Conversational Search: a Short Review

Hongjin Qian

March 2021

1 Introduction

Search engine plays an indispensable role in obtaining information for contemporary human beings. Most of current search engines focus on ad-hoc search which the search result is usually determined by the current query issued by a user. However, there is a notable gap between the user and the search system: the user knows roughly what she needs while not what is available; the system knows what is available while not what the user needs [17]. To bridge the gap, few attempts have been made to manage a interactive processes to help the search engine better understand the user's accurate intent such as query suggestion, query modelling, and personalized search [8]. Nevertheless, the ambiguity of the one-shot query is still unsolved. To alleviate such an ambiguity, researchers turns to implement conversational assistance that helps the search system to accurately locate the actual and accurate search intent of the user.

2 Conversational Search

A conversational search system defines an information retrieval system which allows users to interact with the system to seek information via multi-turn conversations of natural language. Most traditional IR systems are regarded as single-turn information seeking. Though these systems are designed to suggest, auto-complete and recommend queries, interactions between the system and the users are rather limited. On the contrary, a task-oriented dialogue system usually conduct a two-way process that facilitate the both side achieve some goal. A conversational search system, however, is more like a mixture of the two systems. It applies dialogue-like methods to facilitate information seeking.

The major motivation behind the conversational search system is: (1) the initial query issued by a user is ambiguous [1], which leads to rather different results for users; (2) the initial query may be underspecified that lacks details and will lead to diverse search results [34]; (3) the query formulated by the user might not contain the best search term that would lead to semantic bias; (4) the actual intent of a query might be strongly related to other topics which confuses the search system [5]. To battle these long-standing problem in the search engine, the domain of Conversational Search captures tremendous interests.

In fact, conversational search can be date back to early stages of IR community [4, 10]. Early researchers applies heuristic methods such as manual strategies and expert system to interact with the users in the search session. With the rising of neural networks, many end-to-end methods have been proposed to facilitate the conversational search system [19, 41, 2, 29, 32, 21, 23, 20, 35, 30, 9, 6, 33]. Prior to searching-related practices, conversational question answering attract many attentions as it can be considered as a simplified version of in-real information-seeking process. In the following, I would first introduce recent techniques used in conversational question answering and major tasks in conversational search including query Suggestion, query resolution, and query clarification. Except for the technical perspective, I will then introduce the commonly-used datasets and metrics to evaluate conversational search methods.

2.1 Conversational QA

Conversational Question Answering (CQA) is a question-answering (QA) task that is based on not only context passage but the QA histories. CQA aims to model the procedures that human seek information. Machine Reading Comprehension (MRC) is a similar task to CQA. The MRC task usually seeks to find an answer from a context passage. The answer types for MRC are usually multiple choice, text spans, free-form text, or unanswerable. The major difference between CQA and MRC is that the CQA task is conversational while the MRC task usually only have single-turn QA.

Generally, the CQA task seeks to find an answer A_k that maximizes the probability $P(A_k|C, Q_k^*)$. C is the context passage. $Q_k^* = \{Q_1, A_1, ..., Q_{k-1}, A_{k-1}, Q_k\}$ is the QA history. Compared to MRC, CQA takes both the context passage and the QA history into consideration. For each turn in the QA history, the answer can be free-form text, unanswerable, or text span [32].

As the CQA tasks can be considered as a joint task of multi-turn dialogue and MRC, the most intuitive idea to solve such a problem is combining dialogue models and MRC models. Thus, Reddy et al. [32] propose combining the Document Reader (DrQA) model and Pointer-Generator network (PGNet) to first extract a text span from the context passage as rational and use the PGNet to predict the final answer.

Later, inspired by the great generalization ability of large-scale pretrained language models [13, 26], models to solve the CQA task are usually besed on these large pretrained language models [22, 21, 23, 38]. The current state-of-the-art model for CQA task is proposed by Ju et al. [22] in which a Roberta + AT + KD framework is applied. The system is featured by Adversarial Training (AT) and Knowledge Distillation (KD). The Adversarial Training [18] create adversarial examples by making small perturbations into the embedding layers, which improves the robustness of the trained models. The Knowledge Distillation [16] process applies several teacher models that are initiated by several random seeds to distill knowledge into a student model.

Current models for the CQA task are challenged by the various types of answers. A future direction to alleviate such a problem is to combine the extractive models and generative models in a more effective way [2, 22].

2.2 Query Resolution

The query resolution is a long-standing challenge in Conversational search because of the ubiquitous phenomena such as zero anaphora, topic change, and topic return in information-seeking process.

The query resolution task has been early-studied in the context of dialogue system. Raghu et al. [31] termed the situation as Non-Sentential Utterances (NSUs). Specifically, NSUs refers to the commonly-used short sentences that show as bullet sentences while convey a complete semantics. Raghu et al. [31] proposed applying keyword to question (K2Q) framwork to output natural language questions combining the keywords that NSU generate given the context. Similar method is also applied in retreival-based dialogue system [25]. Beyond these heuristic methods, later, the task was formulated as question-in-context rewriting [15]. The task expect to generate a self-contained question given a conversation's history.

Ideas are common in across between the dialogue system and IR system. Early research on query resolution in the IR system also applies term appending method to expand the semantics of an ad-hoc query [27]. Later works focus on encode the search session log into a condensed representation to generate a complete query [15, 25]. Inspired by the query modelling in query modelling [7], some researchers try to formulate the task as query modelling [36, 39], which implicit transform the search history into a hidden representation to conduct information retrieval. Other methods try to tackle such a problem include Term Classification that assign a binary label to each term in the history to generate the refined query [37].

A wildly-used dataset that evaluates query resolution task is the Conversational Assistance Track (CAsT) at TREC 2019 [11]. A major objective for the track is to construct a standard collection of data that can directly make. information-seeking systems comparable

2.3 Query Clarification

One barrier that challenges the success of information-seeking process is the failure to express user's information need accurately. To battle the long-existing barrier, endowing conversational search system the ability to support mixed-initiative interactions become crucial [30, 24]. Such a system is designed to provide active feedbacks to the users such as disclosing new information [24] or posing clarification [1], and the query can be therefore well-refined.

3 Datasets

- OR-QuAC [28]
- CoQA [32]
- Qulac [1] The dataset is for Question for lack of clarity dedicated to query clarification in conversational search.
- MIMICS [40]
- QuAC [9] The QuAC dataset contains Wikipedia passages along with up to 12 questions and corresponding answer spans. It is constructed by crowdsourcing in which two people are asked to perform an interactive dialogue about a specific topic. The dataset is divided into train/dev/test set by [15] which is used to evaluate tasks including query resolution and query clarification.
- CAsT-19 [12] The TREC CAsT dataset is a multi-turn passage retrieval dataset which contains 30 and 50 topics for the train and evaluate sets, respetively. For each topic, it consists of a sequence of multi-turn queries that mimic the procedures people seek information. The passage pool of CAsT-19 is constructed from MS MARCO [3] and the TREC CAR [14].

4 Evaluation Metrics

5 Conclusion

In the short review, I first introduce the fundamental clues of the story of conversational search and justify its differences from traditional IR system. Then, I formulate the major issues that a conversational search system expect to tackle and corresponding techniques. In the next, I introduce the conversational question answering (CQA), query resolution, query clarification in details. Except the technical perspective, I also summarize the commonly-used benchmark datasets and metrics to evaluate these methods. This short review remains a partial work where the major objective is to fill up four thousand words. Hence, the scope fails to cover all related topics which are expected to be supplemented in the future work.

References

[1] Mohammad Aliannejadi, Hamed Zamani, Fabio Crestani, and W. Bruce Croft. 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*, SIGIR'19, page 475–484, New York, NY, USA, 2019. Association for Computing Machinery.

[2] Ashutosh Baheti, Alan Ritter, and Kevin Small. Fluent Response Generation for Conversational Question Answering. 2020.

[3] Payal Bajaj, Daniel Campos, Nick Craswell, Li Deng, Jianfeng Gao, Xiaodong Liu, Rangan Majumder, Andrew McNamara, Bhaskar Mitra, Tri Nguyen, Mir Rosenberg, Xia Song, Alina Stoica, Saurabh Tiwary, and Tong Wang. Ms marco: A human generated machine reading comprehension dataset, 2018.

[4] Nicholas J Belkin, Colleen Cool, Adelheit Stein, and Ulrich Thiel. Cases, scripts, and information-seeking strategies: On the design of interactive information retrieval systems. *Expert systems with applications*, 9(3):379–395, 1995.

[5] Adam Berger and John Lafferty. Information retrieval as statistical translation. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '99, page 222–229, New York, NY, USA, 1999. Association for Computing Machinery.

[6] Keping Bi, Qingyao Ai, Yongfeng Zhang, and W Bruce Croft. Conversational Product Search Based on Negative Feedback 1. 2019.

[7] Ben Carterette, Paul Clough, Mark Hall, Evangelos Kanoulas, and Mark Sanderson. Evaluating retrieval over sessions: The trec session track 2011-2014. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '16, page 685–688, New York, NY, USA, 2016. Association for Computing Machinery.

[8] Yi Chang and Hongbo Deng. *Query Understanding for Search Engines.* Springer.

[9] Eunsol Choi, Mark Yatskar, Percy Liang, and Luke Zettlemoyer. QuAC : Question Answering in Context. pages 2174–2184, 2018.

[10] W. B. Croft and R. H. Thompson. I¡sup¿3¡/sup¿r: A new approach to the design of document retrieval systems. *J. Am. Soc. Inf. Sci.*, 38(6):389–404, November 1987.

[11] Jeffrey Dalton, Chenyan Xiong, and Jamie Callan. Trec cast 2019: The conversational assistance track overview, 2020.

[12] Jeffrey Dalton, Chenyan Xiong, Vaibhav Kumar, and Jamie Callan. Cast-19: A dataset for conversational information seeking. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, page 1985–1988, New York, NY, USA, 2020. Association for Computing Machinery.

[13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

[14] Laura Dietz, Manisha Verma, Filip Radlinski, and Nick Craswell. Trec complex answer retrieval overview. In *TREC*, 2017.

[15] Ahmed Elgohary, Denis Peskov, and Jordan 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)*, pages 5918–5924, Hong Kong, China, November 2019. Association for Computational Linguistics.

[16] Tommaso Furlanello, Zachary C. Lipton, Michael Tschannen, Laurent Itti, and Anima Anandkumar. Born again neural networks, 2018.

[17] Jianfeng Gao, Chenyan Xiong, and Paul Bennett. Recent Advances in Conversational Information Retrieval (CIR) - A review of neural approaches. 2020, 2020.

[18] Ian J. Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples, 2015.

[19] Daya Guo, Duyu Tang, Nan Duan, Ming Zhou, and Jian Yin. Dialog-to-action: Conversational question answering over a large-scale knowledge base. *Advances in Neural Information Processing Systems*, 2018-Decem(NeurIPS):2942–2951, 2018.

[20] Mohit Iyyer, Wen-tau Yih, and Ming-wei Chang. Search-based Neural Structured Learning for Sequential Question Answering. pages 1821–1831, 2017.

[21] Zihang Jiang, Weihao Yu, Daquan Zhou, Yunpeng Chen, Jiashi Feng, and Shuicheng Yan. Convbert: Improving bert with span-based dynamic convolution, 2021.

[22] Ying Ju, Fubang Zhao, and Shijie Chen. Technical report on Conversational Question Answering. 2019.

[23] Aisha Urooj Khan, Amir Mazaheri, Niels da Vitoria Lobo, and Mubarak Shah. Mmft-bert: Multimodal fusion transformer with bert encodings for visual question answering, 2020.

[24] Johannes Kiesel, Arefeh Bahrami, Benno Stein, Avishek Anand, and Matthias Hagen. Toward voice query clarification. In *The 41st International ACM SIGIR Conference on Research amp; Development in Information Retrieval*, SIGIR '18, page 1257–1260, New York, NY, USA, 2018. Association for Computing Machinery.

[25] Vineet Kumar and Sachindra Joshi. Incomplete follow-up question resolution using retrieval based sequence to sequence learning. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '17, page 705–714, New York, NY, USA, 2017. Association for Computing Machinery.

[26] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019.

[27] Mandar Mitra, Amit Singhal, and Chris Buckley. Improving automatic query expansion. In *Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '98, page 206–214, New York, NY, USA, 1998. Association for Computing Machinery.

[28] Chen Qu, Liu Yang, Cen Chen, Minghui Qiu, W. Bruce Croft, and Mohit Iyyer. Open-Retrieval Conversational Question Answering. 2020.

[29] Chen Qu, Liu Yang, Minghui Qiu, Yongfeng Zhang, Cen Chen, W. Bruce Croft, and Mohit Iyyer. Attentive history selection for conversational question answering. *International Conference on Information and Knowledge Management, Proceedings*, (Mc):1391–1400, 2019.

[30] Filip Radlinski and Nick Craswell. A Theoretical Framework for Conversational Search. 2017.

[31] Dinesh Raghu, Sathish Indurthi, Jitendra Ajmera, and Sachindra Joshi. A statistical approach for non-sentential utterance resolution for interactive QA system. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 335–343, Prague, Czech Republic, September 2015. Association for Computational Linguistics.

[32] Siva Reddy, Danqi Chen, and Christopher D. Manning. CoQA: A Conversational Question Answering Challenge. *Transactions of the Association for Computational Linguistics*, 7:249–266, 2019.

[33] Corby Rosset, Chenyan Xiong, Xia Song, Daniel Campos, Nick Craswell, Saurabh Tiwary, and Paul Bennett. Leading Conversational Search by Suggesting Useful Questions. 2:1160–1170, 2020.

[34] Yang Song, Dengyong Zhou, and Li-wei He. Query suggestion by constructing term-transition graphs. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM '12, page 353–362, New York, NY, USA, 2012. Association for Computing Machinery.

[35] M Thomas, Francesco Piccinno, Massimo Nicosia, Peter Shaw, and Yasemin Altun. Answering Conversational Questions on Structured Data without Logical Forms. pages 5902–5910, 2019.

[36] Christophe Van Gysel, Evangelos Kanoulas, and Maarten de Rijke. Lexical query modeling in session search. In *Proceedings of the 2016 ACM International Conference on the Theory of Information Retrieval*, ICTIR '16, page 69–72, New York, NY, USA, 2016. Association for Computing Machinery.

[37] Nikos Voskarides, Dan Li, Pengjie Ren, Evangelos Kanoulas, and Maarten De Rijke. Query Resolution for Conversational Search with Limited Supervision.

[38] Jindou Wu, Yunlun Yang, Chao Deng, Hongyi Tang, Bingning Wang, Haoze Sun, Ting Yao, and Qi Zhang. Sogou Machine Reading Comprehension Toolkit. *arXiv e-prints*, page arXiv:1903.11848, Mar 2019.

[39] Hui Yang, Dongyi Guan, and Sicong Zhang. The query change model: Modeling session search as a markov decision process. *ACM Trans. Inf. Syst.*, 33(4), May 2015.

[40] Hamed Zamani, Gord Lueck, Everest Chen, Rodolfo Quispe, Flint Luu, and Nick Craswell. Mimics: A large-scale data collection for search clarification. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 3189–3196, 2020.

[41] Chenguang Zhu, Michael Zeng, and Xuedong Huang. SDNet: Contextualized Attention-based Deep Network for Conversational Question Answering. pages 1–8, 2018.