Proceedings of the 2023 International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/8/20230282

User interest based contextual conversational recommender system

Yinghui Xia

AlgHub Ltd. Co., No. 1, Shanyuan Street, Zhongguancun, Haidian District, 100080, Beijing

yhxia@alghub.com

Abstract. The Conversational recommender system is for item recommendation during human-machine dialogue, where items can be movies, music, and goods. Scenarios can be intelligent customer service and shopping assistants. The system's goal is to generate high-quality recommendations in short-round interactions. There are two main problems with the system. Firstly, recommending through dialogues is a hard problem as less contextual information is available, resulting in fewer signals for recommendation and less precision for user interest modeling. Secondly, the recommender system considers the absence of interaction between a user and an item as a negative sample. However, this approach can introduce bias in the objectives because non-interaction could potentially be a positive sample. To tackle the previously mentioned concerns, this paper suggests a new model. This model adds an extra knowledge graph to enrich contextual information. Then, this model introduces contrastive learning to separate the distribution between positive and negative samples and improve learning efficiency. This paper further tests the model on the ReDial dataset, and experiments have shown that the method is effective and performs better than previous work.

Keywords: conversational recommender system, contrastive learning, user interest modeling, recommender system, natural language processing.

1. Introduction

As social media and smartphones have advanced, individuals have progressively relied on online conversational systems to acquire information and accomplish tasks. One such system is the conversational recommendation system (CRS), which can recommend content within conversations, provide real-time answers, and help users complete tasks [1-3]. As an AI-based dialogue system, it can provide accurate recommendations based on user input. Due to the increasing use of online dialogue systems, conversational recommendation systems have received considerable attention and become a hot topic of current research.

The difficulty of conversational recommender systems is to recommend content to users within shorter dialogue rounds. Li *et al.* first proposed the ReDial dataset, which serves as an important basis for evaluation [1]. Table 1 shows a sample of user-system dialogue. In addition, Li *et al.* utilized multiple functional analysis modules to train a conversational generative model in an end-to-end framework.

DOI: 10.54254/2755-2721/8/20230282

Table 1. Sample of user and system conversation for movie recommendaton. Important entities are marked in bold font.

User	Hello. How are you?
System	What are your favorite genres or movies in general?
User	I love Scarlett Johannson. Know any of her movies?
System	First movie to my mind is Lucy (2014) and Ghost in the Shell (2017)
User	Hmm. Haven't seen either! Awesome job. You're great!

Different solutions have emerged in recent years to address the conversational recommendation problem. The end-to-end conversational model is widely used after Sordoni and Bengio et al. proposed a hierarchical recurrent encoder-decoder framework (HRED) to build an initial conversational recommendation that integrates different level context information, such as session-level state and query-level state [4]. Li *et al.* use HRED framework to address movie recommender problem in Redial dataset. However, the framework only uses dialogue data as training data, extra knowledge graphs are introduced in later works for conversational recommender systems [5, 6]. For instance, Chen *et al.* import dialogue knowledge for recommendation not only based on mentioned items [7]. Zhou *et al.* extract topics and review information from dialogues to construct a conceptual knowledge graph for content indexing [8]. Huang *et al.* use multi-hop reasoning to leverage knowledge graphs to recommend [9]. Schlichtkrull *et al.* use external knowledge graphs like DBpedia to model items [10].

CRS contains two capabilities, recommendation and dialogue generation. End-to-end framework and external knowledge increase the dialogue context information and external knowledge. With more information, the recommendation capability is improved. In order to further improve the dialogue ability, in the end-to-end framework with dialogue context, encoder based on Transformer can help to improve text-modeling ability. For example, Zhou *et al.* propose KGSF framework [8], which is based on end-to-end dialogue modeling, and technologies of Attention and Transformer technology are used in text understanding and dialogue generation modules.

The recommender in CRS is targeted to retrieve items from the database based on user interest. Previous works adopted cross-entropy as the target function based on user-item interactions. However, user and item interaction is normally sparse and lacks true negative examples. False negative examples would lead to a biased model. Contrastive learning can separate the distance between positive and negative samples and improve the recommendation performance.

Based on above discussion, this paper proposes a new model that uses contrastive learning to improve the recommender component. In addition, knowledge graphs are used to acquire semantics and represent user local preferences. The history of user and content interactions can be used to represent user global preferences. After sufficiently modeling user local and global preferences, content representation can be used to recommend content. This new model has been tested on the ReDial dataset [1]. Experiments have demonstrated that the recommendation aspect of the model surpasses that of prior models.

The rest of this paper has the following structure. Section 2 provides an overview of the proposed methods, including user preference modeling and the contrastive learning method. Section 3 exhibits the experimental outcomes and benchmark analysis on datasets for conversational recommender systems. Section 4 discusses the influences of global user preference and contrastive learning. Section 5 concludes the paper.

2. Methods

This paper follows the KGSF framework [8]. The framework includes three parts: graph-based semantic fusion part, dialogue generation system part, and recommender system part.

The semantic fusion component based on graph theory follows the KGSF methodology, which utilizes ConceptNet as a knowledge graph focused on individual words, and employs graph convolutional neural network (GCN) to effectively represent and learn the semantic relationships among word nodes. GCN can encode every word in KG as embedding through word nodes relations, then entity embeddings in KG can be extracted and used in context modeling process. In addition, DBpedia is used

DOI: 10.54254/2755-2721/8/20230282

as an item-oriented KG. Then, the item KG is encoded by R-GCN. Since DBpedia is a heterogeneous KG, R-GCN can be used to model multi-type relation and entities, such as movies, actors, genres and so on. Both GCN and R-GCN are aimed at solving problem of KG modeling, or getting semantic representation from KG. To align the concept embedding and item embedding, KGSF uses MIM to improve the representation of words and items and get better user embedding. MIM is to simultaneously improve the semantic representation of two sources. By computing the KL divergence between two latent semantic space, the methodology of MIM can establish a connection between words and items in KGs.

The dialogue generation system part in the KGSF framework adopts Transformer to generate a reply. The words and items embedding from graph-based semantic fusion part are also fused into the dialogue generation process.

The recommender system part uses cross-entropy loss to calculate the distance between user embedding and item embedding. Given user embedding, the probability is calculated of recommend item i from the database to a user u:

$$Pr_{rec}(i) = softmax(\mathrm{p}_u^T \cdot \mathrm{n}_i) \tag{1}$$

In the provided equation, n_i represents the embedding for item i, and p_u represents the embedding for user u. During the training process, the similarity between the user embedding in each dialogue and all item embeddings is evaluated. Items that the user has liked are treated as positive samples, while the remaining items are considered negative samples:

$$L_{rec} = -\sum_{j=1}^{N}\sum_{i=1}^{M}\left[-(1-y_{ij})\cdot log(1-Pr_{rec}^{j}(i)) + y_{ij}\cdot log(Pr_{rec}^{j}(i))\right] + \lambda \times L_{MIM} \tag{2}$$

where j is conversation index, i is item index, $\mathrm{L_{MIM}}$ is the MIM loss, λ is weight of MIM los

The equations presented above treat items that have not been interacted with by the user as negative samples, which could introduce bias as these items may actually be potential positive samples. To address this issue, the current paper proposes the implementation of contrastive learning. Specifically, the matching task between the user embedding and item embedding is reformulated as a contrastive learning task. The loss function for contrastive learning can be expressed as shown below:

$$L_{CL}(\mathrm{p}, \mathrm{n}^+) = \log \frac{e^{\mathrm{sim}(\mathbf{p},\mathbf{n}^+)/\tau}}{\sum_{\mathbf{n}_i^- \in \{\mathbf{n}^-\}} e^{\mathrm{sim}(\mathbf{p},\mathbf{n}_i^-)/\tau}} \tag{3}$$

For each conversation, p is user embedding of the conversation, n+ is the recommended item, n− are items of not been recommended. Contrastive learning loss can replaces the cross entropy loss, then reduces bias caused by false negative samples. The final loss can be formulated as following:

$$L_{rec} = -\sum_{j=1}^{N} \log \frac{e^{\mathrm{sim}(\mathbf{p}_j,\mathbf{n}_j^+)/\tau}}{\sum_{\mathbf{n}_i^- \in \{\mathbf{n}^-\}} e^{\mathrm{sim}(\mathbf{p}_j,\mathbf{n}_i^-)/\tau}} + \lambda \times L_{MIM} \tag{4}$$

In the KGSF framework, concepts and entities are extracted from each user utterance in the dialogue. However, the user utterances are not directly used for modeling user embedding. In our model, user utterances are feature-extracted through pre-trained language models (PLMs) and fused into the user embedding, which are then matched with item embedding.

3. Experiments

3.1. Dataset

For the evaluation of the proposed method, the ReDial dataset is utilized as the primary dataset. ReDial, short for Recommendation Dialogues, is an annotated dataset of dialogues where users recommend movies to each other. The dataset involves two individuals in the conversation, where one acts as the recommendation seeker while the other acts as the recommender. Users discuss their movie preferences, which movies they have watched or not, and which ones they like or dislike. The dataset was obtained

DOI: 10.54254/2755-2721/8/20230282

using an interface and pairing mechanism developed by the dataset authors, which were mediated by Amazon Mechanical Turk (AMT). The dataset comprises of over 10,000 conversations, all focused on the theme of providing movie recommendations.

3.2. Metrics

Consistent with prior research on Conversational Recommender Systems, distinct evaluation metrics have been adopted for assessing recommendation and conversation tasks separately. The recommendation task has been evaluated using Recall@k (k=1,10,50), while the conversation task employs Distinct-n (n=2,3,4). For the purposes of this paper, the focus is solely on the recommendation task.

3.3. Evaluation

Table 2 shows the result of experiments. Proposed model is tested on ReDial dataset and compared to baseline (KGSF) and other models (Popularity / TextCNN / ReDial / KBRD / CR-Walker) in metrics of recommendation task. Methods of Popularity and TextCNN are documented in the paper of KGSF. The result shows that contrastive learning can get a better performance.

Table 2. Experiment on ReDial dataset.

Methods	Recall@1	Recall@10	Recall@50
Prior Works			
Popularity	2.0%	9.7%	23.9%
TextCNN	1.1%	8.1%	23.9%
ReDial	2.1%	7.5%	20.1%
KBRD	2.6%	8.5%	24.2%
CR-Walker	**4.0%**	18.7%	**37.6%**
Baseline			
KGSF	3.6%	17.4%	36.6%
Ours			
+contrastive learning	3.9%	**19.3%**	**37.6%**

3.4. Samples ratio

Different ratios of positive and negative samples were assessed, ranging from 1:N (where N=1, 2, 4, 8, 16, etc.). The best performance was achieved at a ratio of 1:8. Figure 1 depicts the correlation between batch negative sampling and model performance in Recall@k (k=1,10,50). When summing up all Recall@k for each sampling rate into a line graph of SUM, the maximum value was obtained at a sampling rate of 1:8. These findings indicate that having a sufficient number of negative samples is crucial for effective contrastive learning in training a recommender system. However, increasing the negative sampling ratio does not always contribute to the system's performance.

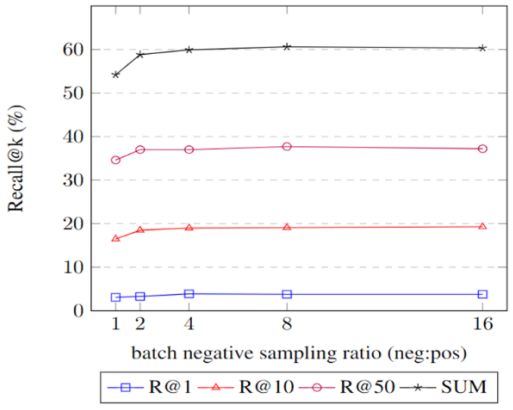

Figure 1. Recall@k of different batch negative sampling ratio.

DOI: 10.54254/2755-2721/8/20230282

3.5. Batch size

Batch size can control the number of samples per batch. In the experiment, different batch size is tested, such as 64, 128, 256, and 512 etc.. The best batch size for the model of this paper in ReDial dataset is 128.

3.6. Embedding size

Embedding size can control the representation ability of the framework. In the experiment, different embedding size of words and items representation is tested, such as 128, 256, and 512 etc.. The best embedding size for the model of this paper in ReDial dataset is 512.

3.7. Experiment

The experimental results have illustrated that the performance of a recommender system can be improved through the application of contrastive learning. During the training process, positive samples (i.e. items that a user has interacted with) are contrasted with randomly selected negative samples (i.e. items that the user has not interacted with) drawn from non-positive samples. Furthermore, multiple ratios of positive and negative samples were tested, with the highest level of performance achieved when the ratio was 1:8. The study also explored different hyperparameters, including varying batch sizes and embedding sizes.

4. Discussion

4.1. User embedding enrichment

The contribution of this paper has shown promising results in improving the performance of conversational recommender systems. However, there are still several aspects that could be improved. One possible way to further enhance user representations is to incorporate additional sources of information, such as historical behavioral data, to better capture the user's preferences and interests. Furthermore, the user-item interaction graph can be extended to incorporate temporal dynamics, as user preferences and interests can evolve over time. This would provide a more accurate representation of user behavior and better inform the recommendation process. In addition, future research could also explore the use of advanced natural language processing techniques to better understand the nuances of user conversations and improve the quality of recommendations. Overall, there is still much potential for further enhancing the performance and effectiveness of conversational recommender systems.

4.2. Dialogue generation enhancement

Regarding the Dialogue Generation Enhancement, while the use of contrastive learning has improved the performance of the model in generating high-quality recommendations. Although the proposed method has shown promising results, there remains potential for further improvement. For example, the system could be designed to better handle long-term dependencies in the dialogue, such as maintaining a consistent topic over multiple turns. Additionally, incorporating knowledge about the user's conversational style, such as their level of formality or preferred tone, could help to generate more natural and personalized recommendations. Moreover, sentiment modeling is a crucial aspect in the generation of dialogues for conversational recommender systems. By analyzing the emotional tendencies of users, the system can gain a deeper understanding of their emotional state and provide recommendations that align with their current mood or attitude. In summary, future research could expand upon the findings of this paper and explore additional approaches to improve the performance of conversational recommender systems.

5. Conclusion

To conclude, this paper presents a novel method to overcome two key challenges in conversational recommender systems: the absence of contextual information and the bias caused by non-interactions. The proposed approach enhances the recommendation process by including global user-item interaction

DOI: 10.54254/2755-2721/8/20230282

data with contextual information. This combination enables a better understanding of user interests and provides more accurate recommendations. Words and items representation are generated from an additional knowledge graph, while user representation is generated from dialogue context. The paper imports contrastive learning to address the bias caused by incomplete interactions and improve learning efficiency. The ReDial dataset was used to test the proposed method, and various hyperparameters were evaluated, such as embedding size and batch size. The results indicate that the proposed method surpasses previous works and presents a viable approach to enhancing conversational recommender systems, especially in contexts such as intelligent customer service and shopping assistants. Furthermore, this study highlights future research opportunities such as enriching user embeddings and enhancing the dialogue generation process by incorporating more advanced natural language processing techniques. Overall, this paper makes a contribution to the field of conversational recommender systems by proposing a new approach that can effectively address the challenges of contextual information and incomplete interactions. The proposed method shows promise in improving the precision of recommendations and enhancing user experience, which can have significant implications for e-commerce.

References

[1] Raymond Li, Samira Ebrahimi Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Christopher Joseph Pal. Towards deep conversational recommendations. In Neural Information Processing Systems, 2018.

[2] Yueming Sun and Yi Zhang. Conversational recommender system. The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, 2018

[3] Lizi Liao, Ryuichi Takanobu, Yunshan Ma, Xun Yang, Minlie Huang, and Tat-Seng Chua. Deep conversational recommender in travel. ArXiv, abs/1907.00710, 2019

[4] Alessandro Sordoni, Yoshua Bengio, Hossein Vahabi, Christina Lioma, Jakob Grue Simonsen, and Jianyun Nie. A hierarchical recurrent encoder-decoder for generative context-aware query suggestion. Proceedings of the 24th ACM International on Conference on Information and Knowledge Management, 2015.

[5] Xiang Wang, Dingxian Wang, Canran Xu, Xiangnan He, Yixin Cao, and Tat-Seng Chua. Explainable reasoning over knowledge graphs for recommendation. In AAAI Conference on Artificial Intelligence, 2018

[6] Wayne Xin Zhao, Gaole He, Hongjian Dou, Jin Huang, Siqi Ouyang, and Ji-Rong Wen. Kb4rec: A data set for linking knowledge bases with recommender systems. Data Intelligence, 1:121–136, 2018

[7] Qibin Chen, Junyang Lin, Yichang Zhang, Ming Ding, Yukuo Cen, Hongxia Yang, and Jie Tang. Towards knowledge-based recommender dialog system. ArXiv, abs/1908.05391, 2019

[8] Kun Zhou, Wayne Xin Zhao, Shuqing Bian, Yuanhang Zhou, Ji rong Wen, and Jingsong Yu. Improving conversational recommender systems via knowledge graph based semantic fusion. Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020

[9] Jin Huang, Zhaochun Ren, Wayne Xin Zhao, Gaole He, Ji-Rong Wen, and Daxiang Dong. Taxonomy-aware multi-hop reasoning networks for sequential recommendation. Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, 2019

[10] Christian Bizer, Jens Lehmann, Georgi Kobilarov, S. Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. Dbpedia - a crystallization point for the web of data. J. Web Semant., 7:154–165, 2009