Emotion-based Conversational Recommendation by Inferring Implicit Users' Preferences from their Subjective Claims

Xuanming Zhang[*], Yonghe Lu[⋆], Jianxing Yu[(✉)], Huaijie Zhu, Wei Liu, Wenqing Chen, Jian Yin

School of Artificial Intelligence, Sun Yat-sen University, Zhuhai, 519082, China
Key Laboratory of Sustainable Tourism Smart Assessment Technology, Ministry of Culture and Tourism of China, Sun Yat-sen University, Zhuhai, 519082, China
Pazhou Lab, Guangzhou, 510330, China
zhangxm236@mail2.sysu.edu.cn
{luyonghe, yujx26, zhuhuaijie, liuw259, chenwq95, issjyin}@mail.sysu.edu.cn

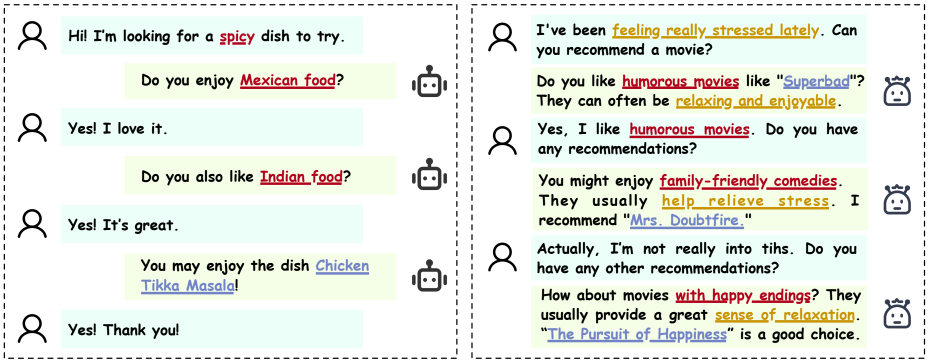
Abstract. This paper focuses on the task of emotion-based *CRSs* (conversational recommendation systems), which aim to infer users' implicit preferences from their emotional conversations without a clearly defined objective, in order to recommend better items that satisfy users' needs. Previous work mostly studies factoid-based *CRSs*, which address cases with explicit needs and capture users' preferences based on expressed entities or attributes. However, in real-world applications, most users express their needs through implicit emotions and feelings, without clearly stated entities or attributes. Existing entity-matching-based methods struggle in this scenario of vague requirements. To address this problem, we propose a novel model that is capable of inferring users' preferences from their subjective expressions. Specifically, we first apply a multifaceted augmentation technique to supplement the missing background knowledge in the conversation, including relevant subjective and objective facts and relations, in order to fully grasp users' implicit needs. Based on this enhanced knowledge, we then construct a user preference tree to capture the relationships between emotions and item attributes. The tree is updated gradually based on users' feedback in each round of conversation. Experiments on two popular datasets demonstrate the effectiveness of our approach.

Keywords: Conversational Recommendation · Implicit Preference Learning · Emotion-driven Preferences

1 Introduction

Conversational recommendation systems (*CRSs*) have become a popular research area, as they can recommend items that users may be interested in during a dialogue. For example, when users mention their preferences for certain musicians during the chat, *CRSs* can detect these needs and recommend relevant music tracks or albums based

[⋆] These authors have contributed equally to this work. Jianxing Yu is the corresponding author.



Hi! I'm looking for a spicy dish to try.

Do you enjoy Mexican food?

Yes! I love it.

Do you also like Indian food?

Yes! It's great.

You may enjoy the dish Chicken Tikka Masala!

Yes! Thank you!

I've been feeling really stressed lately. Can you recommend a movie?

Do you like humorous movies like "Superbad"? They can often be relaxing and enjoyable.

Yes, I like humorous movies. Do you have any recommendations?

You might enjoy family-friendly comedies. They usually help relieve stress. I recommend "Mrs. Doubtfire."

Actually, I'm not really into tihs. Do you have any other recommendations?

How about movies with happy endings? They usually provide a great sense of relaxation. "The Pursuit of Happiness" is a good choice.

Fig. 1: Factoid-based *CRSs* (left) vs. Emotion-based *CRSs* (right). Yellow, red, and blue indicate user queries, item attributes, and recommended items, respectively.

on those preferences. This helps meet users' needs and enhance their experience more effectively. *CRSs* can be applied in many areas, such as e-commerce platforms [12], media services [9], and medical consultations [29]. Currently, factoid-based *CRSs* have been well studied [37]. For example, in the dialog "*I am looking for a spicy dish to try*," the user's goal is clear they are looking for a dish with a spicy attribute. In this case we simply need to retrieve and recommend dishes that match this attribute to likely satisfy the user's needs. However, in real-world applications, users' needs are often vague, unclear, and subjective, with emotions and feelings frequently used to express their preferences. As illustrated in Fig.1, in the dialog "*I've been feeling stressed lately, can you recommend a movie?*" the user seeks a movie to alleviate emotional stress, but their request lacks a specific movie type or attribute. As a result, traditional *CRSs* are unable to retrieve relevant movies through a literal match on type or attribute. In other words, while the user's preference is subjective and emotional, the movie's attributes are objective and descriptive. This mismatch between the two sides leads to a failure in satisfying the user's needs. Furthermore, emotions are often implicit and vague, and users express their preferences in diverse and varying ways. Given the complexity and variability of these expressions, accurately inferring a user's preference is not a trivial task. Moreover, users' needs are typically not communicated all at once; instead, they are often revealed gradually, with only part of the need expressed at a time. These partial expressions make it challenging to fully understand their true preferences. This is more challenging than traditional factoid-based *CRSs*, which we refer to as emotion-based *CRSs*. Although this new task has been less studied, it holds significant commercial potential. For instance, in e-commerce, emotion-based *CRSs* can help users find products based on their feelings, without the need for an explicitly stated target. Thus, we focus on this area to address the existing research gap.

Inspired by the above observations, we propose a novel framework called *EBCR*. It not only infers users' preferences from their implicit emotional expressions using multi-faceted commonsense knowledge, but also analyzes the viewpoints of items to recommend those that align with the users' preferences. In detail, we first analyze the finer-grained entities and emotions of the recommended items from the related reviews.

We then use a retrieval augmentation technique to extract factual and emotional commonsense knowledge from the users' conversational content. That can supplement the missing but crucial clues in their emotional description, enabling a more accurate inference of the users' preferences. To bridge the gap between items and users' needs, we propose an *Inference Strategy Tree* (*IST*). The tree consists of multiple types of nodes organized hierarchically, with parent-child relationships representing fine-grained details. For example, a node might represent 'comedy,' with a child node labeled 'humor,' indicating a specific subgenre of comedy. Given that users' needs are often unclear, constructing the tree in its entirety at once is challenging. Therefore, we adopt an adaptive tree structure that adjusts dynamically based on user feedback and historical interactions as the conversation unfolds. The tree is built incrementally based on the conversational context, clarifying users' true needs by asking item-related questions at each step. Through this process, we can filter out irrelevant requests, infer the items users are most likely to be satisfied with, and ultimately generate appropriate responses for the dialogue. Experiments on two popular datasets demonstrate that our approach significantly outperforms existing baselines. The main contributions of this paper are as follows:

- We propose a new topic of emotion-based *CRSs*. Unlike traditional factoid-based *CRSs*, the users do not need to explicitly describe the desired entity or attribute, and they can obtain satisfactory recommendations based on subjective descriptions such as feelings.
- We propose a new method that can well identify users' implicit preferences, enabling personalized recommendations based on their subjective and emotional conversations.
- Extensive experiments are conducted to fully examine the effectiveness of our approach.

2 Related work

Recently, conversational recommendation systems (*CRSs*) have gained significant attention for their ability to provide personalized recommendations through dialogue with users. Existing *CRSs* methods can broadly be divided into two categories. The first type involves question-based user preference elicitation, following a system-asks, user-responds model. The first type involves question-based user preference elicitation, following a system-asks, user-responds model. This approach focuses on asking predefined attribute-based questions to users in order to gather more detailed information about their preferences [1,39,31], commonly referred to as clarification questions. Some research [33] introduced multiple-choice questions, allowing users to select attribute options based on their interests, especially when they are uncertain about their preferences. However, these *CRSs* often prioritize recommendations over dialogue, limiting users to predefined choices. Other studies [8] have proposed prompt-based learning strategies to enhance the multi-task learning capabilities of unified multi-goal *CRSs*. However, these strategies typically rely on predefined task segmentation, neglecting the potential interdependencies and dynamic relationships between tasks. Additionally, some studies [6] have explored the use of dynamic graphs and set-based clustering to address errors in preference modeling. However, these models often depend on predefined clustering methods, which makes it challenging to adapt to subtle changes in user preferences as the dialogue progresses.

The second category is dialogue-driven user preference elicitation, which follows a user-speaks, system-understands model. This approach highlights the importance of dialogue [30,23,32], aiming to infer user preferences through the semantic interpretation of dialogue content. Due to the limited information available from dialogue alone, external knowledge sources are often incorporated. For example, Lu et al. [19] utilizes review datasets to analyze user utterances and sentiments, Zhou et al. [38] incorporates both word-oriented and entity-oriented knowledge graphs for semantic embedding, and Li et al. [15] emphasizes the importance of similar users, assuming that retrieved similar users share characteristics or behaviors with the target user. Recent studies [21] have highlighted the significance of subjective attributes in recommendation systems, with Long et al. [18] addressing this by constructing datasets that include such attributes. However, these approaches are not scalable and require significant manual effort, making them inadequate for handling the dynamic and complex nature of subjective conversational scenarios. To overcome these challenges, we propose a model that combines generative knowledge supplementation with multifaceted knowledge enhancements. Our approach embeds preference information within a multi-type node tree structure for inference, seamlessly integrating it into the generated responses.

3 Methods

This section introduces the proposed *EBCR* framework, illustrated in Fig. 2. The framework consists of three modules: the contextual information enhancement module, which integrates diverse external knowledge; the preference strategy inference module, which refines user preferences using a multi-type entity tree structure; and the subjective response generation module, which generates responses with empathetic and emotionally rich expressions.

3.1 Contextual Information Enhancement

Subjective Information Enhancement. To address the lack of background information in users' emotion-driven expressions, leveraging extensive external knowledge is essential for enriching the dialogue and enhancing semantic representation. The process begins with the application of the *COMET* [3] model to generate commonsense knowledge. Built upon the *ATOMIC* [24] framework, *COMET* produces triples that encapsulate various facets of commonsense understanding. For a given dialogue history $H = \{D_1, D_2, \ldots, D_n\}$, where n denotes the number of dialogue turns, *COMET* generates corresponding commonsense knowledge C_i for each turn as $C_i = \mathit{COMET}(D_i)$.

An enriched dialogue representation E_i is generated by merging the dialogue content D_i with the corresponding commonsense knowledge C_i. To further process this enhanced representation, we utilize a multi-layer *Transformer* [28] architecture for embedding. The current embedding $n^l(E)$ is computed from the output of the previous

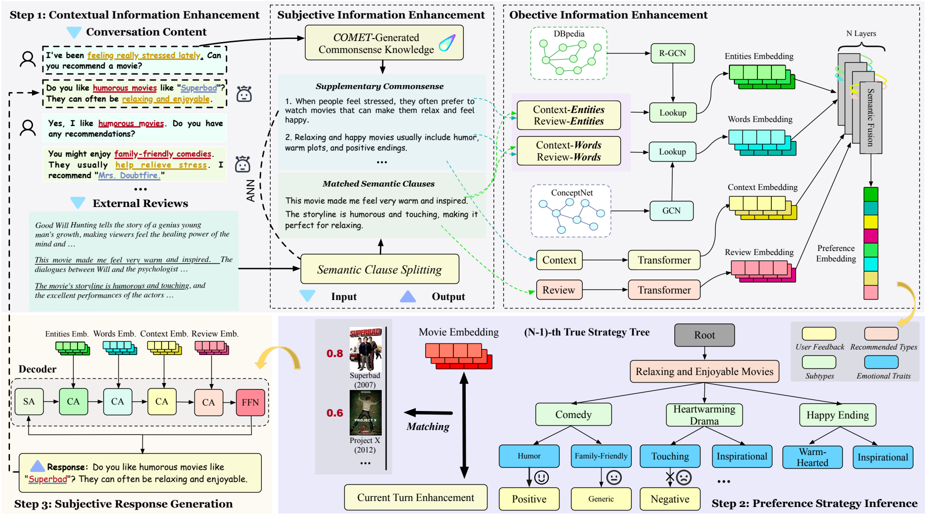

Step 1: Contextual Information Enhancement
Conversation Content
I've been feeling really stressed lately. Can you recommend a movie?
Do you like humorous movies like "Superbad"? They can often be relaxing and enjoyable.
Yes, I like humorous movies. Do you have any recommendations?
You might enjoy family-friendly comedies. They usually help relieve stress. I recommend "Mrs. Doubtfire."
...
External Reviews
Good Will Hunting tells the story of a genius young man's growth, making viewers feel the healing power of the mind and …
This movie made me feel very warm and inspired. The dialogues between Will and the psychologist …
The movie's storyline is humorous and touching, and the excellent performances of the actors …
ANN
Subjective Information Enhancement
COMET-Generated Commonsense Knowledge
Supplementary Commonsense
1. When people feel stressed, they often prefer to watch movies that can make them relax and feel happy.
2. Relaxing and happy movies usually include humor, warm plots, and positive endings.
...
Matched Semantic Clauses
This movie made me feel very warm and inspired.
The storyline is humorous and touching, making it perfect for relaxing.
Semantic Clause Splitting
Input
Output
Obective Information Enhancement
DBpedia
R-GCN
Context-Entities Review-Entities
Lookup
Context-Words Review-Words
Lookup
ConceptNet
GCN
Context
Transformer
Review
Transformer
Entities Embedding
Words Embedding
Context Embedding
Review Embedding
N Layers
Semantic Fusion
Preference Embedding
Entities Emb.
Words Emb.
Context Emb.
Review Emb.
Decoder
SA
CA
CA
CA
CA
FFN
Response: Do you like humorous movies like "Superbad"? They can often be relaxing and enjoyable.
Step 3: Subjective Response Generation
0.8
Superbad (2007)
0.6
Project X (2012)
...
Movie Embedding
Matching
Current Turn Enhancement
(N-1)-th True Strategy Tree
Root
Relaxing and Enjoyable Movies
Comedy
Heartwarming Drama
Happy Ending
Humor
Family-Friendly
Touching
Inspirational
Warm-Hearted
Inspirational
Positive
Generic
Negative
...
User Feedback
Recommended Types
Subtypes
Emotional Traits
Step 2: Preference Strategy Inference

Fig. 2: The overall framework of our method *EBCR*

layer $n^{l-1}(E)$ using a multi-head attention mechanism, enabling the model to attend to various aspects of the input. The process is defined as follows:

$$n^l(E) = \text{MHA}(n^{l-1}(E), n^{l-1}(E), n^{l-1}(E)) \tag{1}$$

For clarity, we define the output of the final transformer layer as the information-enhanced context embedding $\mathbb{F}_c$, where L denotes the number of transformer layers.

$$\mathbb{F}_c = \text{MHA}(n^{L-1}(E), n^{L-1}(E), n^{L-1}(E)) \tag{2}$$

Additionally, external reviews serve as a vital source of authentic information and often contain implicit commonsense knowledge. To utilize this, we use review datasets as supplementary knowledge. Specifically, high-quality reviews are collected from the IMDb[1] platform to construct the external review dataset. For detailed analysis, longer reviews are segmented into individual sentences, which act as retrieval units $R_i = \{S_1, S_2, \ldots, S_t\}$, where t represents the number of segmented sentences. Each sentence S_j is semantically encoded using the *Sentence-BERT* model [22], producing corresponding embedding vectors $V_j = \textit{S-BERT}(S_j)$. These embeddings are then aligned with the enhanced dialogue representation E, and the top n most similar sentences, denoted as $\hat{R}$, are retrieved using *Approximate Nearest Neighbors (ANN)* as follows:

$$\hat{R} = \text{argmax}_{S_j \in R_i} \frac{E \cdot V_j}{\|E\| \|V_j\|} \tag{3}$$

Similarly, the output of the final transformer layer is used as the embedding for external reviews, represented as:

$$\mathbb{F}_r = \text{MHA}(n^{L-1}(\hat{R}), n^{L-1}(\hat{R}), n^{L-1}(\hat{R})) \tag{4}$$

[1] https://www.imdb.com/

Objective Information Enhancement. To capture fine-grained semantic and contextual nuances, both entities and words are processed simultaneously, while external knowledge graphs provide a structured understanding of entity relationships and attributes in the dialogue. *DBpedia* [2] is used to enrich entity-related information. In a given context, entities are extracted and semantically encoded using a *Relational Graph Convolutional Network (R-GCN)* [26], generating corresponding entity embeddings. The final layer representation e^L serves as the consolidated entity embedding e, capturing all contextual and relational information from the preceding layers. Thus, the entity embeddings are expressed as:

$$\mathbb{F}_e = \text{RGCN}(\{e_i\}_{i=1}^{k}) = \{e_1^L, e_2^L, \ldots, e_k^L\} \tag{5}$$

To enrich word-related information, *ConceptNet* [27] is used, capturing various relationships such as synonyms, antonyms, and more. The semantic information is encoded with *Graph Convolutional Network (GCN)* [11], where the representation of the word w at layer $l+1$, denoted as w^{l+1}, is given by:

$$w^{l+1} = \sigma(\hat{A} w^l M^l) \tag{6}$$

Here, $\hat{A}$ corresponds to the normalized adjacency matrix, which captures the semantic relationships between words, and M^l is the learned weight matrix. The final output w^L from the last layer of the *GCN* serves as the word embedding, encapsulating the relevant contextual and relational information. Consequently, the word embeddings are expressed as follows:

$$\mathbb{F}_w = \{w_1^L, w_2^L, \ldots, w_m^L\} \tag{7}$$

Multi-dimensional Embedding Fusion. Following the enhancement of both subjective and objective information, the dialogue content is further enriched through a multifaceted approach to more effectively capture the nuances of users' emotion-driven expressions. Given that preference information across different dimensions is represented in distinct embedding spaces, a mutual information maximization strategy is employed to integrate these embeddings effectively, as shown below:

$$\begin{aligned}&\text{MMI-Fusion}(\cdot) = \\ &\begin{cases} I(T_{\text{pref}}^L) = \max \mathbb{E}_{p(\mathbb{F}_e, \mathbb{F}_w)} \left[\log \frac{p(T_{\text{pref}}^L | \mathbb{F}_e, \mathbb{F}_w)}{p(T_{\text{pref}}^L)} \right], \\ I(T_{\text{core}}^L) = \max \mathbb{E}_{p(\mathbb{F}_c, \mathbb{F}_r)} \left[\log \frac{p(T_{\text{core}}^L | \mathbb{F}_c, \mathbb{F}_r)}{p(T_{\text{core}}^L)} \right], \\ T_{\text{multi}}^L = T_{\text{pref}}^L \oplus T_{\text{core}}^L, \\ P_{\text{multi}} = \tanh(\text{LayerNorm}(T_{\text{multi}}^L)). \end{cases}\end{aligned} \tag{8}$$

Here, I represents mutual information, and $\oplus$ denotes vector concatenation. The mutual information I is calculated by maximizing the expected value. T_{pref}^L represents the embedding of background information; T_{core}^L denotes the embedding of core information; T_{multi}^L is the embedding after multi-dimensional fusion; and P_{multi} represents the final multi-dimensional fusion result after normalization.

3.2 Preference Strategy Inference

To refine user preference information, we constructed a multi-type node preference tree, termed the *Inference Strategy Tree (IST)*, designed around a hierarchical framework of dialogue nodes. In this structure, user preferences are systematically parsed and refined. The nodes of the *IST* are categorized into several distinct types: recommendation-type nodes, which organize recommendations based on user-specific needs; subtype nodes, which further categorize recommendation types; emotional trait nodes, which explore the emotional characteristics of items within each subtype; and user feedback nodes, which capture and integrate user feedback, both positive and negative, across dialogue turns to fine-tune the recommendation strategy. This multi-type entity structure aids in preserving historical preferences and aligning relationships between user emotions and item attributes.

The *IST* is initialized with a pseudo-node at its root, which does not correspond to any specific entity in the dialogue but serves as a structural placeholder. The structure of the *IST* and its associated recommendation paths are dynamically updated through an adaptive mechanism, enabling the system to progressively refine recommendations based on user feedback throughout the conversation. For instance, when the user expresses preferences or aversions toward certain elements, the system adjusts the corresponding weights or prunes them accordingly. Additionally, the *IST* employs an adaptive path optimization algorithm to dynamically recalibrate the weights of recommendation paths, calculating the attention weight for each path and selecting the one with the highest weight for further reasoning.

$$\mathbf{p} = \text{Attn}(\lambda\hat{\mathbf{P}} + (1-\lambda)\mathbf{P}_{\text{multi}}) \tag{9}$$

$$\lambda = \frac{1}{1 + e^{-\left(\mathbf{M}_s \cdot \left[\hat{\mathbf{P}}, \mathbf{P}_{\text{multi}}\right]\right)}} \tag{10}$$

The term $\mathbf{M}_s$ represents a trained parameter, while $\hat{\mathbf{P}}$ denotes the comprehensive representation of the reasoning branch. Furthermore, the structure of the *IST* adapts as the dialogue progresses. After each dialogue round, the system reassesses the effectiveness of each path, recalculating the comprehensive representation $\hat{\mathbf{P}}$ and its corresponding attention weight. This re-evaluation process identifies and prioritizes the paths most likely to align with user needs, ensuring they are emphasized in subsequent reasoning. The specific formulas are as follows:

$$\hat{\mathbf{P}} = \text{Attn}(\mathbf{P}) = \mathbf{P} \cdot \frac{e^{(\mathbf{M}_c \cdot \tanh(\mathbf{M}_r \cdot \mathbf{P}))_i}}{\sum_j e^{(\mathbf{M}_c \cdot \tanh(\mathbf{M}_r \cdot \mathbf{P}))_j}} \tag{11}$$

Where $\mathbf{M}_r$ and $\mathbf{M}_c$ are trainable parameters, interacting with the branch embedding matrix $\mathbf{P}$ to effectively capture and represent relationships between reasoning branches.

3.3 Subjective Response Generation

To effectively complete the dialogue task and generate responses that are highly relevant to the anticipated questions and entities, the decoder network is provided with a set of feature embeddings, including entity embeddings $\mathbb{F}_e$, word embeddings $\mathbb{F}_w$, contextual embeddings $\mathbb{F}_c$, and review embeddings $\mathbb{F}_r$. The decoder's output is subsequently processed through a fully connected feedforward network (*FFN*), enabling the model to produce responses that are both contextually relevant and semantically precise.

$$\Theta^l = \text{Decoder}(\Theta^{l-1}, \mathbb{F}_e, \mathbb{F}_w, \mathbb{F}_c, \mathbb{F}_r, T, H) \tag{12}$$

$$Y_i = \text{FFN}(\text{ReLU}([\Theta^L; \chi]M_1 + b_1)M_2 + b_2) \tag{13}$$

Here, Decoder$(\cdot)$ represents the decoder network, and Θ^L denotes the response generated by the final layer of the decoder. M_1 and M_2 are trainable weight matrices, with b_1 and b_2 as their corresponding bias terms, used to infer the embedding matrix T and the historical sentence embedding matrix H. Additionally, cross-entropy loss is employed to optimize the response generation process. The dialogue loss $\mathcal{L}_c$ is formally defined as:

$$\mathcal{L}_c = -\frac{1}{N}\sum_{t=1}^{N} \log Q(y_t \mid \{y_{1:t-1}\}) \tag{14}$$

where N is the number of dialogue turns, y_t represents the t-th utterance in the dialogue, and $\{y_{1:t-1}\}$ represents the previously generated subsequence. The probability of generating the next token y_t, denoted as $Q(y_t \mid \{y_{1:t-1}\})$, is defined as:

$$\begin{aligned} Q(y_t \mid \{y_{1:t-1}\}) = {} & Q_w(y_t \mid Y) + Q_k(y_t \mid Y, K) \\ & + Q_c(y_t \mid Y, C) \end{aligned} \tag{15}$$

where $Q_w(\cdot)$, $Q_k(\cdot)$, and $Q_c(\cdot)$ represent the probability functions for words, entities, and context, respectively, with Y serving as the input, K denoting the enhanced set of entities, and C representing the enriched contextual information.

4 Experiment

In this section, we assess the performance of *EBCR* through a comprehensive series of experiments, encompassing both quantitative and qualitative analyses. Specifically, we focus on addressing the following five key questions:
Q1: Does *EBCR* outperform the baselines in terms of the recommendation part?
Q2: Does *EBCR* outperform the baselines in terms of the dialogue part?
Q3: How does *EBCR* demonstrate its advantages in emotion-driven scenarios?
Q4: How does each component of the model contribute to the overall system performance?
Q5: How do parameter changes impact the performance of *EBCR*?

4.1 Experimental Configuration

Dataset. We conducted experiments on two datasets specifically curated for emotion-based conversational scenarios. The first dataset is based on the *ReDial* [14] dataset, which was collected via *Amazon Mechanical Turk* and focuses on movie recommendation dialogues. This dataset consists of 10,006 conversations, each featuring over

Scenarios	**ReDial-S**						**SupQA-S**					
	R@1	**R@10**	**R@50**	**Dist-2**	**Dist-3**	**PPL**	**R@1**	**R@10**	**R@50**	**Dist-2**	**Dist-3**	**PPL**
CRFR	0.024	0.103	0.210	0.179	0.412	11.7	0.006	0.035	0.087	0.497	1.092	9.4
CR-Walker	0.022	0.095	0.186	0.152	0.336	14.9	0.004	0.032	0.091	0.421	0.997	11.2
KBRD	0.017	0.076	0.148	0.130	0.262	23.1	-	0.027	0.073	0.192	0.651	20.6
KGSF	0.021	0.092	0.182	0.138	0.275	19.3	0.001	0.026	0.071	0.267	0.836	16.7
ReDial	0.016	0.073	0.143	0.127	0.253	30.2	-	-	0.010	0.089	0.490	24.3
TREA	0.034	0.132	0.261	0.201	0.458	11.9	0.007	0.037	0.097	0.553	1.183	8.9
GPT-3	0.029	0.112	0.246	0.183	0.435	12.1	0.005	0.033	0.083	0.472	0.992	10.0
UniCRS	0.025	0.107	0.217	0.179	0.417	11.9	0.006	0.038	0.099	0.527	1.102	9.1
LOT-CRS	0.028	0.109	0.236	0.181	0.422	12.3	0.003	0.029	0.078	0.463	0.962	13.0
EBCR	**0.039***	**0.154***	**0.306***	**0.237***	**0.474***	**9.6***	**0.012***	**0.040***	**0.107***	**0.570***	**1.253***	**8.1***

Table 1: Results of the automatic evaluation. An asterisk (*) denotes the best performance with statistically significant improvement (t-test, p-value < 0.05).

four distinct movies, totaling 64,362 unique movie entities. The second dataset is derived from the *SupQA* [34] dataset, which includes 48,352 samples across 15 different product domains. Using a semi-automated annotation process, we generated 10,000 shopping-themed dialogues, with each conversation following a specific thematic path.

To adapt these datasets for subjective and emotion-driven contexts, we employed the *GPT-4* [25] generative model to augment 70% of the existing dialogues by creating subjective variations, while preserving the original entities. Emotional features were incorporated using the *RoBERTa* [17] model to further enhance the subjectivity of the dialogues. We then conducted a manual evaluation to ensure the quality and consistency of the modified dialogues. The resulting datasets were named **ReDial-S** and **SupQA-S**.

Baselines. We employed a range of state-of-the-art benchmark methods widely recognized in the field, including multi-hop reasoning, multi-path reasoning, knowledge graph-based approaches, and the classic inference tree-based techniques. These include *CRFR* [36], *CR-Walker* [20], *KBRD* [5], *KGSF* [38], *ReDial* [14], *TREA* [16], *GPT-3* [4], *UniCRS* [7], and *LOT-CRS* [35].

Metrics. We evaluated our model on both recommendation and dialogue tasks. For the recommendation task, performance was measured using *Recall* at various levels (*R@N*, where $N = 1, 10, 50$), which assesses whether the top-N recommended items include the ground truth. In the dialogue task, both automated and manual evaluations were conducted. Automated metrics included *Dist n-gram* (n = 2, 3) [13] and *perplexity (PPL)* [10] to assess language fluency and diversity.

For manual evaluation, eight annotators were invited to assess response quality in terms of *Coherence*, *Relevance*, and *Informativeness*. These annotators were unaware of the specific details of the experiment. Prior to the evaluation, qualification tests were conducted to ensure that the annotators had a strong command of English and good linguistic evaluation skills. All data samples were annotated collectively by the evaluators, who were provided with comprehensive evaluation guidelines, including assessment criteria, scales, and examples. To ensure the validity of the evaluation results, we calculated inter-annotator agreement using *Randolph's kappa*. The consistency scores were as follows: *Coherence*: 0.82, *Relevance*: 0.77, and *Informativeness*: 0.81, indicat-

Scenarios	**ReDial-S(10%)**						**SupQA-S(10%)**					
	R@1	**R@10**	**R@50**	**Dist-2**	**Dist-3**	**PPL**	**R@1**	**R@10**	**R@50**	**Dist-2**	**Dist-3**	**PPL**
CRFR	0.035	0.112	0.226	0.194	0.437	10.2	0.013	0.052	0.109	0.523	1.132	8.5
CR-Walker	0.027	0.108	0.199	0.172	0.361	12.2	0.010	0.046	0.103	0.445	1.104	10.7
KBRD	0.021	0.083	0.159	0.142	0.293	16.7	0.005	0.032	0.081	0.217	0.682	15.3
KGSF	0.025	0.103	0.194	0.152	0.290	15.2	0.006	0.036	0.087	0.290	0.861	14.7
ReDial	0.021	0.079	0.151	0.142	0.237	23.2	0.002	0.019	0.026	0.102	0.517	19.5
TREA	0.040	0.142	0.285	0.221	0.472	10.3	0.014	0.048	0.110	0.571	1.192	8.4
GPT-3	0.032	0.121	0.259	0.192	0.451	10.9	0.009	0.042	0.107	0.493	1.105	9.1
UniCRS	0.030	0.115	0.230	0.195	0.431	10.7	0.012	0.047	0.109	0.544	1.129	8.6
LOT-CRS	0.033	0.114	0.251	0.196	0.443	10.6	0.007	0.037	0.091	0.481	1.103	11.6
EBCR	**0.044***	**0.159***	**0.315***	**0.243***	**0.499***	**9.1***	**0.019***	**0.055***	**0.116***	**0.593***	**1.259***	**7.7***

Table 2: Automatic evaluation results on the 10% emotion-enhanced dataset. An asterisk (*) denotes the best performance with statistically significant improvement (t-test, p-value < 0.05).

ing high reliability in the evaluations.

Experimental Settings. For the baseline models, we adhered to the original parameters as specified in their respective publications. To ensure the rigor of our experimental results, we conducted five runs and averaged the outcomes. Our experiments were implemented using *PyTorch* on four Nvidia RTX 3090 GPU servers. When configuring *EBCR*, we set the embedding dimensions to 256 for the recommendation task and 128 for the conversation task. The embedding layers for both the *GCN* and *R-GCN* models were configured with a depth of 1. The token length for the selected review sentences was set to 20 and initialized using *Word2Vec*[2], with a normalization factor of 1. All other parameters remained at their default settings and were fine-tuned as necessary during the experiments. For training, the batch size was set to 64, and the learning rate was adjusted to 1e-3. In the comparative experiments, we strictly adhered to the single-variable principle and conducted multiple runs to mitigate the effect of random chance.

4.2 Assessment of Recommendation Task (Q1)

As shown in Table 1, the recommendation task was evaluated using *R@N (N=1, 10, 50)*. The results demonstrate that our proposed *EBCR* consistently outperforms all baseline models across the datasets, highlighting its superior effectiveness in conversational recommendation systems for subjective scenarios. The findings also indicate that incorporating extensive external knowledge significantly boosts performance. For instance, models like *KGSF* and *CRFR*, which integrate additional knowledge graphs, achieve better results as the enriched background information compensates for the gaps caused by the missing details in emotion-driven expressions. This underscores the critical role of external knowledge in addressing the complexities of subjective dialogue scenarios. Furthermore, models like *TREA*, which utilize tree structures for entity-based reasoning, show superior performance compared to others. However, *EBCR*, with its nonlinear architecture and multi-type node reasoning, achieves the best overall results.

[2] https://radimrehurek.com/gensim/models/word2vec.html

Methods	Coherence	Relevance	Informat
CRFR	1.65	1.82	2.08
CR-Walker	1.44	1.73	1.93
KBRD	1.25	1.55	1.54
KGSF	1.42	1.69	1.82
ReDial	1.37	1.37	1.39
TREA	1.72	1.92	1.96
GPT-3	1.73	1.81	1.99
UniCRS	1.52	1.74	1.91
LOT-CRS	1.74	1.83	1.92
EBCR	**1.93***	**2.08***	**2.19***

Table 3: Results from human evaluation. An asterisk (*) denotes the best performance with statistically significant improvement (t-test, p-value < 0.05).

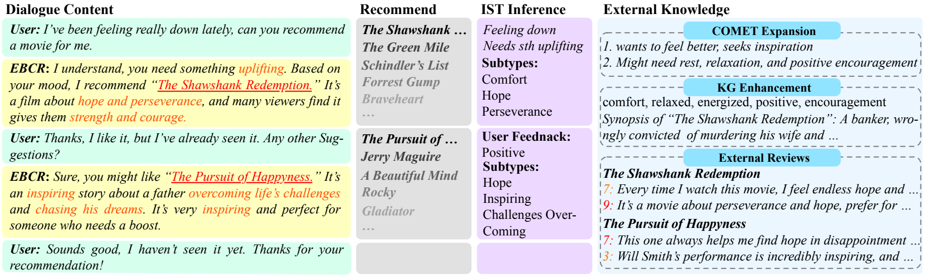
In our experiments, we utilized the *GPT-4* generative model to enhance 70% of the existing dialogues by generating subjective variations, while preserving the original movie entities. To further assess the broad applicability and adaptability of the proposed model, we conducted additional experiments on datasets where 10% of the dialogues were enhanced with emotional expressions. These tests provided valuable insights into the model's performance across diverse scenarios. The experimental results, as shown in Tables 2, demonstrate that our proposed *EBCR* model consistently delivers strong performance. Notably, even with only 10% emotional enhancement, the model effectively captured subtle emotional nuances and generated highly relevant recommendations. As the proportion of subjective dialogue increased, the model maintained exceptional performance, adeptly handling the added emotional complexity.

4.3 Assessment of Dialogue Task (Q2)

The automatic evaluation results in Table 1 demonstrate that *EBCR* consistently outperforms baseline methods across *Dist-2*, *Dist-3*, and *PPL* metrics. Similarly, the human evaluation results in Table 3 indicate that our approach surpasses all baselines in coherence, relevance, and informativeness. These findings suggest that *EBCR* generates dialogue content that is more closely aligned with user preferences and needs. By effectively integrating diverse external knowledge, *EBCR* acquires rich contextual background information and enhances its understanding of users' subjective intentions. Combined with preference path reasoning, this ensures better alignment with the generation module, resulting in more accurate and fluent responses.

4.4 Case Study (Q3)

In this section, we evaluate our method's ability to handle emotion-driven dialogue scenarios by visualizing an example from our experiments. As shown in Fig. 3, when the user expresses '*feeling down*,' *EBCR* leverages external knowledge to supplement the missing background information. *COMET* generates commonsense insights, suggesting that such individuals may need rest, encouragement, or inspiration. The knowledge graph identifies related concepts such as comfort and energy. Additionally, external



Dialogue Content	Recommend	IST Inference	External Knowledge
User: *I've been feeling really down lately, can you recommend a movie for me.*	***The Shawshank …*** ***The Green Mile*** ***Schindler's List*** ***Forrest Gump*** ***Braveheart*** …	*Feeling down* *Needs sth uplifting* **Subtypes:** Comfort Hope Perseverance	**COMET Expansion** *1. wants to feel better, seeks inspiration* *2. Might need rest, relaxation, and positive encouragement* **KG Enhancement** comfort, relaxed, energized, positive, encouragement *Synopsis of "The Shawshank Redemption": A banker, wrongly convicted of murdering his wife and …* **External Reviews** ***The Shawshank Redemption*** *7: Every time I watch this movie, I feel endless hope and …* *9: It's a movie about perseverance and hope, prefer for …* ***The Pursuit of Happyness*** *7: This one always helps me find hope in disappointment …* *3: Will Smith's performance is incredibly inspiring, and …*
EBCR: *I understand, you need something uplifting. Based on your mood, I recommend "The Shawshank Redemption." It's a film about hope and perseverance, and many viewers find it gives them strength and courage.*			
User: *Thanks, I like it, but I've already seen it. Any other Suggestions?*	***The Pursuit of …*** ***Jerry Maguire*** ***A Beautiful Mind*** ***Rocky*** ***Gladiator*** …	**User Feednack:** Positive **Subtypes:** Hope Inspiring Challenges Over-Coming	
EBCR: *Sure, you might like "The Pursuit of Happyness." It's an inspiring story about a father overcoming life's challenges and chasing his dreams. It's very inspiring and perfect for someone who needs a boost.*			
User: *Sounds good, I haven't seen it yet. Thanks for your recommendation!*			

Fig. 3: Case study results: orange highlights descriptive info, underlined text indicates recommended items, and shades in *Recommend* show matching probability.

Scenarios	**ReDial-S**		**SupQA-S**	
	R@10	**R@50**	**R@10**	**R@50**
EBCR	**0.154**	**0.306**	**0.040**	**0.107**
w/o *IST*	0.126	0.255	0.031	0.082
w/o $\mathbb{F}_c$	0.136	0.287	0.037	0.092
w/o $\mathbb{F}_r$	0.138	0.293	0.032	0.085
w/o $\mathbb{F}_e$	0.132	0.288	0.035	0.090
w/o $\mathbb{F}_w$	0.135	0.275	0.033	0.087

Fig. 4: Ablation study results (t-test, p-value < 0.05).

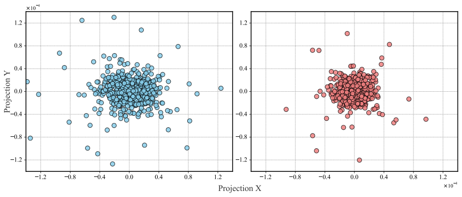


Fig. 5: Scatter plot of *KG* entity embeddings. The plot on the left displays results from *EBCR*, while the plot on the right shows results from w/o *IST* (*PPL* set to 15, 20 iterations).

reviews containing relevant keywords further enrich the emotional context of the conversation. As the dialogue evolves and the user approves a recommendation, the *IST* dynamically adjusts the weight of the current inference path and updates it iteratively. Ultimately, *EBCR* synthesizes the gathered information to generate its responses.

4.5 Ablation Study (Q4)

To evaluate the contribution of each component, we adhered to the single-variable control principle and conducted an ablation study by sequentially removing individual modules from *EBCR*. These included: 1) w/o *IST*: removing the Inference Strategy Tree, 2) w/o $\mathbb{F}_c$: removing the context embedding, 3) w/o $\mathbb{F}_r$: removing the review embedding, 4) w/o $\mathbb{F}_e$: removing the entity embedding, and 5) w/o $\mathbb{F}_w$: removing the word embedding. As shown in Table 4, removing the *Inference Strategy Tree* caused the most significant performance drop, highlighting its critical role in preference inference. Furthermore, performance noticeably declined with the removal of the other four modules, as comprehensive knowledge effectively fills gaps in emotion-based user expressions, demonstrating the overall effectiveness of all modules in *EBCR*.

To further assess *IST* performance, we visualized the entity representations in the knowledge graph for both *EBCR* and its variant without the *IST* module, as shown in

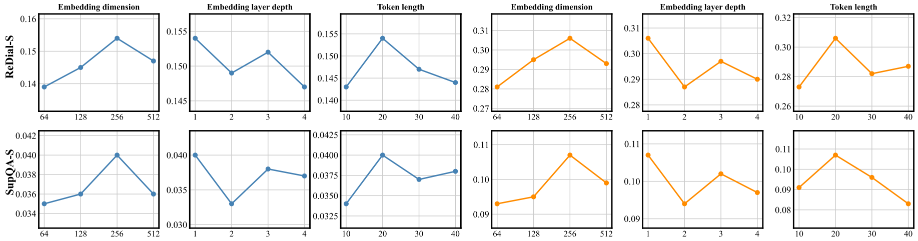


Fig. 6: Results of the Hyperparameter Study, following the single-variable principle (t-test, p-value < 0.05). The orange lines indicate R@50, and the blue lines indicate R@10.

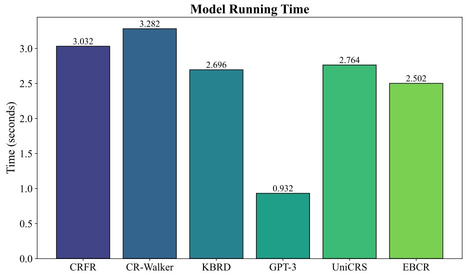


Fig. 7: Results of the time complexity study (Top-3 Turns).

Fig. 5. The visualization shows that the entity embeddings in the w/o *IST* model are more densely clustered. This occurs because, without the guidance of *IST*, all reasoning branches contribute equally to the prediction task, increasing the similarity among entity representations.

4.6 Hyperparameter Study (Q5)

In this section, we analyze the model's performance under different parameters. As shown in Fig. 6, the model is generally insensitive to parameter variations, indicating that *EBCR* has wide applicability. Moreover, as the embedding dimension increases, the model's performance improves, as larger dimensions capture more advanced feature information. The best performance is observed at a dimension of 256, since overly large dimensions may lead to overfitting. Additionally, the model can be fine-tuned by adjusting the embedding layer depth and the token length of the review clauses.

4.7 Time Complexity Study

To evaluate the time complexity of the algorithms, we trained each model for 5 epochs and measured the average response time for the Top 3 recommendations under the same input conditions. The results are shown in the Fig. 7. Overall, *CRFR* and *CR-Walker* had the longest response times, at 3.032 seconds and 3.282 seconds, respectively. In comparison, our model, *EBCR*, had a significantly lower response time of 2.502 seconds, showing a clear improvement. *GPT-3* had the fastest response time, at only 0.932 seconds, as it is a highly optimized generative model specifically designed for rapid natural language generation. In contrast, *EBCR*, despite its higher complexity due to the multi-level inference tree structure and the integration of external knowledge, still maintained a relatively low computational overhead. This demonstrates that our model achieves strong recommendation accuracy while maintaining favorable time complexity.

5 Conclusion

This paper introduced emotion-based *CRSs* as a new topic and presented a novel model, *EBCR*, to address the associated challenges. We first analyzed fine-grained entities and sentiments of recommended items, extracted from both external reviews and dialogue content. Next, we bridged the gap between items and user needs by integrating generative commonsense knowledge and emotional insights through retrieval-augmentation techniques. Leveraging enhanced external knowledge, we developed an *Inference Strategy Tree* (*IST*) to capture potential relationships between user emotions and item attributes. It composed of multiple node types, dynamically updated based on user feedback, refining user preferences and aligning recommendations with their actual needs. Extensive experiments demonstrated that our model significantly outperformed baselines in subjective and emotion-driven scenarios.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (62276279, 62306344, U2001211, U22B2060), Guangdong Basic and Applied Basic Research Foundation (2024B1515020032, 2024A1515010253), and Research Foundation of Science and Technology Plan Project of Guangzhou City (2023B01J0001, 2024B01W0004).

References

1. Bernard, N., Balog, K.: Mg-shopdial: A multi-goal conversational dataset for e-commerce. In: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 2775–2785 (2023)
2. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S.: Dbpedia-a crystallization point for the web of data. Journal of web semantics **7**(3), 154–165 (2009)
3. Bosselut, A., Rashkin, H., Sap, M., Malaviya, C., Celikyilmaz, A., Choi, Y.: Comet: Commonsense transformers for automatic knowledge graph construction. arXiv preprint arXiv:1906.05317 (2019)
4. Brown, T.B.: Language models are few-shot learners. arXiv preprint arXiv:2005.14165 (2020)
5. Chen, Q., Lin, J., Zhang, Y., Ding, M., Cen, Y., Yang, H., Tang, J.: Towards knowledge-based recommender dialog system. arXiv preprint arXiv:1908.05391 (2019)
6. Dai, X., Wang, Z., Xie, J., Liu, X., Lui, J.C.: Conversational recommendation with online learning and clustering on misspecified users. IEEE Transactions on Knowledge and Data Engineering (2024)

7. Deng, Y., Li, Y., Sun, F., Ding, B., Lam, W.: 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. pp. 1431–1441 (2021)
8. Deng, Y., Zhang, W., Xu, W., Lei, W., Chua, T.S., Lam, W.: A unified multi-task learning framework for multi-goal conversational recommender systems. ACM Transactions on Information Systems **41**(3), 1–25 (2023)
9. Dong, Z., Liu, X., Chen, B., Polak, P., Zhang, P.: Musechat: A conversational music recommendation system for videos. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12775–12785 (2024)
10. Jelinek, F., Mercer, R.L., Bahl, L.R., Baker, J.K.: Perplexity—a measure of the difficulty of speech recognition tasks. The Journal of the Acoustical Society of America **62**(S1), S63–S63 (1977)
11. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
12. Kuzi, S., Malmasi, S.: Bridging the gap between information seeking and product search systems: Q&a recommendation for e-commerce. arXiv preprint arXiv:2407.09653 (2024)
13. Li, J., Galley, M., Brockett, C., Gao, J., Dolan, B.: A diversity-promoting objective function for neural conversation models. arXiv preprint arXiv:1510.03055 (2015)
14. Li, R., Ebrahimi Kahou, S., Schulz, H., Michalski, V., Charlin, L., Pal, C.: Towards deep conversational recommendations. Advances in neural information processing systems **31** (2018)
15. Li, S., Xie, R., Zhu, Y., Ao, X., Zhuang, F., He, Q.: User-centric conversational recommendation with multi-aspect user modeling. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 223–233 (2022)
16. Li, W., Wei, W., Qu, X., Mao, X.L., Yuan, Y., Xie, W., Chen, D.: TREA: Tree-structure reasoning schema for conversational recommendation. In: Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 2970–2982 (2023)
17. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
18. Long, Y., Hui, B., Yuan, C., Huang, F., Li, Y., Wang, X.: Multimodal recommendation dialog with subjective preference: A new challenge and benchmark. arXiv preprint arXiv:2305.18212 (2023)
19. Lu, Y., Bao, J., Song, Y., Ma, Z., Cui, S., Wu, Y., He, X.: Revcore: Review-augmented conversational recommendation. arXiv preprint arXiv:2106.00957 (2021)
20. Ma, W., Takanobu, R., Huang, M.: CR-walker: Tree-structured graph reasoning and dialog acts for conversational recommendation. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. pp. 1839–1851 (2021)
21. Radlinski, F., Boutilier, C., Ramachandran, D., Vendrov, I.: Subjective attributes in conversational recommendation systems: challenges and opportunities. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 12287–12293 (2022)
22. Reimers, N., Gurevych, I.: Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084 (2019)
23. Ren, Z., Tian, Z., Li, D., Ren, P., Yang, L., Xin, X., Liang, H., de Rijke, M., Chen, Z.: Variational reasoning about user preferences for conversational recommendation. In: proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval. pp. 165–175 (2022)
24. Sap, M., Le Bras, R., Allaway, E., Bhagavatula, C., Lourie, N., Rashkin, H., Roof, B., Smith, N.A., Choi, Y.: Atomic: An atlas of machine commonsense for if-then reasoning. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 3027–3035 (2019)

25. Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Hambro, E., Zettlemoyer, L., Cancedda, N., Scialom, T.: Toolformer: Language models can teach themselves to use tools. Advances in Neural Information Processing Systems **36** (2024)
26. Schlichtkrull, M., Kipf, T.N., Bloem, P., Van Den Berg, R., Titov, I., Welling, M.: Modeling relational data with graph convolutional networks. In: The semantic web: 15th international conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, proceedings 15. pp. 593–607. Springer (2018)
27. Speer, R., Chin, J., Havasi, C.: Conceptnet 5.5: An open multilingual graph of general knowledge. In: Proceedings of the AAAI conference on artificial intelligence. vol. 31 (2017)
28. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
29. Wahbeh, A., Al-Ramahi, M., El-Gayar, O., Elnoshokaty, A., Nasralah, T.: Conversational agents for mental health and well-being: discovering design recommendations using text mining (2023)
30. Xi, Y., Liu, W., Lin, J., Chen, B., Tang, R., Zhang, W., Yu, Y.: Memocrs: Memory-enhanced sequential conversational recommender systems with large language models. arXiv preprint arXiv:2407.04960 (2024)
31. Zhang, X., Jia, X., Liu, H., Liu, X., Zhang, X.: A goal interaction graph planning framework for conversational recommendation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 19578–19587 (2024)
32. Zhang, X., Xin, X., Li, D., Liu, W., Ren, P., Chen, Z., Ma, J., Ren, Z.: Variational reasoning over incomplete knowledge graphs for conversational recommendation. In: Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining. pp. 231–239 (2023)
33. Zhang, Y., Wu, L., Shen, Q., Pang, Y., Wei, Z., Xu, F., Long, B., Pei, J.: Multiple choice questions based multi-interest policy learning for conversational recommendation. In: Proceedings of the ACM Web Conference 2022. pp. 2153–2162 (2022)
34. Zhang, Y., Yu, J., Rao, Y., Zheng, L., Su, Q., Zhu, H., Yin, J.: Domain adaptation for subjective induction questions answering on products by adversarial disentangled learning. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 9074–9089. Association for Computational Linguistics, Bangkok, Thailand (Aug 2024)
35. Zhao, Z., Zhou, K., Wang, X., Zhao, W.X., Pan, F., Cao, Z., Wen, J.R.: Alleviating the long-tail problem in conversational recommender systems. In: Proceedings of the 17th ACM Conference on Recommender Systems. pp. 374–385 (2023)
36. Zhou, J., Wang, B., He, R., Hou, Y.: Crfr: Improving conversational recommender systems via flexible fragments reasoning on knowledge graphs. In: Proceedings of the 2021 conference on empirical methods in natural language processing. pp. 4324–4334 (2021)
37. Zhou, J., Wang, B., Yang, Z., Zhao, D., Huang, K., He, R., Hou, Y.: Cr-gis: Improving conversational recommendation via goal-aware interest sequence modeling. In: Proceedings of the 29th International Conference on Computational Linguistics. pp. 400–411 (2022)
38. Zhou, K., Zhao, W.X., Bian, S., Zhou, Y., Wen, J.R., Yu, J.: Improving conversational recommender systems via knowledge graph based semantic fusion. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 1006–1014 (2020)
39. Zou, L., Xia, L., Du, P., Zhang, Z., Bai, T., Liu, W., Nie, J.Y., Yin, D.: Pseudo dyna-q: A reinforcement learning framework for interactive recommendation. In: Proceedings of the 13th International Conference on Web Search and Data Mining. pp. 816–824 (2020)