Conversational Search Conversational Search for Learning Technologies

Sharon Oviatt, Laure Soulier

HAL Id: hal-02488525

https://hal.sorbonne-universite.fr/hal-02488525v1

Submitted on 22 Feb 2020

Conversational Search for Learning Technologies

Sharon Oviatt

Monash University, Melbourne, VIC, Australia

sharon.oviatt@monash.edu

Laure Soulier

Sorbonne Universit, CNRS, LIP6, F-75005 Paris, France

laure.soulier@lip6.fr

Abstract

Conversational search is based on a user-system cooperation with the objective to solve an information-seeking task. In this report, we discuss the implication of such cooperation with the learning perspective from both user and system side. We also focus on the stimulation of learning through a key component of conversational search, namely the multimodality of communication way, and discuss the implication in terms of information retrieval. We end with a research road map describing promising research directions and perspectives.

1 Context and background

1.1 What is Learning?

Arguably, the most important scenario for search technology is lifelong learning and education, both for students and all citizens. Human learning is a complex multidimensional activity, which includes procedural learning (e.g., activity patterns associated with cooking, sports) and knowledge-based learning (e.g., mathematics, genetics). It also includes different levels of learning, such as the ability to solve an individual math problem correctly. It also includes the development of meta-cognitive self-regulatory abilities, such as recognizing the type of problem being solved and whether one is in an error state. These latter types of awareness enable correctly regulating ones approach to solving a problem, and recognizing when one is off track by repairing momentary errors as needed. Later stages of learning enable the generalization of learned skills or information from one context or domain to others such as applying math problem solving to calculations in the wild (e.g., calculation of garden space, engineering calculations required for a structurally sound building).

1.2 Human versus System Learning

When people engage an IR system, they search for many reasons. In the process they learn a variety of things about search strategies, the location of information, and the topic about

which they are searching. Search technologies also learn from and adapt to the user, their situation, their state of knowledge, and other aspects of the learning context [CTHH17]. Beyond adaptation, the engagement of the system impacts the search effectiveness: its pro-activity is required to anticipate users need, topic drift, and lower the cognitive load of users [TY19]. For example, when someone is using a keyboard-based IR system of today, educational technologies can adapt to the persons prior history of solving a problem correctly or not, for example by presenting a harder problem next if the last problem was solved correctly, or presenting an easier problem if it was solved incorrectly.

Based on conversational speech IR systems, it is now possible for a system to process a persons acoustic-prosodic and linguistic input jointly, and on that basis a system can adapt to the persons momentary state of cognitive load. The ideal state for engaging in new learning would be a moderate state of load, whereas detection of very high cognitive load might suggest that the person could benefit from taking a break for some period of time or address easier subtopics to decomplexify the search task [AWP$^+$14].

2 Motivation

2.1 How is Learning Stimulated?

Based on the cognitive science and learning sciences literature, it is well known that human thought is spatialized. Even when we engage in problem-solving about temporal information, we spatialize it [JL99]. Since conversational speech is not a spatial modality, it is advantages to combine it with at least one other spatial modality. For example, digital pen input permits handwriting diagrams and symbols that convey spatial location and relations among objects. Further, a permanent ink trace remains, which the user can think about. Tangible input like touching and manipulating objects in a virtual world also supports conveying 3D spatial information, which is especially beneficial for procedural learning (e.g., learning to drive in a simulator). Since learning is embodied and enhanced by a persons physical activity, touch, manipulation, and handwriting can spatialize information and result in a higher level of interactivity, producing more durable and generalizable learning. When combined with conversational input for social exchange with other people, such input supports richer multimodal input.

Based on the information-seeking point of view, the understanding of users information need is crucial to maintain their attention and improve their satisfaction. As of now, the understanding of information need has been evaluated using relevant documents, but it implies a more complex process dealing with information need elicitation due to its formulation in natural language [ASD18] and information synthesis [Mar06, WR09]. There is, therefore, a crucial need to build information retrieval systems integrating human goals.

2.2 How Can We Benefit from Multimodal IR?

Multimodality is the preferred direction for extending conversational IR systems to provide future support for human learning. A new body of research has established that when a person can use multimodal input to engage a system, all types of thinking and reasoning are facilitated, including (1) convergent problem solving (e.g., whether a math problem is solved correctly); (2) divergent ideation (e.g., fluency of appropriate ideas when generating

science hypotheses); and (3) accuracy of inferential reasoning (e.g., whether correct inferences about information are concluded or the information is overgeneralized) [Ovi13]. It is well recognized within education that interaction with multimodal/multimedia information supports improved learning. It also is well recognized that this richer form of information enables accessibility for a wider range of diverse students (e.g., blind and hearing impaired, lower-performing, non-native speakers) [Ovi13].

For these and related reasons, the long-term direction of IR technologies would benefit by transitioning from conversational to multimodal systems that can substantially improve both the depth and accessibility of educational technologies. With respect to system adaptivity, when a person interacts multimodally with an IR system, the system now can collect richer contextual information about his or her level of domain expertise [OGCO19]. When the system detects that the person is a novice in math, for example, it can adapt by presenting information in a conceptually simpler form and with fewer technical terms. In contrast, when a person is detected to be an expert, the system can adapt by upshifting to present more advanced concepts using domain-specific terminology and greater technical detail. This level of IR system adaptivity permits targeting information delivery more appropriately to a given person, which improves the likelihood that he or she will comprehend, reuse, and generalize the information in important ways. The more basic forms of system adaptivity are maintained, but also substantially expanded by the integration of more deeply human-centered models of the person and their existing knowledge of a particular content domain.

Apart from the greater sophistication of user modeling and improved system adaptivity, multimodal IR systems would benefit significantly by becoming more robust and reliable at interpreting a persons queries to the system, compared with a speech-only conversational system [OC15]. This is because fusing two or more information sources reduces recognition errors. There are both human-centered and system-centered reasons why recognition errors can be reduced or eliminated when a person interacts with a multimodal system. First, humans will formulate queries to the IR system using whichever modality they believe is least error-prone, which prevents errors. For example, they may speak a query, but switch to writing when conveying surnames or financial information involving digits. In addition, when they encounter a system error after speaking input, they can switch to another modality like writing information or even spelling a wordwhich leads to recovering from the error more quickly. When using a speech-only system, instead the person must re-speak information, which typically causes them to hyperarticulate. Since hyperarticulate speech departs farther from the systems original speech training model, the result is that system errors typically increase rather than resolving successfully [OC15].

2.3 How can user learning and system learning function cooperatively in a multimodal IR framework?

Conversational search needs to be supported by multimodal devices and algorithmic systems trading off search effectiveness and users satisfaction [TY19]. Figure 1 illustrates how the user, the system, and the multi-modal interface might cooperate. The conversation is initiated by users who formulate their information need through a modality (voice, text, pen,). The system is expected to be proactive by fostering both 1) user revealment by eliciting the information need and 2) system revealment by suggesting what actions are available at the current state of the session [ADHD18]. In response, users are able to clarify their need and the span of the search session, providing them a deeper knowledge with respect to their
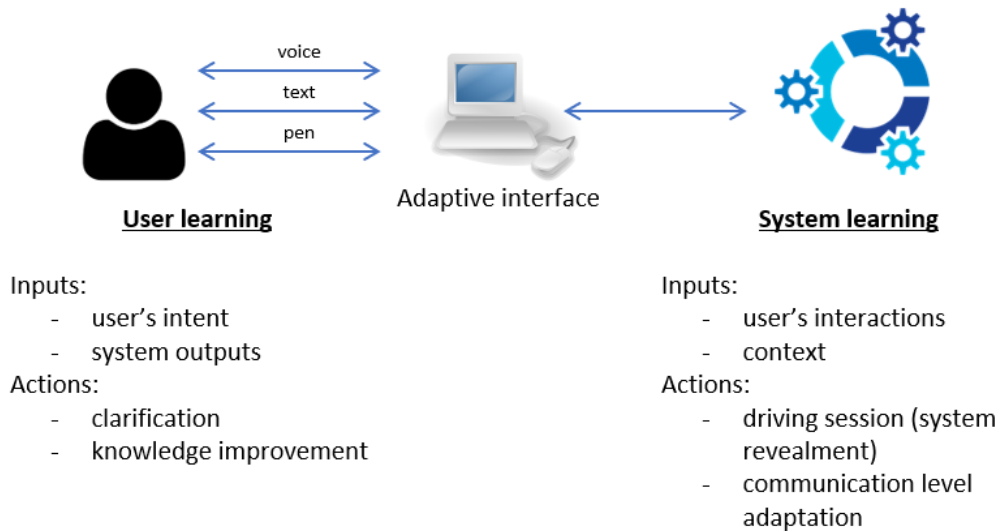
Figure 1: User Learning and System Learning in Conversational Search.

information need. The relevant features impacting both users and systems actions include 1) users intent, 2) users interactions, 3) system outputs, and 4) the context of the session (communication modality, spatial and temporal information). Several advantages of the user and system cooperation might be noticed. First, based on past interactions, the system is able to learn from right and wrong past actions. It is, therefore, more willing to target IR pieces of information that might be relevant to users. This straightforward allows reducing interactions between users and systems and lower the cognitive effort of users. Second, users being driven by increasing their knowledge acquisition experience, the system should be able to learn users satisfaction and therefore bolster new information in the retrieval process. Altogether, these advantages advocate for a more sophisticated and a deeper user modeling regarding both knowledge and retrieval satisfaction.

3 Research Directions and Perspectives

Proposed Research and Challenges: Directions for the Community and Future PhD Topics. Among the key research directions and challenges to be addressed in the next 5-10 years in order to advance conversational search as a more capable learning technology are the following:

- Transforming existing IR knowledge graphs into richer multi-dimensional ones that currently are used in multimodal analytic research which supports integrating information from multiple modalities (e.g., speech, writing, touch, gaze, gesturing) and multiple levels of analyzing them (e.g., signals, activity patterns, representations).

- Integration of multimodal input and multimedia output processing with existing IR techniques

- Integration of more sophisticated user modeling with existing IR techniques, in particular ones that enable identifying the users current expertise level in the content domain

that is the focus of their search and leveraging the span of the search session.

- Conversely, integrating analytics that enable the user to identify the authoritativeness of an information source (e.g., its level of expertise, its credibility or intent to deceive).

- Development of more advanced multimodal machine learning methods that go beyond audio-visual information processing and search. Development of more advanced machine learning methods for extracting and representing multimodal user behavioral models.

Broader Impact. The research roadmap outlined above would result in major and consequential advances, including in the following areas:

- More successful IR system adaptivity for targeting user search goals.

- IR systems that function well based on fewer and briefer interactions between user and system.

- IR system that are more reliable and robust at processing user queries. Expansion of the accessibility of IR technology to a broader population.

- Improved focus of IR technology on end-user goals and values, rather than commercial for-profit aims.

- Improvement of powerful machine learning methods for processing richer multimodal information and achieving more deeply human-centered models.

- Acceleration of the positive impact of lifelong learning technologies on human thinking, reasoning, and deep learning.

Obstacles and Risks.

- Establishing and integrating more deeply human-centered multimodal behavioral models to advance IR technologies risks privacy intrusions that must be addressed in advance.

- Establishing successful multidisciplinary teamwork among IR, user modeling, multimodal systems, machine learning, and learning sciences experts will need to be cultivated and maintained over a lengthy period of time.

- Mutually adaptive systems risk unpredictability and instability of performance, and must be studied to achieve ideal functioning.

- New evaluation metrics will be required that substantially expand those used by IR system developers today.

4 Acknowledgements

References

[ADHD18] Leif Azzopardi, Mateusz Dubiel, Martin Halvey, and Jeff Dalton. Conceptualizing agent-human interactions during the conversational search process, 2018.

[ASD18]   Wafa Aissa, Laure Soulier, and Ludovic Denoyer. A reinforcement learning-driven translation model for search-oriented conversational systems. In *Proceedings of the 2nd International Workshop on Search-Oriented Conversational AI, SCAI@EMNLP 2018, Brussels, Belgium, October 31, 2018*, pages 33–39, 2018.

[AWP$^+$14] Ahmed Hassan Awadallah, Ryen W. White, Patrick Pantel, Susan T. Dumais, and Yi-Min Wang. Supporting complex search tasks. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM 2014, Shanghai, China, November 3-7, 2014*, pages 829–838, 2014.

[CTHH17] Kevyn Collins-Thompson, Preben Hansen, and Claudia Hauff. Search as Learning (Dagstuhl Seminar 17092). *Dagstuhl Reports*, 7(2):135–162, 2017.

[JL99]   Philip N. Johnson-Laird. Space to think. In L. Nadel P. Bloom, M. Peterson and M. Garrett, editors, *Language and Space*, pages 437–462. The MIT Press, Cambridge MA., 1999.

[Mar06]   Gary Marchionini. Exploratory search: from finding to understanding. *Commun. ACM*, 49(4):41–46, 2006.

[OC15]   Sharon L. Oviatt and Philip R. Cohen. *The Paradigm Shift to Multimodality in Contemporary Computer Interfaces*. Synthesis Lectures on Human-Centered Informatics. Morgan & Claypool Publishers, 2015.

[OGCO19] Sharon Oviatt, Joseph Grafsgaard, Lei Chen, and Xavier Ochoa. The handbook of multimodal-multisensor interfaces. chapter Multimodal Learning Analytics: Assessing Learners' Mental State During the Process of Learning, pages 331–374. Association for Computing Machinery and Morgan & Claypool, New York, NY, USA, 2019.

[Ovi13]   S. L. Oviatt. *The Design of Future of Educational Interfaces*. Routledge Press, New York, 2013.

[TY19]   Zhiwen Tang and Grace Hui Yang. Dynamic search - optimizing the game of information seeking. *CoRR*, abs/1909.12425, 2019.

[WR09]   Ryen W. White and Resa A. Roth. *Exploratory Search: Beyond the Query-Response Paradigm*. Synthesis Lectures on Information Concepts, Retrieval, and Services. Morgan & Claypool Publishers, 2009.