RALI@TREC iKAT 2023: Generative Query Reformulation for Conversational Information Seeking

Fengran Mo RALI, Université de Montréal Montréal, Québec, Canada fengran.mo@umontreal.ca Bole Yi RALI, Université de Montréal Montréal, Québec, Canada bole.yi@umontreal.ca

Jian-Yun Nie RALI, Université de Montréal Montréal, Québec, Canada nie@iro.umontreal.ca

ABSTRACT

The Recherche Appliquée en Linguistique Informatique (RALI) team has participated in the 2023 TREC Interactive Knowledge Assistance Track (iKAT). This paper introduces our approaches and reports our results on the passage ranking task. The most challenging in conversational information seeking is to reveal the user's real search intent. To tackle these challenges, we employ a combination of query rewriting and query expansion techniques to rephrase conversational queries using generative language models in both supervised and zero-shot manner. Furthermore, to establish a connection between query reformulation and the retrieval process, we implement a knowledge infusion mechanism to enhance both procedures during training. The outcome of our efforts yields impressive results, with an nDCG@5 score of 16.24% and an MRR of 32.75% in our best-performing experiments. Besides, we also explore the impact of personal information on the search results based on GPT-4, showing that not all query turns are associated with personalized information needs.

KEYWORDS

Query Reformulation, Conversational Search

ACM Reference Format:

Fengran Mo, Bole Yi, and Jian-Yun Nie. 2024. RALI@TREC iKAT 2023: Generative Query Reformulation for Conversational Information Seeking. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnnnnnn

1 INTRODUCTION

Conversational information seeking is an emerging area within information retrieval that is poised to become the future of search engines [6]. The TREC Interactive Knowledge Assistance Track (iKAT) is dedicated to the advancement of collaborative informationseeking conversational agents capable of customizing their responses based on their understanding of and interaction with the user. Specifically, iKAT strives to support multi-path, multi-turn, multi-perspective conversations. This means that the nature of the conversation not only depends on prior responses but also on the user, taking into account their background, perspective, context,

Conference'17, July 2017, Washington, DC, USA

© 2024 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00 https://doi.org/10.1145/nnnnnnnnnnnn and more. As diverse individuals engage in querying on various topics, systems must construct a comprehensive profile of the user to address their information requirements effectively.

In the conventional conversational search, the primary challenge lies in the fact that queries are often context-dependent, emphasizing the need to comprehend the search intent within the conversational context. To tackle this issue, current methods can be roughly divided into two categories. The first group involves using the entire context as a query and training a model to assess the relevance between the context and passages [11, 17, 21, 24, 31]. However, this approach necessitates additional training of the retriever to take the long context as input, which is not always practical [18]. In practice, what is typically available is a general retriever, such as an ad-hoc search retriever, that utilizes a standalone query. The second group of approaches focuses on generating a de-contextualized query using query reformulation techniques [5]. This type of query can be employed with any *off-the-shelf* retrievers.

In the existing literature, two primary categories of query reformulation techniques have been extensively explored, i.e., query rewriting and query expansion. In the case of query rewriting, a generative model is employed to rephrase the current query in a manner that resembles a human-rewritten query, as demonstrated in previous studies [26, 30]. On the other hand, query expansion concentrates on enlarging the current query by incorporating contextually relevant terms [10, 27]. Although both methods exhibit promising outcomes, they have typically been investigated as separate approaches. There are two noteworthy limitations: (1) Query rewriting and query expansion can yield distinct outcomes. Query rewriting is particularly effective for addressing ambiguous queries and adding missing tokens, whereas query expansion focuses on supplementing the query with additional information. Both of these effects are valuable for query reformulation, making it advantageous to utilize both techniques. (2) Previous query rewriting models have primarily been optimized to generate human-rewritten queries, independently of the passage ranking task. While humanrewritten queries often outperform the original ones, research has demonstrated that they may not always serve as the most effective search queries on their own, as highlighted in prior studies[20, 28]. Therefore, it is valuable to incorporate additional criteria directly linked to ranking performance when reformulating a query.

To improve the retrieval effectiveness of the reformulated queries in conversational search, Mo et al. [20] propose **ConvGQR**, a new Generative Query Reformulation framework for **Conv**ersational search, which combines query rewriting with query expansion. We follow this framework and extend the idea to leverage the large language model (LLM) for the same goal.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

In addition to query rewriting based on human-rewritten queries, the ConvGQR method also acquires the capability to generate potential answers to the query, which can be subsequently used for query expansion. This strategy is grounded in the understanding that a passage containing the generated potential answer is more likely to be relevant. This is because either the generated answer itself is the correct answer, or it may co-occur with the correct answer within the same passage. During supervised training, the query reformulation model is trained by incorporating both query rewriting and query expansion criteria into the loss function. Furthermore, the learning process for both query rewriting and expansion is influenced by the information present in relevant passages through a carefully designed knowledge infusion mechanism. This encourages the query generation process to yield improved search performance. We conduct the experiments training on the QReCC [2] dataset and then evaluate the 2023 TREC iKAT test set. Since the ConvGQR framework focuses on pure conversational information seeking rather than considering the personalized elements in the 2023 TREC iKAT, we further implement GPT-4 to implicitly select the personal textual knowledge base (PTKB) provided in the dataset to explore the impact of such information for search results.

2 RELATED WORK

Conversational search is primarily approached through two key methods: conversational query rewriting (CQR) and conversational dense retrieval (CDR). CQR focuses on converting context-dependent queries in a search session into independent queries. This is achieved through techniques such as selecting relevant tokens from the search session [13, 23, 27] and training a generative rewriter model using human-rewritten queries paired with their corresponding sessions [12, 14, 26, 30]. Some studies incorporate reinforcement learning [3, 29] or ranking signals [15, 20] to align the generation process with the downstream search task. In contrast, CDR employs conversational search session data to carry out end-to-end dense retrieval. To enhance conversational search performance, advanced techniques like context denoising [9, 16, 18, 21, 22, 24, 31], data augmentation [4, 11, 17], and the identification of challenging negative examples [8] have been explored.

3 METHODOLOGY

We implement two methods with the same idea of combining query rewriting and query expansion for query reformulation. On one hand, we inherit the *ConvGQR* framework to train the model which includes three parts: (1) Query rewriting to approach the humanrewritten query (Sec. 3.1.1). (2) Query expansion based on the generated potential answer (Sec. 3.1.2). (3) Knowledge infusion mechanism that connects the optimization for both query reformulation and passage retrieval (Sec. 3.2). On the other hand, we try to leverage the powerful capacity of the large language model to perform zero-shot query reformulation (Sec. 3.3), which directly generates the rewritten queries and the expansion part based on the given conversational session context.

3.1 Query Reformulation

Both query rewriting and expansion leverage the historical context $C^t = \{q_t, a_t\}_{t=1}^{i-1}$ combined with the current query q_i as their input. Similar to the input format employed in Mo et al. [20], a separation token, denoted as "[SEP]", is introduced between each turn, and the turns are concatenated in reverse order, as described in Eq. 1.

$$[\mathsf{CLS}] q_i [\mathsf{SEP}] q_{i-1} \cdots q_1 [\mathsf{SEP}]. \tag{1}$$

3.1.1 Query Rewriting. The query rewriting model aims to establish a function $\mathcal{H}(C^t, q_i) = q^*$ using a generative pre-trained language model (PLM), where q^* represents the sequence employed as the supervision signal, typically a human-rewritten query from the training data. This function, \mathcal{H} , essentially seeks to incorporate the information present in C^t but absent in q_i in order to approximate q^* . In essence, the overall objective can be understood as the optimization of the parameter $\theta_{\mathcal{H}}$ of the function \mathcal{H} through maximum likelihood estimation:

$$\theta_{\mathcal{H}} = \arg \max_{\theta_{\mathcal{H}}} \prod_{t=1}^{i-1} \Pr\left(q^* | \mathcal{H}\{C^t, q_i\}, \theta_{\mathcal{H}}\right).$$
(2)

3.1.2 Query Expansion. Recent research has shown that the PLMs possess the capability to directly provide responses to questions, functioning as a closed-book question-answering system [1] by utilizing their stored knowledge. Although the accuracy of these generated answers is not guaranteed, the potential answers can still serve as valuable expansion terms [19]. These expansion terms can guide the search process toward a passage containing the potential answer or a similar response. To train the generation process, we employ the correct answer, denoted as a^* , for each query turn as the training objective. Depending on the dataset, a^* could be a concise entity, a continuous text segment, or even non-contiguous text segments. During the inference stage for a new query, the potential answers are generated by the query expansion model and then used to augment the previously rewritten query.

The resulting reformulated query is created by combining both the rewritten query and the potential answer that is generated. Both of the generative PLMs used for the rewriting and expansion goal are fine-tuned by minimizing the negative log-likelihood loss to predict the respective target. This is done with an input sequence $\{w_l\}_{l=1}^{L}$ as described in Eq. 3, although they are trained with distinct data.

$$\mathcal{L}_{\text{gen}} = -\sum_{l=1}^{L} \log \left(\Pr(w_l | w_{1:l-1}, C^t, q_i) \right).$$
(3)

3.2 Knowledge Infusion

A significant drawback in the current generative conversational query reformulation methods is their lack of consideration for the dependency between generation and retrieval, where they are typically trained in isolation. To overcome this limitation, a knowledge infusion mechanism is introduced aiming at enhancing the training of both query reformulation and search tasks. The underlying idea is to require the generative model to produce a query representation that is close to that of a relevant passage. By ensuring that the hidden states of the generative model encapsulate information

Method	MRR	NDCG@3	NDCG@5	NDCG	P@20	Recall@20	Recall	MAP
ConvGQR	0.3275	0.1654	0.1624	0.1518	0.1421	0.0611	0.2034	0.0551
LLMConvGQR	0.3224	0.1318	0.1338	0.1200	0.1169	0.0523	0.1621	0.0461

Table 1: Performance of dense retrieval with two generative query reformulation methods on 2023 TREC iKAT test set. The version of the relevance judgment file is the newest released "pruned_qrels".

Dataset	Split	#Conv.	#Turns(Qry.)	#Collection
QReCC	Train Test	10,823 2,775	63,501 16,451	54M
iKAT-23	Train Test	11 25	95 332	12M

Table 2: Statistics of conversational search datasets.

from relevant passages, the queries generated using these representations have the potential to enhance search performance due to the increased semantic similarity.

To accomplish this objective, an effective approach involves integrating the knowledge embedded in the relevant passage representation into the query representation while fine-tuning the generative PLMs. Specifically, we initiate the process by employing an off-the-shelf retriever, serving as an encoder, to generate a representation $\mathbf{v}_{p_{+}}$ for the relevant passage. To ensure consistency, the retriever used here is identical to the one employed for the search process. Consequently, the representation space for passages remains consistent for both query reformulation and retrieval phases. After encoding the session query representation v_S using the generative model, we distill the knowledge from v_{p_+} and integrate it into v_S by minimizing the Mean Squared Error (MSE), as indicated in Eq. 4. Both \mathbf{v}_S and \mathbf{v}_{p_+} are sequence-level representations based on the first special token "[CLS]". Ultimately, the overall training objective, denoted as \mathcal{L}_{total} , encompasses both the query generation loss \mathcal{L}_{OG} , and the retrieval loss \mathcal{L}_{R} . To balance the impact of query generation and retrieval, a weight factor α is employed.

$$\mathcal{L}_{R} = MSE(\mathbf{v}_{S}, \mathbf{v}_{p_{+}}), \tag{4}$$

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{QG}} + \alpha \cdot \mathcal{L}_{\text{R}}.$$
 (5)

3.3 Large Language Model for Zero-Shot Query Reformulation

Both query rewriting and query expansion involve text generation processes. Consequently, we have extended our approach to harness Large Language Models (LLM) for zero-shot query reformulation. Specifically, we have devised instructions to guide the LLM in generating the desired text, which includes the rewritten query and expansion parts (potential answers). The process involves initially generating the rewritten query and then using it to generate the expanded portion. The instruction templates follow the format [Instruction, Input Context, Restriction] for both query rewriting and query expansion. The instructions for these tasks are as follows: "Given the historical queries [$q_1, q_2, ..., q_{n-1}$], rewrite the current query q_n to be under 64 tokens" and "Given the rewritten query q'_n , generate an answer that is under 64 tokens".

To explore the impact of PTKB information on the search results, we implement GPT-4 to implicitly select PTKB for personalized query reformulation. The procedure contains two stages: selection then reformulation. Specifically, we first prompt the GPT-4 to select the relevant sentences in PTKB for the current query turn, then incorporate the selected user information with the input context to perform the query reformulation.

4 EXPERIMENT

4.1 Experimental Setup

4.1.1 Datasets. We use the relevant judgment for passage ranking in the test set from TREC iKAT for evaluation. For the supervised learning manner, we use the human-rewritten query and gold answer annotations in the QReCC dataset to train the models. The statistic is shown in Table 2.

4.1.2 Implementation Details. For supervised learning manner, we implement the generative language models for query reformulation based on T5-base [25] models. We use the QReCC dataset to fine-tune the generative PLMs and keep the dense retriever frozen when acting as a passage encoder. For the zero-shot manner, on one hand, we directly leverage ChatGPT (gpt-turbo-3.5) to generate a rewritten query and its answer as an expansion for each query turn. On the other hand, we leverage GPT-4 to select PTKB for personalized query reformulation. The dense retrieval is performed using Faiss [7].

4.2 Experimental Results

The experimental results of our two main methods without considering personal information are evaluated officially and reported in Table 1. Initially, we can see the overall performance of ConvGQR is better than LLMConvGQR, which shows the fine-tuning paradigm is still efficient and the LLM-based method can be further optimized by designing a better prompt template. Besides, we observe that the NDCG@1000 is better than NDCG@100 in both methods, which means the relevant passages are ranked behind and the ranking results still have room for improvement. Nevertheless, we do not consider the user profiles for query reformulation in these two methods, since the relevant judgment annotation does not consider the personalized information.

4.3 Impact of PTKB for Query Reformulation

The results of the impact of using PTKB for personalized query reformulation are shown in Table [?]. We observe that not using PTKB performs better than with such information, which indicates that not all query turns are necessarily associated with personalized

Method	MRR	NDCG@3	NDCG@5	NDCG	P@20	Recall@20	Recall	MAP
GPT-4 w/. PTKB - expansion	0.1418 0.2748	0.0562 0.1217	$0.0564 \\ 0.1222$	$0.1108 \\ 0.2001$	0.0526 0.1263	0.0260 0.0642	0.2006 0.3269	0.0249 0.0613
GPT-4 w/o PTKB - expansion	0.2151 0.2968	0.0910 0.1295	0.0903 0.1138	0.1436 0.1992	0.0805 0.1207	0.0425 0.0606	0.2354 0.3213	0.0429 0.0529

Table 3: Performance of dense retrieval using GPT-4 with or without PTKB information on 2023 TREC iKAT test set.

information needs in this dataset and more proper approaches should be designed for the usage of PTKB. Besides, we find that not incorporating the generated answers as an expansion achieves better results. This might be attributed to the generated answers by LLM containing additional noise and a filter mechanism is desired before leveraging the generated expansion.

5 CONCLUSION AND FUTURE WORK

In this paper, we present our solution to the 2023 TREC iKAT passage ranking task focusing on implementing the query reformulation technique. The query reformulation procedure combines query rewriting and query expansion based on the supervised trained model or large language model in a zero-shot manner. The overall results show that the small-size fine-tuned model performs better than the large language model. Besides, we analyze the impact of considering the personal text knowledge base (PTKB) for query reformulation, which indicates that not all query turns are associated with personalized information needs. Thus, a mechanism to determine whether a query is involved with the personalized element is desirable in future work.

REFERENCES

- Vaibhav Adlakha, Shehzaad Dhuliawala, Kaheer Suleman, Harm de Vries, and Siva Reddy. 2022. TopiOCQA: Open-domain Conversational Question Answering with Topic Switching. *Transactions of the Association for Computational Linguistics* 10 (2022), 468–483.
- [2] Raviteja Anantha, Svitlana Vakulenko, Zhucheng Tu, Shayne Longpre, Stephen Pulman, and Srinivas Chappidi. 2021. Open-Domain Question Answering Goes Conversational via Question Rewriting. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 520–534.
- [3] Zhiyu Chen, Jie Zhao, Anjie Fang, Besnik Fetahu, Rokhlenko Oleg, and Shervin Malmasi. 2022. Reinforced Question Rewriting for Conversational Question Answering. (2022).
- [4] Zhuyun Dai, Arun Tejasvi Chaganty, Vincent Y Zhao, Aida Amini, Qazi Mamunur Rashid, Mike Green, and Kelvin Guu. 2022. Dialog Inpainting: Turning Documents into Dialogs. In International Conference on Machine Learning. PMLR, 4558–4586.
- [5] Ahmed Elgohary, Denis Peskov, and Jordan Boyd-Graber. 2019. Can You Unpack That? Learning to Rewrite Questions-in-Context. 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). 5918–5924.
- [6] Jianfeng Gao, Chenyan Xiong, Paul Bennett, and Nick Craswell. 2022. Neural approaches to conversational information retrieval. arXiv preprint arXiv:2201.05176 (2022).
- [7] Jeff Johnson, Matthijs Douze, and Hervé Jégou. 2019. Billion-scale similarity search with gpus. IEEE Transactions on Big Data 7, 3 (2019), 535–547.
- [8] Sungdong Kim and Gangwoo Kim. 2022. Saving dense retriever from shortcut dependency in conversational search. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 10278–10287.
- [9] Antonios Minas Krasakis, Andrew Yates, and Evangelos Kanoulas. 2022. Zeroshot Query Contextualization for Conversational Search. In Proceedings of the 45th International ACM SIGIR conference on research and development in Information Retrieval (SIGIR).

- [10] Vaibhav Kumar and Jamie Callan. 2020. Making Information Seeking Easier: An Improved Pipeline for Conversational Search. In *Empirical Methods in Natural Language Processing*.
- [11] Sheng-Chieh Lin, Jheng-Hong Yang, and Jimmy Lin. 2021. Contextualized Query Embeddings for Conversational Search. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. 1004–1015.
- [12] Sheng-Chieh Lin, Jheng-Hong Yang, Rodrigo Nogueira, Ming-Feng Tsai, Chuan-Ju Wang, and Jimmy Lin. 2020. Conversational question reformulation via sequence-to-sequence architectures and pretrained language models. arXiv preprint arXiv:2004.01909 (2020).
- [13] Sheng-Chieh Lin, Jheng-Hong Yang, Rodrigo Nogueira, Ming-Feng Tsai, Chuan-Ju Wang, and Jimmy Lin. 2021. Multi-stage conversational passage retrieval: An approach to fusing term importance estimation and neural query rewriting. ACM Transactions on Information Systems (TOIS) 39, 4 (2021), 1–29.
- [14] Kelong Mao, Zhicheng Dou, Haonan Chen, Fengran Mo, and Hongjin Qian. 2023. Large Language Models Know Your Contextual Search Intent: A Prompting Framework for Conversational Search. arXiv preprint arXiv:2303.06573 (2023).
- [15] Kelong Mao, Zhicheng Dou, Bang Liu, Hongjin Qian, Fengran Mo, Xiangli Wu, Xiaohua Cheng, and Zhao Cao. 2023. Search-Oriented Conversational Query Editing. In Findings of the Association for Computational Linguistics: ACL 2023. 4160–4172.
- [16] Kelong Mao, Zhicheng Dou, and Hongjin Qian. 2022. Curriculum Contrastive Context Denoising for Few-shot Conversational Dense Retrieval. In Proceedings of the 45th International ACM SIGIR conference on research and development in Information Retrieval (SIGIR).
- [17] Kelong Mao, Zhicheng Dou, Hongjin Qian, Fengran Mo, Xiaohua Cheng, and Zhao Cao. 2022. ConvTrans: Transforming Web Search Sessions for Conversational Dense Retrieval. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. 2935–2946.
- [18] Kelong Mao, Hongjin Qian, Fengran Mo, Zhicheng Dou, Bang Liu, Xiaohua Cheng, and Zhao Cao. 2023. Learning Denoised and Interpretable Session Representation for Conversational Search. In Proceedings of the ACM Web Conference 2023. 3193– 3202.
- [19] Yuning Mao, Pengcheng He, Xiaodong Liu, Yelong Shen, Jianfeng Gao, Jiawei Han, and Weizhu Chen. 2021. Generation-Augmented Retrieval for Open-Domain Question Answering. 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). 4089–4100.
- [20] Fengran Mo, Kelong Mao, Yutao Zhu, Yihong Wu, Kaiyu Huang, and Jian-Yun Nie. 2023. ConvGQR: Generative Query Reformulation for Conversational Search. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. 4998–5012.
- [21] Fengran Mo, Jian-Yun Nie, Kaiyu Huang, Kelong Mao, Yutao Zhu, Peng Li, and Yang Liu. 2023. Learning to Relate to Previous Turns in Conversational Search. In 29th ACM SIGKDD Conference On Knowledge Discover and Data Mining (SIGKDD).
- [22] Fengran Mo, Chen Qu, Kelong Mao, Tianyu Zhu, Zhan Su, Kaiyu Huang, and Jian-Yun Nie. 2024. History-Aware Conversational Dense Retrieval. arXiv preprint arXiv:2401.16659 (2024).
- [23] Hongjin Qian and Zhicheng Dou. 2022. Explicit Query Rewriting for Conversational Dense Retrieval. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. 4725–4737.
- [24] Chen Qu, Liu Yang, Cen Chen, Minghui Qiu, W Bruce Croft, and Mohit Iyyer. 2020. Open-retrieval conversational question answering. In Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval. 539–548.
- [25] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.* 21, 140 (2020), 1–67.
- [26] Svitlana Vakulenko, Shayne Longpre, Zhucheng Tu, and Raviteja Anantha. 2021. Question rewriting for conversational question answering. In Proceedings of the 14th ACM International Conference on Web Search and Data Mining. 355–363.
- [27] Nikos Voskarides, Dan Li, Pengjie Ren, Evangelos Kanoulas, and Maarten de Rijke. 2020. Query resolution for conversational search with limited supervision.

In Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval. 921–930.
[28] Zeqiu Wu, Yi Luan, Hannah Rashkin, David Reitter, and Gaurav Singh Tomar.

- [28] Zeqiu Wu, Yi Luan, Hannah Rashkin, David Reitter, and Gaurav Singh Tomar. 2021. CONQRR: Conversational Query Rewriting for Retrieval with Reinforcement Learning. arXiv preprint arXiv:2112.08558 (2021).
- [29] Zeqiu Wu, Yi Luan, Hannah Rashkin, David Reitter, and Gaurav Singh Tomar. 2022. CONQRR: Conversational Query Rewriting for Retrieval with Reinforcement Learning. (2022).
- [30] Shi Yu, Jiahua Liu, Jingqin Yang, Chenyan Xiong, Paul Bennett, Jianfeng Gao, and Zhiyuan Liu. 2020. Few-shot generative conversational query rewriting. In Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval. 1933–1936.
- [31] Shi Yu, Zhenghao Liu, Chenyan Xiong, Tao Feng, and Zhiyuan Liu. 2021. Fewshot conversational dense retrieval. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 829–838.