Asking Clarifying Questions for Conversational Search

Ansh Bisht and **Jordan Dickson** and **Jack Harrison** and **Ruben Lazell** and **Andrew Taison**
School of Computing, Engineering and the Built Environment,
Edinburgh Napier University
Matric Numbers: 40527530, 40545300, 40537035, 40679914, 40538519

Abstract

Short abstract.

1 Introduction

Brief context of problem - why clarifying Q's matter. What is the goal. Structure of the paper.

2 Related Work

This project is motivated by recent work in conversational search, particularly where ambiguity in user queries is addressed by asking clarifying questions. Across the discussed literature it is commonly noted that users often struggle to express their intent clearly in a single query and find that even simple follow up questions can help search systems return more relevant results. Whilst our system simplifies many of the models used in previous work, it is inspired by several of their core ideas, adapting them to be lightweight for demonstration purposes and time constraints.

2.1 Clarification as a Ranking Problem and the Qulac Dataset

Aliannejadi et al. (2019) examine how clarifying questions can improve conversational search, showing that even a single, well targeted question can significantly improve retrieval performance. To support this, they introduce the Qulac dataset, which extends TREC Web Track topics with crowd sourced question and answer pairs organised by query facets (categories). This dataset has been widely used in related research. In addition to the dataset, they propose a retrieval framework in which both documents and candidate questions are retrieved and ranked based on the user's query and conversational context. Question ranking is performed using a BERT model pre-trained on Wikipedia and fine-tuned on Qulac.

Our system builds on this framework, making use of the Qulac dataset for its high quality question-answer pairs and Aliannejadi et al.'s multi-turn conversation extension dataset. Like their model, we treat clarification as a ranking problem, selecting the most relevant question from a fixed set based on the current context, rather than generating questions from scratch.

2.2 Simulated Interaction and Clarification Behaviour

Alongside ranking-based clarification frameworks, other studies have focused on how user behaviour can affect the system and effectiveness. Two notable examples are the CoSearcher system introduced by Salle et al. (2022) and the risk-aware decision model proposed by Wang and Ai (2022). Whilst these works differ in their objectives, both incorporate user simulators and address when and how clarification should take place in conversational search.

Salle et al. (2022) present CoSearcher, a user simulator designed to evaluate search intent refinement in conversational search by modelling different user behaviours. Their main focus is on testing how different facet-ranking strategies perform under varying conditions such as low user patience or reduced cooperativeness. The authors experiment with a range of ranking models, including LexVec, and use BERT to classify user responses and determine whether further clarification is necessary. Although SBERT was initially considered, it was not used, as LexVec was found to slightly outperform it.

Wang and Ai (2022) approach the clarification process as a decision making problem, where the system has to decide whether to ask a question or return a result based on how confident it is. To explore this, they train a reinforcement learning model using a reward function that balances retrieval performance with the length and usefulness of the conversation. Like CoSearcher, they implement a user simulator to represent different be-

haviours, including tolerance and patience. They use these two parameters to determine how many questions a user will put up with before giving up.

Both studies highlight the importance of choosing not just what to ask, but when to ask it. Whilst our system does not include a user simulator or learning-based decision model, we were influenced by how these papers handle ambiguity and user behaviour. In our case, the decision to clarify or move on is made using a straightforward confidence check from the BM25 module. If the score is high, we assume the summary is a good enough match for the user's intent. If it is low, the system switches to question selection instead. Whilst this is a much simpler method, it follows the same general idea that clarification should depend on context and confidence, rather than being the default.

3 Methodology

System architecture. What models/algorithms used. How are the questions generated. What inputs/outputs are there.

(Aliannejadi et al., 2019)

4 Evaluation

How did we check the system works. Describe testing process.

5 Results and Discussion

Sample outputs. Summary of system behaviour. What worked well, what didn't? Possible improvements.

6 Conclusion

start here.

Limitations

Required by ACL format, and should be AFTER conclusion. Discuss honest limitations of the work.

Ethics Statement

Required by ACL format. Could just be a sentence or two. Explicit ethics statement on the broader impact of the work, or other ethical considerations.

References

Mohammad Aliannejadi, Hamed Zamani, Fabio Crestani, and W. Bruce Croft. 2019. Asking clarifying questions in open-domain information-seeking conversations. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 475–484. ACM.

Alexandre Salle, Shervin Malmasi, Oleg Rokhlenko, and Eugene Agichtein. 2022. Cosearcher: studying the effectiveness of conversational search refinement and clarification through user simulation. *Information Retrieval Journal*, 25:209–238.

Zhenduo Wang and Qingyao Ai. 2022. Simulating and modeling the risk of conversational search. *ACM Transactions on Information Systems*, 40:1–33.

A Appendix

Possibly not needed.