

Social competence in conversational retrieval

Paul Thomas, Mary Czerwinski, Daniel McDuff, and Nick Craswell || Microsoft

Problem

Conversational IR systems—here, meaning systems that work over multiple turns of natural language—are rapidly getting better at the basic functions of understanding searcher utterances, tracking context, and finding and presenting relevant information. However, conversation is much more than an exchange of facts: humans start learning conversational norms in utero, and unlike web UIs, conversational exchanges are laden with social signals. As humans, we cannot help but treat computers as social agents (Nass et al., 1994), so as developers and researchers we must be aware of social conventions.

A recent survey (Thomas et al., in press) has many examples, including:

- Politeness norms, which vary from culture to culture, but which constrain what we say to whom and when. Leech suggests a “politeness principle”; Brown and Levinson have rules for “face” (see Watts, 2003, for a summary). We have not yet understood what it is for an IR agent to be “polite”, nor formalised any design rules to help.
- Non-verbal behaviours, which can carry as much information as the words used (Scherer and Ceschi, 2020) and which are important for feeling an agent is “helpful” (McDuff et al., 2017), are again not well studied in conversational IR.
- Style and alignment are important for feelings of empathy and efficiency, in talking with people (Tannen, 1987; Brennan, 1996) and in talking with machines (Dubiel et al., 2020; Pickering and Garrod, 2004; Branigan et al., 2010). We have seen this in our own work on conversational retrieval as well (Thomas et al., 2018).

Conversational agents are rapidly getting more useful, and it is time to think about going beyond “mere” correctness to the full gamut of conversational phenomena.

Goal

An overall goal is simple to state, if hard to get to: a conversational IR agent should not only be competent at retrieving information, formulating coherent utterances, and managing an exchange of information; it should also be competent in at least the basic social nuances of conversation (Reeves, 2010).

This means considering the (sizable!) body of work from linguistics, psychology, philosophy, and computer science that analyses social phenomena in conversation, and that demonstrates and probes social effects in human:agent as well as human:human settings. It also means deciding which of these phenomena apply (or should apply) in the agent’s setting, and designing accordingly.

State of the art

We do already know a great deal about human-to-human conversation, including anthropomorphism, “personality”, social norms, and social signalling beyond simple question-answering. There are some good examples outside of IR, for example, in conversational recommendation systems (Walker et al., 1997) or in-car systems (Stier et al., 2020). To the best of our knowledge, however, no one is yet designing conversational retrieval systems for this sort of competence.

We can also develop guidelines for conversational IR systems, based on social norms and what we learn from the work above. For example, Gnewuch et al. (2017) have design rules based on Grice’s

well-known maxims for conversation (quantity, quality, relation and manner), but to our knowledge these have not been well tested.

Steps

As well as the general goal above, we can suggest some concrete steps:

1. IR researchers should be aware of social phenomena that may apply to their conversational systems; what it would mean if these phenomena were seen in this new context; and how their designs might change as a result.
2. Agents should be designed with effective expression of emotion, as well as of fact. This might be through facial expressions and body gestures, through style matching speech, or through careful selection of text.
3. We should consider how to build an agent which can track, and adapt, to conversational style.
4. We should consider when “embodied” agents—those which present e.g., a “head” or “body” plus gestures via each—might be useful, to provide a richer set of social cues.
5. IR researchers should collect and use corpora which include data on aspects such as tone and rate of speech, lexical variation, facial and body gestures, and other channels, beyond just text. These channels are important for understanding social norms and signalling.
6. Guidelines such as Gnewuch et al.’s should be tested in an IR setting, and as a community we need to develop something similar for our own systems.

Importantly, past research on conversation wasn’t carried out with computers in mind. We should also consider the possible drawbacks of applying this research here. Are there opportunity costs or particular risks with regards to ethics, privacy and security?

Select references

Holly P. Branigan, Martin J. Pickering, Jamie Pearson, and Janet F. McLean. 2010. Linguistic alignment between people and computers. *J. Pragmatics* 42 (2010), 2355–2368.

Susan E Brennan. 1996. Lexical entrainment in spontaneous dialog. In *Proc. International Symposium on Spoken Dialogue*.

Mateusz Dubiel, Pilar Oplustil, Martin Halvey and Simon King. 2020. Persuasive synthetic speech: Voice perception and user behaviour. In *Proc. Conversational User Interfaces*.

Daniel McDuff, Paul Thomas, Mary Czerwinski, and Nick Craswell. 2017. Multimodal analysis of vocal collaborative search: A public corpus and results. In *Proc. Multimodal Interaction*.

Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proc. SIGCHI Conference on Human Factors in Computing Systems*, 72–78.

Martin J. Pickering and Simon Garrod. 2004. *Toward a mechanistic psychology of dialogue*. Behavioral and Brain Sciences 27, 2 (2004), 169–225.

Byron Reeves. 2010. The benefits of interactive, online characters. *The Madison Avenue J.*

Klaus R Scherer and Grazia Ceschi. 2000. Criteria for emotion recognition from verbal and nonverbal expression: Studying baggage loss in the airport. *Personality and social psychology bulletin* 26, 3 (2000), 327–339.

Daniela Stier, Katherine Munro, Ulrich Heid, and Wolfgang Minker. 2020. Towards Situation-Adaptive In-Vehicle Voice Output. In *Proc. Conversational User Interfaces*.

Deborah Tannen. 1987. Conversational style. In *Psycholinguistic models of production*, Hans W Dechert and Manfred Raupach (eds.). Ablex, Norwood, NJ.

Paul Thomas, Mary Czerwinski, Daniel McDuff, Nick Craswell, and Gloria Mark. 2018. Style and alignment in information seeking conversation. In *Proc. Conference on Human Information Interaction and Retrieval*, pp42–51.

Paul Thomas, Mary Czerwinski, Daniel McDuff, and Nick Craswell. In press. Theories of conversation for conversational IR. To appear in *ACM Transactions on Information Systems*.

Marilyn A Walker, Janet E Cahn, and Stephen J Whittaker. 1997. Improvising linguistic style: Social and affective bases for agent personality. In *Proc. Autonomous Agents*, pp96–105.

Richard J Watts. 2003. *Politeness*. Cambridge University Press, Cambridge.