ConQRR: Conversational Query Rewriting for Retrieval with Reinforcement Learning

Zeqiu Wu ♢* **Yi Luan** ♠ **Hannah Rashkin** ♠ **David Reitter** ♠
Hannaneh Hajishirzi ♢♣ **Mari Ostendorf** ♢ **Gaurav Singh Tomar** ♠
♢University of Washington ♠Google Research ♣Allen Institute for AI
{zeqiuwu1,hannaneh,ostendor}@uw.edu
{luanyi,hrashkin,reitter,gtomar}@google.com

Abstract

Compared to standard retrieval tasks, passage retrieval for conversational question answering (CQA) poses new challenges in understanding the current user question, as each question needs to be interpreted within the dialogue context. Moreover, it can be expensive to re-train well-established retrievers such as search engines that are originally developed for non-conversational queries. To facilitate their use, we develop a query rewriting model ConQRR that rewrites a conversational question in the context into a standalone question. It is trained with a novel reward function to directly optimize towards retrieval using reinforcement learning and can be adapted to any off-the-shelf retriever. ConQRR achieves state-of-the-art results on a recent open-domain CQA dataset containing conversations from three different sources, and is effective for two different off-the-shelf retrievers. Our extensive analysis also shows the robustness of ConQRR to out-of-domain dialogues as well as to zero query rewriting supervision.

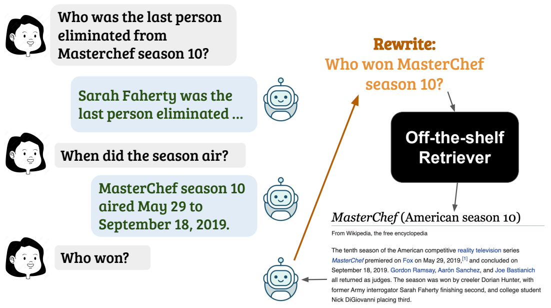

Figure 1: A CQA agent rewrites the current user question into a more effective one (in orange) for the given *off-the-shelf* retriever to find the most relevant passage.

1 Introduction

Passage retrieval in an open-domain conversational question answering (CQA) system (Anantha et al., 2021), compared to standard retrieval tasks (Voorhees and Tice, 2000; Bajaj et al., 2016), poses new challenges of understanding user questions within the dialogue context. Most existing conversational retrieval models (Yu et al., 2021; Lin et al., 2021; Kim and Kim, 2022) rely on training specific retrievers like dual encoders (Karpukhin et al., 2020). However, re-training well-established retrievers for conversational queries can be expensive or even infeasible due to their complicated system designs (e.g., those used in search engines). Moreover, the preference and availability of such off-the-shelf retrievers can vary depending on the end users.

The task of question-in-context rewriting or query rewriting (QR) in a conversation (Elgohary et al., 2019; Dalton et al., 2020) is to convert a context-dependent question into a self-contained question. It enables the use of any off-the-shelf retriever (Table 1), which we define as a retriever that cannot be fine-tuned or provide access to any of its internal architecture design or intermediate results (i.e., can only be seen as a black-box).

Therefore, in this paper, we focus on *query rewriting* for the task of *conversational passage retrieval* in a CQA dialogue with *any off-the-shelf* retrieval system that can only be used as a black box. Specifically, we seek to build a QR model that rewrites a user query into the input of the retriever, in such a way that optimizes for passage retrieval performance. Figure 1 shows an example of our task, where given an off-the-shelf retriever, the agent rewrites the current user query "Who won?" into a more effective query for retrieval.

Recent work that leverages QR for conversational passage retrieval (Anantha et al., 2021; Dalton et al., 2020) collects human-rewritten queries to train a supervised QR model. However, humans are usually instructed to rewrite conversational queries to be unambiguous to a human outside the dialogue context, which does not necessarily align with *the goal in our task*—to optimize the retrieval perfor-

*Work done during an internship at Google Research.

mance. We conduct comprehensive experiments in Section 4.3 to confirm these human rewrites indeed sometimes omit information from the dialogue context that is useful to the retriever. This limitation of human query rewrites impacts supervised training. In addition, prior supervised QR models are agnostic to downstream retrievers as they are separately trained before their predicted rewrites being used for retrieval during inference.

We propose a reinforcement learning (RL)-based model ConQRR (**Con**versational **Q**uery **R**ewriting for **R**etrieval). It directly optimizes the rewritten query towards retrieval performance, using only weak supervision from retrieval. We adopt a novel reward function that computes an approximate but effective retrieval performance metric on in-batch passages at each training step. Our reward function does not assume any specific retriever model design, and is generic enough for ConQRR to adapt to any off-the-shelf retriever.

We show ConQRR outperforms existing QR models on a recent large-scale open-domain CQA dataset QReCC (Anantha et al., 2021) by over 12% and 14% for BM25 and a neural dual encoder retriever model (Ni et al., 2021) respectively, averaging over three retrieval metrics. We observe the performance boost on all three QReCC subsets from different conversation sources, including one that only appears in the test set (i.e., out-of-domain).

To conclude, our contributions are as follows. 1) We introduce ConQRR as the first RL-based QR model that can be adapted to and optimized towards any off-the-shelf retriever for conversational retrieval. 2) We demonstrate that ConQRR achieves state-of-the-art results with off-the-shelf retrievers on QReCC with conversations from three sources, and is effective for two retrievers including BM25 and a dual encoder model. 3) Our analysis shows ConQRR trained with *no human rewrite supervision* provides better retrieval results than strong baselines trained with full supervision, and is robust to out-of-domain dialogues, topic shifts and long dialogue contexts. 4) We conduct a novel quantitative study to analyze the limitations and utility of human rewrites in retrieval performance, which are largely unexplored in prior work.

2 Related Work

2.1 Conversational Question Answering

Most existing CQA datasets (Choi et al., 2018; Reddy et al., 2019) are designed for the task of reading a document to answer questions in a conversation, which does not require the retrieval step. In contrast, QReCC (Anantha et al., 2021) is a recent open-domain CQA dataset where a conversational agent retrieves the most relevant passage(s) before generating an answer to the question.

	Fine-Tune	Arch Type
ConvDR (Yu et al., 2021)	Part	Dual Encoder
CQE (Lin et al., 2021)	Part	Dual Encoder
Kim and Kim (2022)	Yes	Dual Encoder
QR for Retrieval	*No*	*No Limit*

Table 1: Retriever requirements of different frameworks for conversational retrieval. *Arch Type* stands for *Retriever Architecture Type*.

2.2 Conversational Retrieval

A few recent works (Dalton et al., 2020; Qu et al., 2020) collect retrieval datasets for *conversational search* tasks (Belkin et al., 1995; Solomon, 1997) where each dialogue context consists of a sequence of previous user questions only. Dalton et al. (2020) annotate 80 conversations for the TREC CAsT-19 task and Qu et al. (2020) derive their dataset based on QuAC (Choi et al., 2018) and propose to fine-tune a dual encoder retriever (Guu et al., 2020; Karpukhin et al., 2020). In contrast, the dialogue context in a CQA conversation, which is the focus of our work, consists of both user and agent turns. Each user query in a CQA conversation can be more challenging to de-contextualize as it depends on both previous user and agent turns.

Most existing conversational retrieval models require fine-tuning a retriever of a specific type (Table 1). Yu et al. (2021), Lin et al. (2021) and Kim and Kim (2022) attempt to fine-tune a dual encoder retriever (Xiong et al., 2021; Karpukhin et al., 2020) to handle conversational queries. Kumar and Callan (2020) propose a framework focusing on improving the passage re-ranker after the retrieval.

Query Rewriting (QR) In order to directly use an *off-the-shelf* retriever as we aim to do, conversational QR (Elgohary et al., 2019) has been applied in prior work (Vakulenko et al., 2021; Lin et al., 2020; Yu et al., 2020; Voskarides et al., 2020) to first convert a conversational query into a standalone one. Yu et al. (2020) propose a supervised QR model trained with human rewrites and weak QR supervisions specifically for conversational search tasks that are generated from additional search session resources. Lin et al. (2020) and

Vakulenko et al. (2021) also use human rewrites to train a supervised QR model based on pre-trained language models like T5 (Raffel et al., 2020) or GPT2 (Radford et al., 2019). Voskarides et al. (2020) use human rewrites to train a model that classifies whether each token in the dialogue context should be used to construct the query for retrieval. In contrast, we show the limitations of human rewrites used as QR supervision and design an RL-based QR model which can achieve better performance than supervised models even without human rewrites. Similar to our finding with details in Appendix A.4, Ishii et al. (2022) claim that using rewritten queries as an intermediate step does not necessarily outperform fine-tuning the end task model (e.g., retriever). However, we provide strong evidence in Section 4.3 to support the importance of QR in the *off-the-shelf* retriever setting.

2.3 RL for Text Generation

RL-based QR for Retrieval Nogueira and Cho (2017) and Adolphs et al. (2021) apply RL based on gold passage labels to do *non-conversational* query reformulation for retrieval. In contrast, to the best of our knowledge, we are the first to apply RL for rewriting *conversational* queries, and we only use weak retrieval supervision and an approximate retrieval metric for computational efficiency. Additionally, our model rewrites the query based on the dialogue context, while their models require multiple rounds of retrieval in order to reformulate a query, which can be time-consuming.

Other Applications Prior work also applies RL approaches to address text generation tasks like machine translation (Ranzato et al., 2016; Wu et al., 2016), text summarization (Paulus et al., 2018; Celikyilmaz et al., 2018) and image captioning (Rennie et al., 2017; Fisch et al., 2020) by training a model directly optimized towards generation quality metrics like BLEU, ROUGE or CIDEr. Buck et al. (2018) use RL to rewrite a *non-conversational* query for the task of question answering model.

3 Approach

Problem Definition We focus on the task of *query rewriting (QR)* for *conversational passage retrieval* in a CQA dialogue, with an *off-the-shelf* retriever. The task inputs include a dialogue context x consisting of a sequence of previous utterances $(u_1, u_2, \ldots, u_{n-1})$, the current user question u_n, a passage corpus P and an off-the-shelf retriever R.[1] R cannot be fine-tuned but returns a ranked list of top-k passages when given a query string and a passage corpus, and no other assumption about the model architecture of R can be made. The task aims to rewrite x into a query q such that R can take q as the input query to retrieve passages relevant to x from P. Specifically, a passage p is relevant to x if p provides enough information to answer u_n in the context of $(u_1, u_2, \ldots, u_{n-1})$.

In this section, we first describe a supervised QR model based on T5 (T5QR) (Lin et al., 2020) that applies a generic Seq2Seq training objective with QR labels (Section 3.1). Then we introduce our RL-based framework CONQRR (**Con**versational **Q**uery **R**ewriting for **R**etrieval) that trains a QR model to optimize towards retrieval and is adaptable to any given off-the-shelf retriever, with weak retrieval supervision (Section 3.2).

3.1 T5QR

T5 is an encoder-decoder model that is pre-trained on large textual corpora (Raffel et al., 2020). Following Lin et al. (2020), we fine-tune T5 to rewrite a conversational query with the input as the concatenation of utterances in the dialogue context x and the output as the human rewrite $\hat{q}$. Note that we concatenate the utterances in a reversed order such that u_n becomes the first one in the input string and any truncation impacts more distant context. Utterances are separated with a separator token "[SEP]" in the concatenated string. The model is then trained with a standard cross entropy (CE) loss to maximize the likelihood of generating $\hat{q}$, which is a self-contained version of the query u_n that can be interpreted without knowing previous turns $(u_1, u_2, \ldots, u_{n-1})$ in x.

3.2 CONQRR

QR models trained with a standard CE loss are agnostic to the retriever. In addition, human rewrites are not necessarily the most effective ones for passage retrieval (see Section 4.3 for an exploration).

This motivates us to design our RL-based framework CONQRR (Figure 2) that trains a QR model directly optimized for the retrieval performance and can be adapted to any given off-the-shelf retriever. Here, the RL environment includes the retriever model, dialogue context and passage candidates, in

[1]To mimic practical use cases, R is usually assumed to be general purpose retriever with standard search queries.