Recent Advances in Conversational Information Retrieval

Jianfeng Gao, Chenyan Xiong, Paul Bennett
Microsoft Research AI
One Microsoft Way, Redmond
WA 98052, USA
jfgao, chenyan.xiong, paul.n.bennett@microsoft.com

ACM Reference Format:
Jianfeng Gao, Chenyan Xiong, Paul Bennett. 2020. Recent Advances in Conversational Information Retrieval. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR ’20), July 25–30, 2020, Virtual Event, China.* ACM, San Francisco, USA, 4 pages. https://doi.org/10.1145/3397271.3401418

1 MOTIVATION

A conversational information retrieval (CIR) system is an information retrieval (IR) system with a conversational interface which allows users to interact with the system to seek information via multi-turn conversations of natural language (in spoken or written form). CIR provides a more natural interface for information seeking than traditional, single-turn, search engines and is particularly useful for search on modern devices with small or no screen.

CIR is a long-standing topic, which we can trace back to the 1960’s. However, the research in CIR remained in its infancy until the 2010’s due to the limitations in conversational data, natural language processing (NLP) technologies, commercial needs etc. Even today, commercial search engines only provide some dialogue capabilities considering the history of research in this area.

Recent progress in deep learning has brought tremendous improvements in conversational AI, leading to a plethora of commercial conversational services that allow naturally spoken interactions, increasing the need for more human-centric interactions in IR. As a result, we have witnessed a resurgent interest in developing modern CIR systems in research communities and industry.

This tutorial presents recent advances in CIR, focusing mainly on neural approaches and new applications developed in the past five years. Our goal is to provide a thorough and in-depth overview of the general definition of CIR, the components of CIR systems, new applications raised for its conversational aspects, and the (neural) techniques recently developed for it.

The tutorial is partly based on our in-progress textbook, with IR and NLP communities as the primary target audience. Audiences with other backgrounds will also find it an accessible introduction to CIR. The materials will cover both the basics of CIR and the advanced neural techniques recently developed; the presentation is accessible to a broad cross-section of SIGIR attendees with different types of backgrounds.

SIGIR ’20, July 25–30, 2020, Virtual Event, China

ACM ISBN 978-1-4503-8016-4/20/07.
https://doi.org/10.1145/3397271.3401418

2 OBJECTIVES

This tutorial will provide a survey-like foundation and build a deep understanding in select areas of recent CIR advances to SIGIR attendees; it will also attract more colleagues in the IR community to this important field, and prove a valuable resource for students, researchers, and practitioners.

CIR is a rapidly growing field; many new concepts and techniques are being developed constantly. It is also a topic of interest from multiple communities, e.g., information retrieval, natural language processing, speech, dialogue, human computer interaction, and deep learning – a truly multi-discipline topic. However, the rapid growth and the multi-discipline nature unavoidably leads to entangled concepts and fast-moving views of CIR, which could confuse researchers and practitioners new to this field.

This tutorial aims to achieve the following goals.

(1) *To speed up the adoption of CIR:* Provide an overview of recent advances in CIR in IR and related communities, clarify important concepts, and present a comprehensive survey of neural techniques developed recently.
(2) *To facilitate the future research in CIR:* Overview the landscape of CIR, categorize its research topics, and discuss the important problems raised in recent research.
(3) *Unify the interests and views in academia and industry.* Highlight the process in both academia and industry, from both a researcher’s point of view and an industry practitioner’s point of view.
(4) *Provide a reference point for future research.* Survey related work from multiple research communities, with the potential to serve as a general reference point.

3 RELEVANCE AND RELATED WORK

CIR is an emerging topic in the core information retrieval community. It is considered as one of the most important emerging research areas in the SWIRL 2018 workshop report on future trends in Information Retrieval [8]. Commercial search engines are shifting from the traditional “ten blue links” user experience to a more conversational search experience — thanks to the rapid development of AI techniques and the availability of portable devices. For example, in 2019, half of Google’s search queries no longer end with a click on the SERP but provide users with more distilled and immersive results, as well as a more interactive experience [10].

This tutorial surveys the recent advances in CIR in industry and academia for SIGIR audiences. Although many tutorials on dialog systems, QA systems, and neural approaches to these dialog and QA systems have been presented in recent conferences, to the best of our knowledge, this is the first tutorial on Conversational IR.

Related tutorials include the following.

The tutorial on *Neural Approaches to Conversational AI* was co-presented by one of our presenters at ACL 2018, SIGIR 2018, and ICML 2019 [11]. It covers the state-of-the-art neural techniques developed for three types of dialog systems: question answering, task-oriented dialogue, and social chat bots. In comparison, CIR, the scope of this tutorial, is related to but differs significantly from the above three types of dialog systems. For example, in CIR, a user often starts with a search goal (similar to that of QA and task-oriented dialog systems), but then shifts her interest and starts to explore new topics based on the result returned by CIR (similar to that of social chatbots). This tutorial aims to give a comprehensive view of CIR by consolidating task definitions and problem formulations of previous works, and reviewing related neural techniques to CIR.

The *Deep Chit-Chat: Deep Learning for ChatBots* tutorial was presented in EMNLP 2018, WebConf 2019, and SIGIR 2019 [23]. That tutorial surveys deep learning models used developed for chit-chat dialog assistants, and includes an in-depth overview of the techniques behind Microsoft's chatbot, Xiaoice.

The *Deep Learning for Conversational AI* tutorial at NAACL 2018 reviewed goal-oriented spoken dialogue systems [20]. It surveyed dialog component techniques such as spoken utterance understanding, state tracking, and dialogue management.

The *Data Collection and End-to-End Learning for Conversational AI* tutorial at EMNLP 2019 surveyed the datasets developed for Conversational AI [22].

4 FORMAT AND SCHEDULE

This half-day tutorial will include three hour materials scheduled as in the following parts.

4.1 Introduction of CIR

This part motivates the research of CIR by reviewing the studies on how people search and how information seeking can be cast as human-machine conversations.

How People Search. We first define task of information seeking as including information lookup and exploratory search [6]; then we present the problem formulation, with information seeking as a process of sense-making which is an iterative process of formulating a conceptual representation from a large collection of information. We show the conversational aspect of information seeking by presenting a dynamic model motivated by users dynamic search behaviors: users learn as they search, and their information needs adjust as they see retrieval results and other document surrogates. The dynamic process is also referred to as the berry picking model of search [3]. Effective access to information often requires interaction between the user and the system, where both the user and the system should play a role. Even more effective information access requires a system that actively supports effective interaction and explicitly models the interaction. We will also describe the properties of CIR, the definition of a CIR system [18] and present a CIR example.

Background. We review recent advances in related dialogue systems [12], and briefly discuss the early works in CIR [7]. We draw the connection between early approaches to CIR and modern neural approaches to CIR. Most of these early works are based on traditional, keyword-based IR systems. But the concepts and design principles that were explored are instrumental in building new generations of CIR systems. Below are a few examples.

(1) [15]: interactive search and retrieval methods using automatic information display
(2) [2]: Information search tactics, JASIS, 1979
(3) [4]: understanding intermediary interactions
(4) [1]: Relevance feedback interactions
(5) [19]: Clarifying search – A user interface framework for text searches.
(6) [21]: Advantages of query biased summaries in information retrieval.
(7) [9]: Scatter/Gather: a cluster-based approach to browsing large document collections.
(8) [17]: Information Retrieval through man-machine dialog.

4.2 Conversational Search Systems

The second part of this tutorial discusses the core components in CIR systems, with the focuses on the new challenges raised from the conversational nature in comparison with classic, keyword-based, IR systems. There are three sections in this part. The first section describes the general framework of a CIR system; the remaining two present two core CIR components, conversational query understanding and conversational document ranking, respectively.

We start with a description of the general framework of a typical CIR system, including its core components, the information and data flow, and the differences from keyword-based IR systems. Then, we present an example of a vanilla set-up of a CIR system.

Conversational Query Understanding differs from traditional query understanding in that the former needs to take into account contextual information in conversation history (i.e., query session). We present the definition of the task and its available benchmarks. Then we present two categories of solutions. One is based on classic IR query understanding techniques (e.g., session-based query expansion), and the other on natural language generation (e.g., contextual query rewriting). Most of these solutions are developed based on the recent Conversational Assistant Track in TREC 2019 and various on-going works afterwards.

Conversational Document Ranking. We survey the recent advances in conversational document ranking. We define the conversational ranking task and distinguish it from ranking in ad-hoc retrieval. Then, we discuss the benchmarks and evaluation metrics in conversational ranking. After that, we present the background on ad-hoc retrieval and ranking, focusing on neural ranking models.

After that, we review conversational ranking models, focusing the discussion on new challenges compared to the models developed for ad-hoc retrieval, including contextual ranking, and new training techniques (e.g., multi-task learning, end-to-end learning) that have been developed and adapted for conversational query understanding.

4.3 Interactive Information Retrieval Methods

The third part presents a set of methods that aim to make CIR systems play a more active role in human-machine interactions of information seeking. These methods collectively allow a CIR system to actively help a user clarify her search intent by asking clarification questions, guide the user to discover and learn new knowledge by suggesting and recommending related information and search in new directions, and synthesize retrieved content to form a new understanding. Finally, we present the challenges and solutions when evaluating such an interactive CIR system.

Learning to Ask Clarifying Questions. The interactive nature of CIR presents new opportunities for the system to directly ask users to clarify what they want. This is referred to as the "learning to ask" task, which has recently attracted a lot of attention in both the industry and research community. We first define the task and the benchmarks for evaluation. Then we present two main challenges of the task and the solutions being developed to address them: (1) how to model the ambiguity of information needs, and (2) how to generate clarification questions.

Conversation Leading Recommendation and Suggestions. In natural conversations, the participants not only "passively" respond but each party can lead, switch, or initialize new conversations. Being able to not only respond to user's queries, but also make suggestions and recommendations to lead the conversation, is another appealing feature of CIR. We will first motivate the task and discuss the evaluation benchmarks; then present two groups of techniques in conversational leading: text-based and structure-based.

CIR Result Presentation. A CIR system needs to present the search result in a concise form such that a user can easily investigate whether the desired result is included and the search session can be ended, or new search needs to be issued. This part describes the two groups of methods that have been developed to better organize and present the results based on the dialog history (search session). The first is to group search results into different categories. The second is to summarize multiple search results into a concise answer.

Evaluation of Conversational Search Systems. We wrap up the discussion on CIR systems by presenting evaluation strategies, including the evaluation of CIR's sub-tasks (e.g. query rewriting accuracy, ranking accuracy) and the overall end-to-end success metrics (e.g., utility, engagement, and the balance of interaction and task completion).

4.4 Conversational QA over Texts

This part presents recent developments in conversational Question Answering (QA) systems that allow users to query a document collection in natural language. Such text-QA agents are much easier to use in mobile devices than traditional search engines in that they provide concise, direct answers to user queries, as opposed to a list of document links.

The heart of these systems is a neural machine reading comprehension (MRC) model that generates an answer to an input query based on a set of passages and dialog history. We start with an introduction to the conversational text-QA task. Using CoQA and QuAC as examples, we discuss the dialog features such as topic shift, drill down, clarification.

Then, we present an overview of a typical architecture of neural MRC models that are originally developed for single-turn QA. Then, we review extensions of these models that can handle dialog features of conversational QA. We use state-of-the-art MRC models (BiDAF++, BERT, DrQA etc.) as examples to describe the model components that are developed for each of the three steps: document/query embedding, fusion and answer generation.

We conclude this part with a discussion on the use of large-scaled pre-trained models for conversational QA. On many conversational MRC leaderboards[1], the top-performing systems all use large-scaled pre-trained models e.g., BERT and RoBERTa [14]. We discuss the strength and weaknesses of these models, and the recent improvements on robustness, interpretability, easy-deployment etc.

4.5 Conversational QA over Structured Databases

This part presents recent developments in conversational Question Answering over structured databases, i.e. Knowledge Bases (KBs).

We start with an introduction to the task, describing the forms of KB and the types of the questions. Then, we give an overview of a typical architecture of KB Question Answering (KBQA) systems that are based on semantic parsing.

After that, we describe each of the three main components in a typical semantic parser based KBQA system, including the semantic parser, dialog manager, and response generator. We present in detail a Dynamic Neural Semantic Parser (DynSP) based on [13], and several methods of extending DynSP or similar neural semantic parsers to cope with ellipsis phenomena (i.e., co-reference and ellipsis resolution) in conversational QA.

We also survey the response generation techniques with structured groundings. A response generator converts system actions to natural language responses. It is an important component that affects naturalness of a conversational system, and user experience. While it is traditionally implemented using template-based methods, we focus on the modern approaches that use neural generative language models.

We wrap up the discussions with a brief review of the approaches that are not based on semantic parsing, such as the graph neural network based approaches and the unsupervised inferences [5, 16].

4.6 Case Study of Commercial Systems.

The last part of this tutorial reviews a variety of commercial systems for CIR and related tasks. Due to the proprietary nature of many of these systems, We limit our review to summarizing published material about the systems.

Research Platforms and Toolkits. With the growth of CIR, the need to build customized search experiences for a variety of domains has arisen. This has given rise to a variety of toolkits that enable researchers and engineers to quickly create robust conversational interfaces. In this section we review such toolkits including: DialogFlow, Conversation Learner, ConvLab, Alexa etc.

Commercial Applications. We provide case studies of various commercial applications of CIR.

(1) Chatbots: We briefly review several popular *chatbots* whose goal was not primarily information seeking but as AI companions. We will review social chatbots such as Microsoft's XiaoIce [24].
(2) Conversational Search Engine: With the advances in CIR, even standard GUI interfaces to search have become more

[1] https://stanfordnlp.github.io/coqa/ and https://quac.ai/

conversational. These shifts involve QA where instant answers on a page present extracted information directly on the Search Engine Result Page (SERP) as well as other features like "People Also Ask" (PAA), and clarifying questions among others.

(3) Productivity-Focused Agents: Some of the first CIR experiences developed around the notion of a personal assistant that could act on a person's private information store. These assistants were often focused towards productivity tasks such as calendar management, task management, contact management, etc.
(4) Device-based Assistants: Strongly related to productivity-focused assistants are device-based assistants. These assistants typically offer CIR integrated throughout their functionality and often are integrated with personal devices (e.g. mobile phones).
(5) Hybrid-Intelligence Assistants: These assistants typically blend automated systems with crowd-powered systems to provide capabilities that are beyond automated systems current capabilities but with an overall impression to the user that the entire experience is completely AI-powered. Examples we review include Facebook's M and Microsoft's Calendar.help.

5 SUPPORT MATERIALS

This tutorial is self-contained and we will share a detailed list of references to surveyed papers.

REFERENCES

[1] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, et al. 1999. *Modern information retrieval*. Vol. 463. ACM press New York.
[2] Marcia J Bates. 1979. Information search tactics. *Journal of the American Society for information Science* 30, 4 (1979), 205–214.
[3] Marcia J Bates. 1989. The design of browsing and berrypicking techniques for the online search interface. *Online review* 13, 5 (1989), 407–424.
[4] Helen M Brooks, Penny J Daniels, and Nicholas J Belkin. 1986. Research on information interaction and intelligent information provision mechanisms. *Journal of Information Science* 12, 1-2 (1986), 37–44.
[5] Philipp Christmann, Rishiraj Saha Roy, Abdalghani Abujabal, Jyotsna Singh, and Gerhard Weikum. 2019. Look before you Hop: Conversational Question Answering over Knowledge Graphs Using Judicious Context Expansion. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 729–738.
[6] Kevyn Collins-Thompson, Preben Hansen, and Claudia Hauff. 2017. Search as learning (dagstuhl seminar 17092). In *Dagstuhl reports*, Vol. 7. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
[7] W. Bruce Croft. 2019. The Importance of Interaction for Information Retrieval. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, New York, NY, USA, 1–2.
[8] J. Shane Culpepper, Fernando Diaz, and Mark Smucker. 2018. Research Frontiers in Information Retrieval: Report from the Third Strategic Workshop on Information Retrieval in Lorne (SWIRL 2018). *SIGIR Forum* 52, 1 (2018), 34–90.
[9] Douglass R Cutting, David R Karger, Jan O Pedersen, and John W Tukey. 2017. Scatter/gather: A cluster-based approach to browsing large document collections. In *ACM SIGIR Forum*, Vol. 51. ACM New York, NY, USA, 148–159.
[10] Rand Fishkin. 2019. Less than Half of Google Searches Now Result in a Click. sparktoro.com/blog/less-than-half-of-google-searches-now-result-in-a-click/. Accessed:2020-01-23.
[11] Jianfeng Gao, Michel Galley, and Lihong Li. 2018. Neural Approaches to Conversational AI. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. 1371–1374.
[12] Jianfeng Gao, Michel Galley, Lihong Li, et al. 2019. Neural approaches to conversational AI. *Foundations and Trends® in Information Retrieval* 13, 2-3, 127–298.
[13] Mohit Iyyer, Wen-tau Yih, and Ming-Wei Chang. 2017. Search-based neural structured learning for sequential question answering. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1821–1831.
[14] Ying Ju, Fubang Zhao, Shijie Chen, Bowen Zheng, Xuefeng Yang, and Yunfeng Liu. 2019. Technical report on Conversational Question Answering. *arXiv preprint arXiv:1909.10772* (2019).
[15] Michael E Lesk and Gerard Salton. 1969. Interactive search and retrieval methods using automatic information displays. In *Proceedings of the May 14-16, 1969, spring joint computer conference*. ACM, 435–446.
[16] Thomas Müller, Francesco Piccinno, Massimo Nicosia, Peter Shaw, and Yasemin Altun. 2019. Answering Conversational Questions on Structured Data without Logical Forms. *arXiv preprint arXiv:1908.11787* (2019).
[17] Robert N Oddy. 1977. Information retrieval through man-machine dialogue. *Journal of documentation* 33, 1 (1977), 1–14.
[18] Filip Radlinski and Nick Craswell. 2017. A theoretical framework for conversational search. In *Proceedings of the 2017 conference on conference human information interaction and retrieval*. ACM, 117–126.
[19] Ben Shneiderman, Don Byrd, and W Bruce Croft. 1997. Clarifying search: A user-interface framework for text searches. *D-lib magazine* 3, 1 (1997), 18–20.
[20] Pei-Hao Su, Nikola Mrkšić, Iñigo Casanueva, and Ivan Vulić. 2018. Deep learning for conversational AI. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorial Abstracts*. 27–32.
[21] Anastasios Tombros, Mark Sanderson, and Phil Gray. 1998. Advantages of query biased summaries in information retrieval. In *SIGIR*, Vol. 98. 2–10.
[22] Tsung-Hsien Wen, Pei-Hao Su, Paweł Budzianowski, Iñigo Casanueva, and Ivan Vulić. 2019. Data Collection and End-to-End Learning for Conversational AI. 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): Tutorial Abstracts*. ACL.
[23] Wei Wu and Rui Yan. 2019. Deep Chit-Chat: Deep Learning for Chatbots. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1413–1414.
[24] Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum. 2018. The design and implementation of XiaoIce, an empathetic social chatbot. *arXiv preprint arXiv:1812.08989* (2018).