, movie introduction has the most critical impact on MACR, further proving that introducing more external entities can effectively solve the problem of insufficient contextual information. The results of "-DBMG" and "-genre" also show that strengthening the category association between entity embeddings through DBMG and using users' category preferences to compensate for entity preferences significantly improve for the recommendation effectiveness.

Table 2
Automatic evaluation results on conversation task. The best results are in bold. "-intro" means remove introduction information, "*-DBMG*" means use DBpedia.

Models	Dist-2	Dist-3	Dist-4	PPL
Transformer	0.148	0.151	0.137	17.0
ReDial	0.225	0.236	0.228	28.1
KBRD	0.263	0.368	0.423	17.9
KGSF	0.289	0.434	0.519	9.8
RevCore	0.424	0.558	0.612	10.2
C^2-CRS	0.413	0.619	0.767	-
MACR	**0.523**	**0.731**	**0.842**	10.3
-intro	0.403	0.613	0.770	**9.5**
-DBMG	0.478	0.710	0.841	10.4

Table 3
Human evaluation results on conversation task. The best results are in bold.

Models	Fluency	Informativeness
Transformer	0.94	0.89
ReDial	1.21	1.07
KBRD	1.26	1.19
KGSF	1.48	1.32
RevCore	1.51	1.37
C^2-CRS	1.54	1.44
MACR	**1.56**	**1.49**

6.2. Evaluation on Conversation Task

The evaluation of the conversation task was divided into automatic and human evaluation, and the results are shown in Tables 2 and 3, respectively. In the automatic evaluation, ReDial achieves better results than Transformer because it applies a pre-trained RNN to represent the conversation history. KBRD outperforms ReDial due to the utilization of KG in the dialogue module to enhance entity word generation. KGSF achieves better performance than KBRD by using cross-attention to fuse information from the aligned two KGs into the context. RevCore and C^2-CRS further improve the performance of the dialogue module by introducing reviews, and the latter uses contrastive learning to better fuse external information.

Our MACR further enhances item representation and improves the diversity of system responses in dialogue contexts through DBMG and item introduction. In terms of automatic evaluation, as shown in Table 2, MACR improved Dist-2 scores by 23.3% compared to baseline RevCore. Compared to C^2-CRS, MACR improved Dist@3 scores by 18.1% and Dist-4 scores by 9.8%, respectively. For PPL, MACR achieved scores comparable to the baseline KGSF and RevCore. In terms of human evaluation, Table 3 also

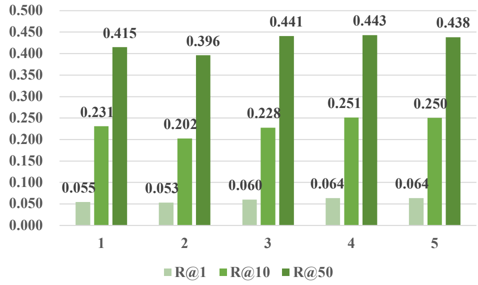


Figure 5: The effect of the number of movie genres on the recommendation task.

Table 4
The effect of introduction length on the recommendation task and the conversation task. Bold indicates the best results.

Len	Recommendation			Recommendation		
	R@1	R@10	R@50	Dist-2	Dist-3	PPL
10	0.043	0.243	0.426	0.462	0.644	10.4
20	0.061	0.234	0.442	0.511	0.710	**10.3**
30	**0.064**	**0.251**	**0.443**	0.523	**0.731**	**10.3**
40	0.061	0.220	0.426	**0.551**	0.724	10.7

demonstrates the better fluency and diversity of the responses generated by MACR.

To prove the contribution of each component in the dialogue module, we performed two ablation experiments, as shown in the last two rows of Table 2. The ablation experiment "-intro" showed that the movie introduction could significantly increase the diversity of system responses. However, there are contextual differences between movie introductions and dialogue utterances, so integrating movie introductions also harms the fluency of the response generation task. And "-DBMG" indicates that KG containing category information can also enhance the effectiveness of the dialogue task.

6.3. Discussion

Effect of the number of item categories. In our approach, a movie can have multiple genres, and different movies contain different numbers of genres. To explore the effect of the number of movie genres on MACR, we conducted a series of experiments on the maximum number of movie genres. As shown in Figure 5, we set the maximum number of extracted movie genres to 1, 2, 3, 4, and 5. MACR achieves the best results on the recommendation task when the maximum number of movie genres is 4. This result indicates that, to some extent, movies containing a more number of genres help MACR to understand users' preferences more accurately. However, too many genres (*e.g.,* when the number is 5) may introduce more noise and mislead the model's judgment, thus reducing the recommendation performance.

Effect of item introduction length. We executed a series of experiments to explore the impact of introduction sentence length by setting the maximum length of introduction sentences to 10, 20, 30, and 40. As shown in Table 4, most evaluation metrics achieved the best results at a maximum length of 30 for the introduction sentence. For this result, we believe that longer introduction sentences may contain more external entities, which helps the model somewhat predict users' preferences. However, too many external entities also introduce more noise, which reduces the model's performance. In addition, there are some differences between the external text and the original text, and excessive introduction of external text may also suppress the original text. Therefore, we set the maximum length to 30 to maintain a balance between them.

6.4. Case Study

In this section, we will use a case study to illustrate how our model works in practice, as shown in Figure 6. First, MACR extracts the movie entity "*Harry Potter*" in the context. Then, the information retrieval module obtains the genres to which the movie belongs, *i.e.,* "*Fantasy*" and "*Adventure*", and gets the embedding of the categories and entities through the knowledge graph DBMG. The recommendation module predicts movies that may interest the user, such as "*Maleficent*" and "*The BFG*", based on the user representation of the fusion of genre embedding and entity embedding. The dialogue module uses the prediction results of the recommendation module and the introduction sentences selected by the information retrieval module to generate responses. If the user does not like or has already seen the recommended movie, the system will continue interacting with the user and update the user's preferences. For example, in Figure 6, the user dislikes the recommended movie and indicates that he/she prefers "*thrilling adventure stories*". The system re-recommends movies based on the user's favorite "*Harry Potter*" in the historical conversation and the information in the new conversation.

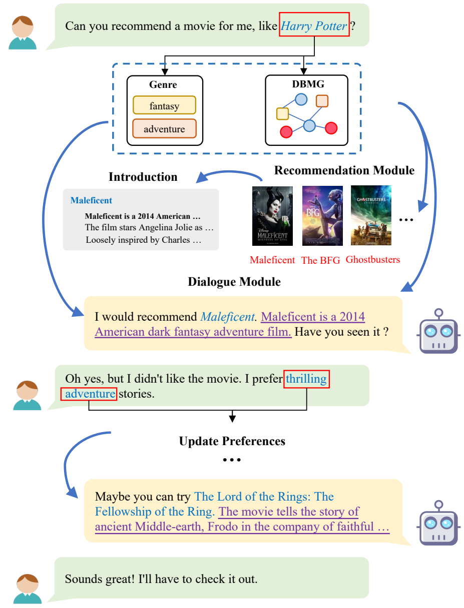


Figure 6: Case Study. Bold in the introduction box indicates the selected introduction sentence. Blue font indicates entities and purple font indicates descriptive sentences.

7. Conclusion

In this paper, we noticed potential internal information in the conversation and desirable external information, *i.e.,* category preferences and item introductions, and proposed a novel model MACR. We constructed a knowledge graph DBMG with category information to enhance category associations between entity embeddings, then extracted users' category preferences in context and fused them into the user representation. In addition, we extracted entities from relevant item introductions and fused item introductions into the context using an introduction-attentive encoder-decoder. Extensive experiments show that our approach outperforms previous state-of-the-art methods.

Our MACR proves to be effective in alleviating the shortcomings of short conversation histories through categories and introductions. However, the limitation of MACR is the lack of scalability to other external information. In the future, we will consider a more general model that can directly exploit other external data.

Acknowledgements

This work is supported in part by the National Natural Science Foundation of China [grant number 62141201], Chongqing Ba'nan District Science and Technology Bureau [grant number 2020QC403], and Postgraduate Research and Innovation Projects of Chongqing [grat number CYS22659]. We thank all the anonymous reviewers who generously contributed their time and efforts. Their professional recommendations have greatly enhanced the quality of the manuscript.

References

Alam, M., Iana, A., Grote, A., Ludwig, K., Müller, P., Paulheim, H., 2022. Towards analyzing the bias of news recommender systems using sentiment and stance detection, in: Companion Proceedings of the Web Conference 2022, Association for Computing Machinery, New York, NY, USA. p. 448–457. URL: https://doi.org/10.1145/3487553.3524674, doi:10.1145/3487553.3524674.

Chen, Q., Lin, J., Zhang, Y., Ding, M., Cen, Y., Yang, H., Tang, J., 2019. Towards knowledge-based recommender dialog system, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Association for Computational Linguistics, Hong Kong, China. pp. 1803–1813. URL: https://aclanthology.org/D19-1189, doi:10.18653/v1/D19-1189.

Chen, Y., Chen, B., Duan, X., Lou, J.G., Wang, Y., Zhu, W., Cao, Y., 2018. Learning-to-ask: Knowledge acquisition via 20 questions, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery; Data Mining, Association for Computing Machinery, New York, NY, USA. p. 1216–1225. URL: https://doi.org/10.1145/3219819.3220047, doi:10.1145/3219819.3220047.

Deng, Y., Li, Y., Sun, F., Ding, B., Lam, W., 2021. Unified conversational recommendation policy learning via graph-based reinforcement learning, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Association for Computing Machinery, New York, NY, USA. p. 1431–1441. URL: https://doi.org/10.1145/3404835.3462913, doi:10.1145/3404835.3462913.

Gomes, J., de Mello, R.C., Ströele, V., de Souza, J.F., 2022. A hereditary attentive template-based approach for complex knowledge base question answering systems. Expert Systems with Applications 205, 117725. URL: https://www.sciencedirect.com/science/article/pii/S0957417422010089, doi:https://doi.org/10.1016/j.eswa.2022.117725.

Gu, J., Lu, Z., Li, H., Li, V.O., 2016. Incorporating copying mechanism in sequence-to-sequence learning, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Berlin, Germany. pp. 1631–1640. URL: https://aclanthology.org/P16-1154, doi:10.18653/v1/P16-1154.

Gulcehre, C., Ahn, S., Nallapati, R., Zhou, B., Bengio, Y., 2016. Pointing the unknown words, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Berlin, Germany. pp. 140–149. URL: https://aclanthology.org/P16-1014, doi:10.18653/v1/P16-1014.

Hu, X., Mi, H., Wen, Z., Wang, Y., Su, Y., Zheng, J., de Melo, G., 2021. R2D2: Recursive transformer based on differentiable tree for interpretable hierarchical language modeling, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Online. pp. 4897–4908. URL: https://aclanthology.org/2021.acl-long.379, doi:10.18653/v1/2021.acl-long.379.

Jelinek, F., Mercer, R.L., Bahl, L.R., Baker, J., 1977. Perplexity—a measure of the difficulty of speech recognition tasks. Journal of the Acoustical Society of America 62.

Ji, H., Huang, M., 2021. DiscoDVT: Generating long text with discourse-aware discrete variational transformer, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Online and Punta Cana, Dominican Republic. pp. 4208–4224. URL: https://aclanthology.org/2021.emnlp-main.347, doi:10.18653/v1/2021.emnlp-main.347.

Karimi Mahabadi, R., Ruder, S., Dehghani, M., Henderson, J., 2021. Parameter-efficient multi-task fine-tuning for transformers via shared hypernetworks, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Online. pp. 565–576. URL: https://aclanthology.org/2021.acl-long.47, doi:10.18653/v1/2021.acl-long.47.

Kawai, M., Sato, H., Shiohama, T., 2022. Topic model-based recommender systems and their applications to cold-start problems. Expert Systems with Applications 202, 117129. URL: https://www.sciencedirect.com/science/article/pii/S0957417422005279, doi:https://doi.org/10.1016/j.eswa.2022.117129.

Kim, Y., 2014. Convolutional neural networks for sentence classification, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar. pp. 1746–1751. URL: https://aclanthology.org/D14-1181, doi:10.3115/v1/D14-1181.

Lei, W., He, X., Miao, Y., Wu, Q., Hong, R., Kan, M.Y., Chua, T.S., 2020. Estimation-Action-Reflection: Towards Deep Interaction Between Conversational and Recommender Systems. Association for Computing Machinery, New York, NY, USA. p. 304–312. URL: https://doi.org/10.1145/3336191.3371769.

Li, J., Galley, M., Brockett, C., Gao, J., Dolan, B., 2016a. A diversity-promoting objective function for neural conversation models, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, San Diego, California. pp. 110–119. URL: https://aclanthology.org/N16-1014, doi:10.18653/v1/N16-1014.

Li, J., Galley, M., Brockett, C., Gao, J., Dolan, B., 2016b. A diversity-promoting objective function for neural conversation models, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, San Diego, California. pp. 110–119. URL: https://aclanthology.org/N16-1014, doi:10.18653/v1/N16-1014.

Li, J., Song, H., Li, J., 2022. Transformer-based question text generation in the learning system, in: 2022 the 6th International Conference on Innovation in Artificial Intelligence (ICIAI), Association for Computing Machinery, New York, NY, USA. p. 50–56. URL: https://doi.org/10.1145/3529466.3529484, doi:10.1145/3529466.3529484.

Li, R., Kahou, S., Schulz, H., Michalski, V., Charlin, L., Pal, C., 2018. Towards deep conversational recommendations, in: Proceedings of the 32nd International Conference on Neural Information Processing Systems, Curran Associates Inc., Red Hook, NY, USA. p. 9748–9758.

Li, S., Lei, W., Wu, Q., He, X., Jiang, P., Chua, T.S., 2021. Seamlessly unifying attributes and items: Conversational recommendation for cold-start users. ACM Trans. Inf. Syst. 39. URL: https://doi.org/10.1145/3446427, doi:10.1145/3446427.

Liang, Z., Hu, H., Xu, C., Miao, J., He, Y., Chen, Y., Geng, X., Liang, F., Jiang, D., 2021. Learning neural templates for recommender dialogue system, in: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Online and Punta Cana, Dominican Republic. pp. 7821–7833. URL: https://aclanthology.org/2021.emnlp-main.617, doi:10.18653/v1/2021.emnlp-main.617.

Lu, Y., Bao, J., Song, Y., Ma, Z., Cui, S., Wu, Y., He, X., 2021. RevCore: Review-augmented conversational recommendation, in: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, Association for Computational Linguistics, Online. pp. 1161–1173. URL: https://aclanthology.org/2021.findings-acl.99, doi:10.18653/v1/2021.findings-acl.99.

Nayeri, M.R., Araabi, B.N., Yazdanpanah, M., Moshiri, B., 2022. Design, implementation and evaluation of an expert system for operating regime detection in industrial gas turbine. Expert Systems with Applications 203, 117332. URL: https://www.sciencedirect.com/science/article/pii/S0957417422006923, doi:https://doi.org/10.1016/j.eswa.2022.117332.

Ng, C.B.R., Bil, C., Sardina, S., O'bree, T., 2022. Designing an expert system to support aviation occurrence investigations. Expert Systems with Applications 207, 117994. URL: https://www.sciencedirect.com/science/article/pii/S0957417422012192, doi:https://doi.org/10.1016/j.eswa.2022.117994.

Ren, X., Yin, H., Chen, T., Wang, H., Huang, Z., Zheng, K., 2021. Learning to ask appropriate questions in conversational recommendation, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Association for Computing Machinery, New York, NY, USA. p. 808–817. URL: https://doi.org/10.1145/3404835.3462839, doi:10.1145/3404835.3462839.

Schlichtkrull, M., Kipf, T.N., Bloem, P., van den Berg, R., Titov, I., Welling, M., 2018. Modeling relational data with graph convolutional networks, in: Gangemi, A., Navigli, R., Vidal, M.E., Hitzler, P., Troncy, R., Hollink, L., Tordai, A., Alam, M. (Eds.), The Semantic Web, Springer International Publishing, Cham. pp. 593–607.

Seo, Y.D., Kim, Y.G., Lee, E., Kim, H., 2021. Group recommender system based on genre preference focusing on reducing the clustering cost. Expert Systems with Applications 183, 115396. URL: https://www.sciencedirect.com/science/article/pii/S0957417421008198, doi:https://doi.org/10.1016/j.eswa.2021.115396.

Serban, I.V., Sordoni, A., Lowe, R., Charlin, L., Pineau, J., Courville, A., Bengio, Y., 2017. A hierarchical latent variable encoder-decoder model for generating dialogues, in: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, AAAI Press. p. 3295–3301.

Shen, Q., Wen, H., Tao, W., Zhang, J., Lv, F., Chen, Z., Li, Z., 2022. Deep interest highlight network for click-through rate prediction in trigger-induced recommendation, in: Proceedings of the ACM Web Conference 2022, Association for Computing Machinery, New York, NY, USA. p. 422–430. URL: https://doi.org/10.1145/3485447.3511970, doi:10.1145/3485447.3511970.

Su, Y., Zhang, R., M. Erfani, S., Gan, J., 2021. Neural graph matching based collaborative filtering, in: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Association for Computing Machinery, New York, NY, USA. p. 849–858. URL: https://doi.org/10.1145/3404835.3462833, doi:10.1145/3404835.3462833.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, Curran Associates Inc., Red Hook, NY, USA. p. 6000–6010.

Zhang, Y., Chen, X., Ai, Q., Yang, L., Croft, W.B., 2018. Towards conversational search and recommendation: System ask, user respond, in: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Association for Computing Machinery, New York, NY, USA. p. 177–186. URL: https://doi.org/10.1145/3269206.3271776, doi:10.1145/3269206.3271776.

Zhou, K., Zhao, W.X., Bian, S., Zhou, Y., Wen, J.R., Yu, J., 2020a. Improving conversational recommender systems via knowledge graph based semantic fusion, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery; Data Mining, Association for Computing Machinery, New York, NY, USA. p. 1006–1014. URL: https://doi.org/10.1145/3394486.3403143, doi:10.1145/3394486.3403143.

Zhou, K., Zhou, Y., Zhao, W.X., Wang, X., Wen, J.R., 2020b. Towards topic-guided conversational recommender system, in: Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Online). pp. 4128–4139. URL: https://aclanthology.org/2020.coling-main.365, doi:10.18653/v1/2020.coling-main.365.

Zhou, Y., Zhou, K., Zhao, W.X., Wang, C., Jiang, P., Hu, H., 2022. C²-crs: Coarse-to-fine contrastive learning for conversational recommender system, in: Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, Association for Computing Machinery, New York, NY, USA. p. 1488–1496. URL: https://doi.org/10.1145/3488560.3498514, doi:10.1145/3488560.3498514.

Zou, J., Chen, Y., Kanoulas, E., 2020. Towards Question-Based Recommender Systems. Association for Computing Machinery, New York, NY, USA. p. 881–890. URL: https://doi.org/10.1145/3397271.3401180.

Zou, J., Kanoulas, E., 2019. Learning to ask: Question-based sequential bayesian product search, in: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Association for Computing Machinery, New York, NY, USA. p. 369–378. URL: https://doi.org/10.1145/3357384.3357967, doi:10.1145/3357384.3357967.

Highlights

- Constructing a knowledge graph DBMG with item category information
- Extracting user preferences for item categories from conversations
- Fusing category preferences and entity preferences through gating mechanisms
- Using item introductions to alleviate the lack of information in short dialogues

Credit Author Statement

Chengyang Zhang: Conceptualization, Methodology, Software, Writing - Original Draft, Writing - Review & Editing, Investigation.
Xianying Huang: Writing - Review & Editing, Supervision, Project administration.
Jiahao An: Formal analysis , Data Curation, Writing - Review & Editing, Validation, Resources.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: