CTS-CRL: Context Turn Selection for Conversational Rewriting with Joint Learning

Yunpeng Zhang[1,2], Guan Bei[*,1,2], Sun Zhe[1,2], Junyu Li[1,2], Yongji Wang[1,2,3]
[1]Integration Innovation Center, Institute of Software, Chinese Academy of Sciences, Beijing, China
[2]University of Chinese Academy of Sciences, Beijing, China
[3]Key Laboratory of Systems Software, Institute of Software, Chinese Academy of Sciences, Beijing, China
zhangyunpeng231@mails.ucas.ac.cn, guanbei@iscas.ac.cn, sunzhe24@mails.ucas.ac.cn,
lijunyu25@mails.ucas.ac.cn, ywang@itechs.iscas.ac.cn
[*]Corresponding author

***Abstract*—Context Query Rewriting (CQR) is pivotal in conversational search for resolving context dependencies. However, existing methods typically employ a brute-force strategy that feeds the entire dialogue history into Large Language Models (LLMs). This indiscriminate approach introduces severe noise—particularly in multi-turn conversations with frequent topic shifts, which hinders both model robustness and training efficiency. To overcome these limitations, we propose CTS-CRL (Context Turn Selection for Conversational Rewriting with Joint Learning). Unlike previous methods, CTS-CRL incorporates a novel semantic-temporal context selection mechanism that dynamically filters irrelevant turns while preserving essential dependencies. Furthermore, we integrate this mechanism into a unified training framework combining Supervised Fine-Tuning (SFT) and Group Relative Policy Optimization (GRPO). Experimental results on two benchmarks demonstrate the superiority of our approach. Notably, CTS-CRL establishes a new state-of-the-art on the TopiOCQA dataset, significantly outperforming baselines in handling complex topic transitions, while maintaining high efficiency with fewer parameters.**

***Index Terms*—Context Selection, Conversational Query Rewriting, Language Models, Reinforcement Learning**

I. Introduction

Conversational Query Rewriting (CQR) has emerged as a critical component in conversational search systems [1]. Its primary goal is to reformulate ambiguous, context-dependent user utterances into self-contained queries that can be effectively processed by standard search engines [2]. In natural conversations, users frequently omit crucial information by implicitly referring to entities or events mentioned in previous turns. For instance, a user might ask ”What about the second one?” expecting the system to resolve the reference based on dialogue history. Current approaches to CQR predominantly adopt a ”concatenate-all” strategy, feeding the entire dialogue history or a fixed window of turns into sequence-to-sequence models. While effective for short, single-topic sessions, this approach faces severe challenges in real-world scenarios characterized by long durations and frequent topic shifts. When a conversation drifts from one subject to another (e.g., from ”sports cars” to ”traffic laws”), historical turns associated with the previous topic become irrelevant noise. Indiscriminately including these turns not only dilutes the semantic focus of the LLM but also significantly degrades training efficiency, as the model must waste capacity learning to ignore substantial amounts of distractor information.

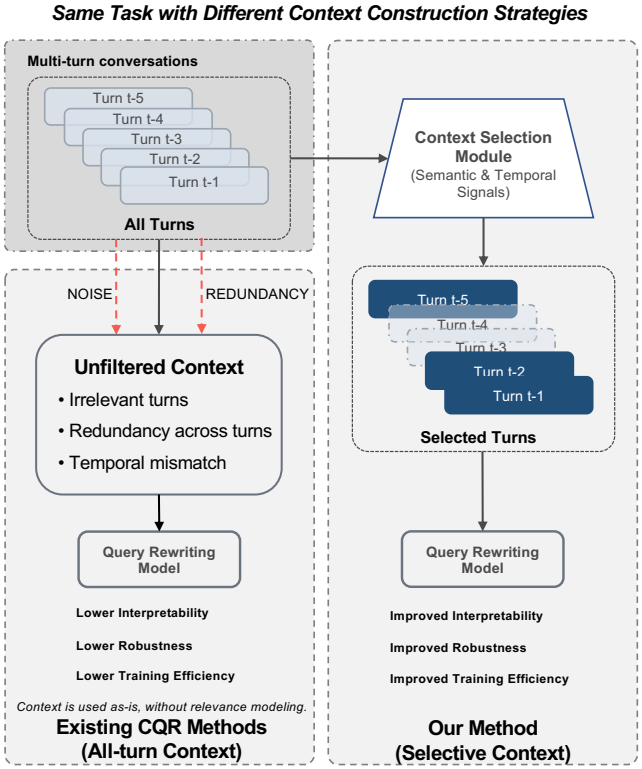

Fig. 1. Existing CQR methods use all historical turns as context, while our method selectively constructs relevant context via a dedicated selection module.

To address these challenges, we propose CTS-CRL (Context Turn Selection for Conversational Rewriting with Joint Learning). Unlike previous methods that rely on passive context concatenation, CTS-CRL introduces an active Semantic-Temporal Context Selection mechanism. This approach dynamically filters out irrelevant turns while preserving essential dependencies, effectively mitigating the noise interference caused by topic shifts. Furthermore, we integrate this mechanism into a unified training framework that combines Supervised

Fine-Tuning (SFT) and Group Relative Policy Optimization (GRPO) [3] to align context reasoning with retrieval objectives.

The key contributions of our work are summarized as follows:

- We propose **CTS-CRL**, a novel framework that shifts the CQR paradigm from indiscriminate context concatenation to precise context selection. By explicitly modeling context relevance, our approach effectively resolves the conflict between capturing long-range dependencies and avoiding noise accumulation in multi-turn conversations with frequent topic shifts.
- We design a **semantic-temporal collaborative mechanism** to implement this selection strategy. This dual-stream module synergizes a recency-based hard constraint (temporal) with a retrieval-enhanced filter (semantic) to accurately identify informative turns.
- We present a unified training paradigm that combines "Filter-Generate-Verify" Supervised Fine-Tuning (SFT) with Group Relative Policy Optimization (GRPO). This strategy enables the model to jointly optimize context selection and query rewriting tailored for downstream retrieval performance.

II. RELATED WORK

Conversational Query Rewriting (CQR) is essential for bridging the gap between context-dependent user utterances and stateless retrieval engines. The primary goal is to reformulate ambiguous queries into self-contained natural language expressions that preserve the original semantic intent while resolving ellipses and coreferences within the dialogue history. Existing approaches can be broadly categorized into three paradigms: supervised learning on compact models, LLM-based generation, and alignment tuning via Reinforcement Learning.

Supervised Learning for Compact Models. Early research predominantly treated CQR as a sequence-to-sequence generation task trained on human-annotated datasets. Initial efforts, such as Voskarides et al., modeled the resolution process as a term-level classification task using distant supervision. Subsequently, Lin et al. demonstrated the effectiveness of applying pretrained encoder-decoder architectures (e.g., T5) to conversational query reformulation. To further improve retrieval utility, Wu et al. proposed CONQRR, a framework that selectively incorporates dialogue history via reinforcement learning to optimize retrieval effectiveness [4]. More recently, researchers have focused on distilling capabilities from LLMs into smaller, more efficient rewriters. For instance, Lai et al. introduced ADACQR, utilizing LLM-generated candidates and contrastive loss to align rewriters with both sparse and dense retrieval signals [5]. Similarly, Mo et al. proposed CHIQ, utilizing open-source LLMs to enhance ambiguity resolution through hybrid strategies including fine-tuning and fusion [6].

Large Language Models for Rewriting and Expansion. With the emergence of Large Language Models (LLMs), prompt engineering has become a dominant paradigm for performing CQR without parameter updates. Beyond simple coreference resolution, recent works utilize the world knowledge of LLMs to enrich queries. Ye et al. proposed an informative query rewriting framework, prompting LLMs to not only resolve ambiguities but also explicitly supplement missing keywords necessary for retrieval [7]. Similarly, in the broader context of retrieval enhancement, Wang et al. introduced Query2Doc, which prompts LLMs to generate pseudo-documents that expand the query's semantic coverage [8]. Other works, such as LLM4CS by Mao et al., employ prompting to generate multiple rewrite candidates and hypothetical responses, aggregating them to robustly capture search intent [9]. While effective, these large-scale prompting approaches often suffer from high latency and token costs.

Alignment and Reinforcement Learning. A growing trend involves fine-tuning models to better align rewriting outputs with downstream retriever preferences rather than solely imitating human references.WANG et al. propose a reinforcement-learning-based query rewriting framework for conversational RAG that leverages multi-aspect feedback from retrieval and generation to provide dense and stable supervision for rewrite optimization [10]. Yoon et al. introduced RETPO, which utilizes Direct Preference Optimization (DPO) to align open-source LLMs with retriever feedback using a constructed dataset [11]. In a similar vein, Zhang et al. proposed AdaQR, adapting query rewriters with limited annotations by modeling retriever preferences through the marginal probabilities of retrieved answers [12]. Recent work has further explored reinforcing these models to better leverage complex dialogue contexts [13].

Limitations and Our Approach. Despite these advancements, effectively modeling dialogue history remains a persistent challenge. The prevailing "concatenate-all" strategy—simply feeding the entire history into the model—often introduces semantic noise and dilutes relevant signals, particularly in long conversations with topic shifts. Conversely, heuristic truncation risks losing critical long-range dependencies. While some recent studies explore implicit token reweighting, the explicit and quantitative selection of relevant context turns prior to rewriting has received limited attention. In contrast to existing methods, we propose a novel **Context Turn Selection (CTS)** mechanism that quantitatively filters noise before generation. By integrating this module into a joint Supervised Fine-Tuning (SFT) and Group Relative Policy Optimization (GRPO) framework, our method explicitly learns to balance context sufficiency with noise reduction, achieving superior robustness and retrieval performance.

III. METHODOLOGY

The overall architecture of our proposed framework is illustrated in Fig. 2. The framework consists of three integral components: a Relevance-Aware Context Selector, a Retrieval-Oriented SFT stage, and a Context-Aware Reinforcement Learning stage via GRPO. In this section, we first formulate the Conversational Query Rewriting (CQR) problem and then detail each component.

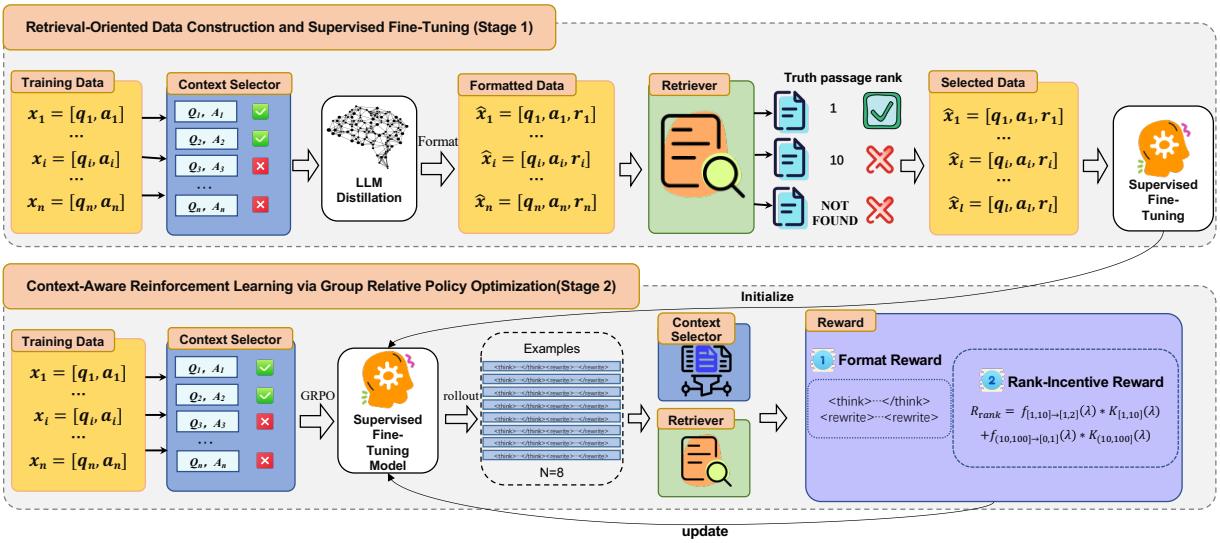

Fig. 2. Illustration of the proposed context-aware CQR framework. Given an input query with dialogue history, a Semantic-Temporal Context Selection module first identifies the most relevant turns. In Stage 1, the selected context is used to distill high-quality query rewriting (CQR) samples from a stronger model, and only those rewrites that retrieve the ground-truth at top-1 are kept to supervise fine-tuning of the base CQR model. In Stage 2, the fine-tuned model is further optimized with Group Relative Policy Optimization (GRPO). At each rollout, new samples filtered by the context selector are rewritten and retrieved to update the policy. The reinforcement learning reward combines a format quality term and a rank-based retrieval incentive, driving the model to generate contextually relevant rewrites that improve retrieval performance.

A. Problem Formulation

Formally, let $\mathcal{D} = \{d_1, d_2, ..., d_M\}$ denote a large-scale corpus containing M passages. In a conversational search scenario, we consider a dialogue session consisting of a sequence of history turns $\mathcal{H} = \{(q_1, a_1), ..., (q_{t-1}, a_{t-1})\}$ and a current user query q_t, where q_i and a_i represent the query and answer at the i-th turn, respectively. Due to the ellipsis and coreference phenomena in conversations, the current query q_t is often ambiguous and context-dependent, making it unsuitable for direct retrieval.

The goal of CQR is to learn a rewriting function $f_\theta : (\mathcal{H}, q_t) \rightarrow \hat{q}$, which maps the conversational context and current query to a self-contained, de-contextualized query $\hat{q}$. Ideally, the rewritten query $\hat{q}$ should preserve the original user intent while explicitly resolving ambiguities. From the perspective of information retrieval, the ultimate objective is to maximize the probability of retrieving the ground-truth passage d^+ from $\mathcal{D}$ using a retriever $\mathcal{R}(\cdot)$. This can be formulated as maximizing the ranking score:

$$\hat{q}^* = \arg\max_{\hat{q}} \text{Score}(\mathcal{R}(\hat{q}), d^+) \tag{1}$$

where Score$(\cdot)$ denotes the relevance metric. Unlike traditional generation tasks that focus on linguistic fluency, our problem highlights the *retrievability* of the generated query.

B. Semantic-Temporal Context Selection

To effectively address the noise accumulation problem caused by topic shifts, CTS-CRL incorporates a specialized Semantic-Temporal Context Selection mechanism. Instead of feeding the entire history H into the rewriter, this mechanism dynamically selects a subset of informative turns. As illustrated in Fig. 3 , the selection process operates through two collaborative streams:

1) Temporal Stream (Recency-Based Hard Constraint): Empirical studies suggest that a majority of coreference resolutions and ellipsis recoveries depend on the most recent dialogue turns. To guarantee the capture of these immediate dependencies regardless of semantic shifts, we establish a *Temporal Stream*. This stream enforces a hard constraint to retain the most recent N turns of history, denoted as $\mathcal{H}_{temp}$:

$$\mathcal{H}_{temp} = \{(q_{t-k}, a_{t-k}) \mid k = 1, \ldots, N\} \tag{2}$$

where N is a hyperparameter governing the size of the recency buffer.

2) Semantic Stream (Two-Stage Retrieval): To address long-range dependencies where the referenced object appears in distant history, we design a *Semantic Stream*. Directly matching q_t with history is often ineffective due to the brevity and ambiguity of the current query. Therefore, we adopt a coarse-to-fine retrieval pipeline:

i) Context-Aware Query Expansion: We first construct an enriched search query q_{search} to mitigate the semantic gap. By concatenating the current session topic T, the immediate previous response a_{t-1}, and the current query q_t, we obtain:

$$q_{search} = \text{Concat}(T, a_{t-1}, q_t) \tag{3}$$

Including a_{t-1} helps bridge the semantic connection when q_t is a short follow-up question.

ii) Bi-Encoder Coarse Retrieval: We employ a Bi-Encoder to map both the search query q_{search} and each historical

turn $h_i = (q_i, a_i) \in \mathcal{H}$ into a shared embedding space. The relevance score is calculated via cosine similarity:

$$S_{bi}(q_{search}, h_i) = \cos(\mathbf{E}(q_{search}), \mathbf{E}(h_i)) \tag{4}$$

We retain a candidate set $\mathcal{C}_{cand}$ containing the top-K turns with the highest S_{bi} scores.

iii) Cross-Encoder Reranking: Since Bi-Encoders may miss subtle semantic interactions, we apply a Cross-Encoder for fine-grained reranking. The original query q_t and each candidate $c_i \in \mathcal{C}_{cand}$ are concatenated and fed into the Cross-Encoder to compute a precise relevance score:

$$Score(q_t, c_i) = \sigma(\text{CrossEnc}(q_t, c_i)) \tag{5}$$

where σ is the sigmoid function. We filter these candidates using a relevance threshold τ, resulting in the semantic context set $\mathcal{H}_{sem} = \{c_j \mid Score(q_t, c_i) > \tau\}$.

3) Temporal-Semantic Fusion: Large Language Models are sensitive to the temporal order of inputs. Simply concatenating the retrieved contexts may disrupt the logical flow. Therefore, we merge the outputs of both streams and perform a temporal re-ordering operation:

$$\mathcal{H}_{final} = \text{Sort}_{time}(\mathcal{H}_{temp} \cup \mathcal{H}_{sem}) \tag{6}$$

The strictly time-ordered $\mathcal{H}_{final}$ is then used as the clean context for the subsequent rewriting generation.

C. Stage 1: Retrieval-Oriented Data Construction and Supervised Fine-Tuning

This stage aims to initialize the model with the ability to perform reasoning-aware rewriting. We employ a "Filter-Generate-Verify" pipeline to construct a high-quality dataset, ensuring that the training samples are not only linguistically fluent but also effective for retrieval tasks.

1) Data Synthesis via Teacher Model: Based on the refined context $\mathcal{H}_{final}$ obtained from the previous section, we utilize an advanced Large Language Model (e.g., Gemini-3-Pro) as the teacher. We construct a specific prompt that instructs the teacher model to first output a "chain-of-thought" analysis (wrapped in `<think>` tags) to resolve ambiguities in the current query q_t, and then generate the final rewritten query $\hat{q}$. This process creates a preliminary dataset containing the context, the reasoning path, and the candidate rewrite.

2) Retrieval-Consistency Verification: A key challenge in CQR is that a rewrite may look correct to humans but fail to retrieve the correct document. To address this, we filter the synthesized data using the specific retriever targeted in our experiments (e.g., BM25 for sparse retrieval or ANCE for dense retrieval). For each candidate rewrite $\hat{q}$, we execute a retrieval query over the corpus. We enforce a strict Top-1 Hit constraint: a sample is added to the final training set $\mathcal{D}_{sft}$ if and only if the ground-truth passage is ranked at the first position by the retriever. This ensures that the student model learns from successful retrieval behaviors tailored to the specific search engine. **Supervised Fine-Tuning** We fine-tune a smaller student model on the verified dataset $\mathcal{D}_{sft}$. The model is trained to generate both the reasoning tokens and the rewrite tokens autoregressively. We minimize the standard cross-entropy loss:

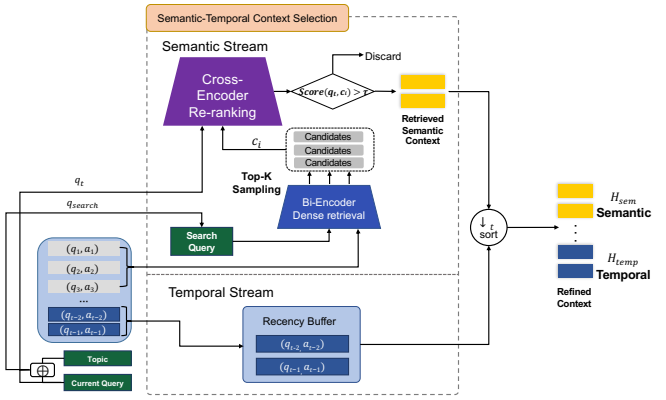

Fig. 3. The architecture of the Semantic-Temporal Context Selection mechanism within CTS-CRL. The model employs a dual-stream strategy to capture both long-range semantic dependencies (Semantic Stream) and immediate coreference cues (Temporal Stream).

$$\mathcal{L}_{SFT} = -\sum_{t} \log P(y_t \mid \mathcal{H}_{final}, y_{<t}) \tag{7}$$

where y represents the concatenated sequence of the thought chain and the rewritten query. This supervised stage provides a stable starting point (warm-start) for the subsequent reinforcement learning stage.

D. Stage 2: Context-Aware Reinforcement Learning via Group Relative Policy Optimization

While SFT initializes the model with basic rewriting capabilities, it is trained on static data and does not directly interact with the retrieval environment. To further align the model's generation with the ultimate retrieval objective, we employ Group Relative Policy Optimization (GRPO) [3]. This stage optimizes the policy π_θ to maximize the expected reward derived from retrieval feedback.

1) GRPO Algorithm: Unlike traditional PPO which relies on a value function (Critic) to estimate advantages, GRPO estimates the baseline from a group of outputs sampled from the same input. This significantly reduces memory usage and training instability. For each query input, we sample a group of G outputs $\{y_1, y_2, \ldots, y_G\}$ from the old policy $\pi_{\theta_{old}}$. The advantage for each output is calculated by normalizing its reward against the group's mean and standard deviation. This encourages the model to generate rewrites that perform relatively better than the group average, progressively refining the reasoning and rewriting quality.

2) Reward Function: The design of the reward signal is crucial for guiding the reinforcement learning process. Following the methodology proposed in [13] , we adopt a composite reward function consisting of a format constraint and a ranking incentive.

i) Format Reward (R_{fmt}): To maintain the structural integrity of the generation, we impose a format reward. The

model receives a positive reward if and only if the output strictly follows the fixed structure, containing both the reasoning process (wrapped in <think>) and the final rewrite (wrapped in <rewrite>). This ensures the model generates valid inputs for the retriever.

ii) Rank-Incentive Reward (R_{rank}): To directly optimize retrieval performance, we compute a reward based on the rank of the ground-truth passage. As defined in [13], the ranking reward is formulated as:

$$R_{rank}(\lambda) = f_{[1,10]\to[1,2]}(\lambda) \cdot K_{[1,10]}(\lambda) + f_{(10,100]\to[0,1]}(\lambda) \cdot K_{(10,100]}(\lambda) \tag{8}$$

where $f : A \to B$ denotes the mapping function from interval A to interval B, and $K_A(\lambda)$ is the indicator function, which equals 1 if λ belongs to set A and 0 otherwise. This design aims to guide the model in generating rewrites that are more conducive to recalling relevant documents, while preserving the core semantics of the query.

The final total reward is the weighted sum: $R_{total} = R_{rank}(\lambda) \cdot R_{fmt}(\gamma = 1) + \Phi \cdot R_{fmt}(\gamma = 0)$, where R_{rank} represents the retrieval effectiveness reward and γ is the format reward. Note that Φ is constantly set to -0.1, and $\gamma \in \{0, 1\}$ indicates whether the formatting rules are violated ($\gamma = 0$) or satisfied ($\gamma = 1$).

IV. EXPERIMENTS

A. Datasets

We evaluate our proposed framework on two widely-used open-domain Conversational Question Answering (CQA) benchmarks: **QReCC** [14] and **TopiOCQA** [15]. The detailed statistics of these datasets are summarized in Table I.

1) QReCC (Question Rewriting in Conversational Context): is a large-scale dataset constructed from TREC CAsT, QuAC, and Google Natural Questions. It provides context-independent query rewrites as ground truth, which enables the supervised training of our rewriting model in Stage 1. The retrieval corpus consists of approximately 54 million passages from Web pages, requiring the model to handle open-domain search challenges.

2) TopiOCQA (Topic-Switching Open-domain CQA): features multi-turn conversations with frequent topic transitions. Unlike QReCC, where a conversation is often grounded in a single document, TopiOCQA requires the system to retrieve evidence from the entire English Wikipedia (chunked into $\sim$25 million passages). This dataset serves as a rigorous testbed for our *Context Selector*, assessing its ability to filter irrelevant context during topic shifts.

B. Implementation Details

1) Retrieval Setup: To verify robustness across retrieval paradigms, we configure two environments: sparse retrieval using **BM25** (via Pyserini) and dense retrieval using **ANCE** [16]. For the Context Selector, we employ `all-MiniLM-L6-v2` [17] as the Bi-Encoder and `ms-macro-MiniLM-L6-v2` as the Cross-Encoder.

TABLE I
STATISTICS OF THE QRECC AND TOPIOCQA DATASETS.

Dataset	Dialogues	QA Pairs	Avg. Turns	Corpus Size
QReCC	13,809	81,132	5.9	54M (Web)
TopiOCQA	3,920	50,264	12.8	25M (Wiki)

2) Training Setup: We adopt Qwen3-4B [18] as the backbone model. Training is implemented using **LLaMA-Factory** [19] for SFT and **Verl** [20] for GRPO. In the *SFT stage*, we apply LoRA ($r = 16$) and train for 6 epochs with a learning rate of 5e-5 and a batch size of 16. In the GRPO stage, we set the group rollout size $G = 8$ and increase the effective batch size to 64 to stabilize gradient estimation. The max sequence length is set to 2048 to accommodate multi-turn context history.

C. Metrics

Following previous CQR studies, we evaluate the retrieval performance of the rewritten queries. Since our experiments involve both sparse and dense retrieval scenarios, we adopt standard Information Retrieval (IR) metrics: **Recall@K** (R@10, R@100) to measure the proportion of relevant passages retrieved within the top-K results, **MRR@3** (Mean Reciprocal Rank) to evaluate the ranking quality of the first correct answer, and **NDCG@3** (Normalized Discounted Cumulative Gain) to account for graded relevance (if applicable) or position bias at shallow depths [21]. The evaluation is conducted using the Pyserini toolkit [22] to ensure reproducibility.

D. Baselines

We consider three categories of baselines in our experiments to comprehensively evaluate the performance of our proposed framework.

The first category comprises Basic Baselines that do not involve any query rewriting optimization. This includes Human Rewrite (Oracle), which uses manually annotated rewrites to represent the performance upper bound; Raw Query, which directly uses the current user query for retrieval without modification; and Untrained LLMs, which utilize the base capability of large language models without task-specific fine-tuning.

The second category consists of Small-scale Fine-tuned Models that adapt lightweight architectures for the rewriting task. Representative methods include QuReTec [23], which classifies term relevance for query expansion; T5QR [24], which fine-tunes T5 for sequence-to-sequence rewriting; CONQRR [4], which employs reinforcement learning for query reformulation; ConvGQR [25], which combines generation and retrieval supervision; CHIQ-Fusion [6], which uses to enable effective query rewriting; REPTO [11], which uses DPO for query rewriting; and IterCQR [26], which iteratively refines the query. The third category is Training-free Prompt-based Methods, leveraging prompt engineering with powerful LLMs (e.g., ChatGPT [27]) for query rewriting without parameter

TABLE II
MAIN RESULTS ON THE QRECC AND TOPIOCQA DATASETS UNDER THE **SPARSE RETRIEVAL (BM25)** AND **DENSE RETRIEVAL (ANCE)**SETTING. THE BEST RESULTS ARE HIGHLIGHTED IN **BOLD**.

Type	Method	TopiOCQA				QReCC			
		MRR@3	NDCG@3	R@10	R@100	MRR@3	NDCG@3	R@10	R@100
Sparse	*Basic Baselines*								
	Human Rewrite	-	-	-	-	39.2	38.7	63.8	**98.5**
	Raw Query	2.1	1.8	4.0	9.2	6.5	5.5	11.1	21.5
	Qwen3-4B	10.4	9.5	23.3	42.1	31.3	30.9	44.7	67.3
	DS-R1-Distill-Qwen-7B	9.4	8.7	19.4	41.1	29.8	28.5	45.4	66.5
	Fine-tuned Rewriters								
	QuReTec	8.5	7.3	16.0	-	34.0	30.5	55.5	-
	T5QR	11.3	9.8	22.1	44.7	33.4	30.2	53.8	86.1
	CONQRR	-	-	-	-	38.3	-	60.1	88.9
	ConvGQR	12.4	10.7	23.8	45.6	44.1	41.0	64.4	88.0
	IterCQR	16.5	14.9	29.3	54.1	46.7	44.1	64.4	85.5
	CHIQ-Fusion	25.6	23.5	44.7	-	**54.3**	**51.9**	**78.5**	-
	LLM-based Methods								
	LLM4CS	18.9	17.7	33.7	-	47.8	45.0	69.1	-
	InfoCQR	-	-	-	-	48.9	46.3	66.4	-
	REPTO	28.3	26.5	48.3	73.1	50.0	47.3	69.5	89.5
	CTS-CRL (Ours)	**29.5**	**28.4**	**50.4**	**74.6**	51.7	47.9	76.7	89.4
Dense	*Basic Baselines*								
	Human Rewrite	-	-	-	-	38.4	35.6	58.6	78.1
	Raw Query	4.1	3.8	7.5	13.8	10.2	9.3	15.7	22.7
	Qwen3-4B	19.2	18.3	33.0	46.7	29.3	27.2	44.2	59.0
	DS-R1-Distill-Qwen-7B	21.0	19.9	36.4	53.4	28.4	25.9	43.9	59.0
	Fine-tuned Rewriters								
	QuReTec	11.2	10.5	20.2	-	35.0	32.6	55.0	-
	T5QR	23.0	22.2	37.6	54.4	34.5	31.8	53.1	72.8
	CONQRR	-	-	-	-	41.8	-	65.1	84.7
	ConvGQR	25.6	24.3	41.8	58.8	42.0	39.1	63.5	81.8
	IterCQR	26.3	25.1	42.6	62.0	42.9	40.2	65.5	84.1
	CHIQ-Fusion	38.0	37.0	61.6	-	**47.2**	44.2	**70.7**	-
	LLM-based Methods								
	LLM4CS	27.7	26.7	43.3	-	44.8	42.1	66.4	-
	InfoCQR	-	-	-	-	43.9	41.3	65.6	-
	REPTO	30.0	28.9	49.6	68.7	44.0	41.1	66.7	84.6
	CTS-CRL (Ours)	**45.8**	**43.7**	**65.9**	**78.4**	46.1	**45.5**	67.7	**85.2**

updates. This category includes LLM4CS [9], which utilizes prompting strategies for conversational search, and InfoCQR [7], which incorporates information-theoretic incentives into the prompting process.

E. Results and Discussion

1) Overall Performance: The experimental results on QReCC and TopiOCQA under sparse (BM25) and dense (ANCE) retrieval settings are presented in Table II , respectively. Overall, our proposed framework, CTS-CRL, achieves consistent improvements over baselines across most metrics. Notably, in the challenging TopiOCQA benchmark which involves frequent topic transitions, our method demonstrates a significant advantage, establishing a new state-of-the-art performance.

2) Performance under Sparse Retrieval: Under the BM25 setting, CTS-CRL significantly outperforms the base model and fine-tuned baselines. On TopiOCQA, our method achieves an MRR@3 of 29.5, surpassing the backbone model Qwen3-4B (10.4) by a substantial margin. This confirms that the "Raw Query" and standard LLM generation are insufficient for handling complex context dependencies. Compared to strong fine-tuned methods like CHIQ-Fusion and ConvGQR, CTS-CRL yields an improvement of 4.0% and 18.4% in MRR@3 on TopiOCQA, respectively. On QReCC, our method also achieves a competitive NDCG@3 of 47.9, outperforming most baselines while remaining slightly below CHIQ-Fusion. One possible reason is that, in QReCC, most historical context remains strongly relevant to the current query. As a result, the proposed context selection mechanism may occasionally discard a small amount of useful information in this setting. Overall, these results indicate that the proposed two-stage training strategy effectively instills conversational rewriting capability into the local model, enabling it to generate more precise search queries than general-purpose LLMs.

3) Performance under Dense Retrieval: When equipped with the ANCE retriever, the performance of all models generally improves. CTS-CRL continues to dominate on TopiOCQA, achieving an MRR@3 of 45.8 and NDCG@3 of

TABLE III
ABLATION STUDY OF THE PROPOSED CTS-CRL METHOD ON **TOPIOCQA** AND **QRECC DATASETS**. THE BEST PERFORMANCE IN EACH CATEGORY IS HIGHLIGHTED IN **BOLD**.

Type	Method	TopiOCQA				QReCC			
		MRR@3	NDCG@3	R@10	R@100	MRR@3	NDCG@3	R@10	R@100
Sparse	CTS-CRL	**29.5**	**28.4**	**50.4**	**74.6**	**51.7**	**47.9**	**76.7**	**89.4**
	w/o Context Selection	24.5	23.7	44.7	62.6	49.6	47.0	75.5	88.7
	w/o RL-based Training	20.7	21.6	35.9	56.4	36.4	37.4	57.9	70.2
Dense	CTS-CRL	**45.8**	**43.7**	**65.9**	**78.4**	**46.1**	**45.5**	**67.7**	**85.2**
	w/o Context Selection	37.2	35.8	57.3	71.9	44.3	42.1	63.7	79.3
	w/o RL-based Training	27.4	26.8	45.7	60.9	39.7	38.2	53.6	74.6

43.7. Crucially, our method outperforms the strongest baseline (CHIQ-Fusion) by over 2 points in MRR@3 on this dataset. We attribute this success to the proposed Relevance-Aware Context Selection Mechanism. Baselines like CHIQ-Fusion typically concatenate full context or use heuristic summarization, which introduces noise during topic shifts. In contrast, our dual-stream selection module effectively filters out irrelevant history while preserving semantic dependencies, allowing the dense retriever to focus on the current search intent.

4) Efficiency and Robustness: It is worth noting that while CHIQ-Fusion achieves marginally higher results on QReCC (MRR@3 47.2 vs. ours 46.1), the performance gap is minimal. This result effectively validates the precision of our context filtering mechanism. Despite explicitly reducing the context input by selecting specific turns, our model maintains competitive retrieval performance. This demonstrates that our method successfully identifies and preserves the most relevant historical dependencies, ensuring that critical information is not lost during the filtering process.

F. Ablation Study

To validate the rationality of our model design,we performed ablation studies on the TopiOCQA dataset to assess the impact of the Semantic-Temporal Context Selection mechanism and RL-based Training components.

Specifically, we compared the full CTS-CRL model with the following variants:

- without the Semantic-Temporal Context Selection mechanism: This variant removes Context Selection mechanism, retaining all the historical turns.
- without RL-based Training: This variant eliminates the reinforcement learning training and retains only the context filtering and the SFT training stage.

The experimental results are summarized in TableIII. It can be observed that the full CTS-CRL model consistently achieves the best performance. Observations from Table III reveal that both modules are indispensable for optimal performance. Specifically, removing the Semantic-Temporal Context Selection mechanism leads to a noticeable performance decline across all metrics; for instance, the R@10 on TopiOCQA (Dense) drops from 65.9 to 57.3. This substantiates the mechanism's efficacy in filtering historical noise, especially in complex topic-switching scenarios. More prominently, the absence of RL-based Training triggers the most significant degradation (e.g., MRR@3 plunging from 45.8 to 27.4 in the Dense setting), underscoring that autonomous exploration is the primary driver for generating high-quality, retrieval-oriented query rewrites.

In conclusion, each component plays an indispensable role, and their synergistic combination enables CTS-CRL to handle complex conversational contexts effectively.

V. CONCLUSION

In this paper, we present CTS-CRL, a novel and efficient framework for conversational query rewriting. Our work addresses the critical challenges of context ambiguity and topic shifts in multi-turn conversational search through two key technical contributions.

First, we introduced a *Semantic-Temporal Context Selection mechanism*. By synergizing a temporal hard-constraint stream with a semantic retrieval stream, this module effectively identifies and retains the most valuable historical turns while filtering out noise. This is particularly crucial for handling complex conversations with frequent topic transitions, as evidenced by our significant performance gains on the TopiOCQA benchmark. Second, we established a retrieval-oriented training paradigm that combines *Context-Aware SFT* with *Group Relative Policy Optimization (GRPO)*. Unlike previous methods that rely solely on linguistic supervision, our "Filter-Generate-Verify" data synthesis pipeline ensures high data quality, while the GRPO stage directly aligns the model's generation with the ranking metrics of the search engine. This allows our lightweight model to learn explicit reasoning paths and generate retrieval-friendly queries.

Extensive experiments on QReCC and TopiOCQA demonstrate that CTS-CRL significantly outperforms strong baselines, including much larger LLMs and specialized rewriting models. Our findings confirm that coupling precise context refinement with reinforcement learning is a promising direction for building effective and efficient conversational search systems. In future work, we plan to extend this framework to end-to-end conversational retrieval models and explore its effectiveness in domain-specific applications.

ACKNOWLEDGMENT

This research was supported by the National Key Research and Development Program of China, under Grant No. 2022ZD0120201 - "Unified Representation and Knowledge Graph Construction for Science Popularization Resources".

REFERENCES

[1] A. Elgohary, D. Peskov, and J. Boyd-Graber, "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)*, K. Inui, J. Jiang, V. Ng, and X. Wan, Eds. Hong Kong, China: Association for Computational Linguistics, Nov. 2019, pp. 5918–5924.

[2] S. Vakulenko, S. Longpre, Z. Tu, and R. Anantha, "Question rewriting for conversational question answering," *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, 2020.

[3] Z. Shao, P. Wang, Q. Zhu, R. Xu, J.-M. Song, M. Zhang, Y. K. Li, Y. Wu, and D. Guo, "DeepSeekMath: Pushing the limits of mathematical reasoning in open language models," *arXiv preprint arXiv:2402.03300*, 2024.

[4] Z. Wu, Y. Luan, H. Rashkin, D. Reitter, and G. S. Tomar, "Conqrr: Conversational query rewriting for retrieval with reinforcement learning," in *Conference on Empirical Methods in Natural Language Processing*, 2021.

[5] Y. Lai, J. Wu, C. Zhang, H. Sun, and D. Zhou, "AdaCQR: Enhancing query reformulation for conversational search via sparse and dense retrieval alignment," in *Proceedings of the 31st International Conference on Computational Linguistics*, O. Rambow, L. Wanner, M. Apidianaki, H. Al-Khalifa, B. D. Eugenio, and S. Schockaert, Eds. Abu Dhabi, UAE: Association for Computational Linguistics, Jan. 2025, pp. 7698–7720.

[6] F. Mo, A. Ghaddar, K. Mao, M. Rezagholizadeh, B. Chen, Q. Liu, and J.-Y. Nie, "CHIQ: Contextual history enhancement for improving query rewriting in conversational search," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, Eds. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 2253–2268.

[7] F. Ye, M. Fang, S. Li, and E. Yilmaz, "Enhancing conversational search: Large language model-aided informative query rewriting," in *Findings of the Association for Computational Linguistics: EMNLP 2023*, H. Bouamor, J. Pino, and K. Bali, Eds. Singapore: Association for Computational Linguistics, Dec. 2023, pp. 5985–6006.

[8] L. Wang, N. Yang, and F. Wei, "Query2doc: Query expansion with large language models," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 2023, pp. 9414–9423.

[9] K. Mao, Z. Dou, F. Mo, J. Hou, H. Chen, and H. Qian, "Large language models know your contextual search intent: A prompting framework for conversational search," in *Findings of the Association for Computational Linguistics: EMNLP 2023*, H. Bouamor, J. Pino, and K. Bali, Eds. Singapore: Association for Computational Linguistics, Dec. 2023, pp. 1211–1225.

[10] Y. Wang, H. Zhang, L. Pang, B. Guo, H. Zheng, and Z. Zheng, "Maferw: query rewriting with multi-aspect feedbacks for retrieval-augmented large language models," in *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence*, ser. AAAI'25/IAAI'25/EAAI'25. AAAI Press, 2025.

[11] C. Yoon, G. Kim, B. Jeon, S. Kim, Y. Jo, and J. Kang, "Ask optimal questions: Aligning large language models with retriever's preference in conversation," in *Findings of the Association for Computational Linguistics: NAACL 2025*, L. Chiruzzo, A. Ritter, and L. Wang, Eds. Albuquerque, New Mexico: Association for Computational Linguistics, Apr. 2025, pp. 5899–5921.

[12] T. Zhang, K. Li, H. Luo, X. Wu, J. R. Glass, and H. M. Meng, "Adaptive query rewriting: Aligning rewriters through marginal probability of conversational answers," in *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, Y. Al-Onaizan, M. Bansal, and Y.-N. Chen, Eds. Miami, Florida, USA: Association for Computational Linguistics, Nov. 2024, pp. 13 444–13 461.

[13] C. Zhu, S. Wang, R. Feng, K. Song, and X. Qiu, "ConvSearch-r1: Enhancing query reformulation for conversational search with reasoning via reinforcement learning," in *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, C. Christodoulopoulos, T. Chakraborty, C. Rose, and V. Peng, Eds. Suzhou, China: Association for Computational Linguistics, Nov. 2025, pp. 26 547–26 564.

[14] R. Anantha, S. Vakulenko, Z. Tu, S. Longpre, S. Pulman, and S. Chappidi, "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*, K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, and Y. Zhou, Eds. Association for Computational Linguistics, Jun. 2021, pp. 520–534.

[15] V. Adlakha, S. Dhuliawala, K. Suleman, H. de Vries, and S. Reddy, "Topiocqa: Open-domain conversational question answering with topic switching," *Transactions of the Association for Computational Linguistics*, vol. 10, pp. 468–483, 2021.

[16] L. Xiong, C. Xiong, Y. Li, K.-F. Tang, J. Liu, P. Bennett, J. Ahmed, and A. Overwijk, "Approximate nearest neighbor negative contrastive learning for dense text retrieval," *arXiv preprint arXiv:2007.00808*, 2020.

[17] W. Wang, F. Wei, L. Dong, H. Bao, N. Yang, and M. Zhou, "Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 5776–5788.

[18] A. Yang, A. Li, B. Yang, B. Zhang, B. Hui, B. Zheng, and et al., "Qwen3 technical report," arXiv preprint arXiv:2505.09388, 2025.

[19] Y. Zheng, R. Zhang, J. Zhang, Y. Ye, and Z. Luo, "LlamaFactory: Unified efficient fine-tuning of 100+ language models," in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, Y. Cao, Y. Feng, and D. Xiong, Eds. Bangkok, Thailand: Association for Computational Linguistics, Aug. 2024, pp. 400–410.

[20] G. Sheng, C. Zhang, Z. Ye, X. Wu, W. Zhang, R. Zhang, Y. Peng, H. Lin, and C. Wu, "Hybridflow: A flexible and efficient rlhf framework," in *Proceedings of the Twentieth European Conference on Computer Systems*, ser. EuroSys '25. ACM, Mar. 2025, p. 1279–1297.

[21] P. Jiang, J. Lin, L. Cao, R. Tian, S. Kang, Z. Wang, J. Sun, and J. Han, "Deepretrieval: Hacking real search engines and retrievers with large language models via reinforcement learning," *arXiv preprint arXiv:2503.00223*, 2025.

[22] J. Lin, X. Ma, S. Lin, J. Yang, R. Pradeep, and R. Nogueira, "Pyserini: A python toolkit for reproducible information retrieval research with sparse and dense representations," in *Proceedings of the 44th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2021, pp. 2356–2362.

[23] N. Voskarides, D. Li, P. Ren, E. Kanoulas, and M. de Rijke, "Query resolution for conversational search with limited supervision," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020, p. 921–930.

[24] S.-C. Lin, J.-H. Yang, R. Nogueira, M.-F. Tsai, C.-J. Wang, and J. J. Lin, "Conversational question reformulation via sequence-to-sequence architectures and pretrained language models," *arXiv preprint arXiv:2004.01909*, 2020.

[25] F. Mo, K. Mao, Y. Zhu, Y. Wu, K. Huang, and J.-Y. Nie, "ConvGQR: Generative query reformulation for conversational search," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, A. Rogers, J. Boyd-Graber, and N. Okazaki, Eds. Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 4998–5012.

[26] Y. Jang, K.-i. Lee, H. Bae, H. Lee, and K. Jung, "IterCQR: Iterative conversational query reformulation with retrieval guidance," in *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, K. Duh, H. Gomez, and S. Bethard, Eds. Mexico City, Mexico: Association for Computational Linguistics, Jun. 2024, pp. 8121–8138.

[27] OpenAI, "GPT-4 technical report," 2023.