

# the morning paper

a random walk through Computer Science research, by Adrian Colyer

Made delightfully fast by

---

≡ MENU

---

## Ten challenges in highly-interactive dialog systems

JULY 8, 2016 ~ ADRIAN COLYER

[Ten challenges in highly-interactive dialog systems](#) Ward & De Vault AAI 2015

It's time to end our look into the technology behind chatbots and dialog systems – there's a huge crop of SIGMOD 2016 papers waiting to be explored for starters! To end the mini-series, today I've chosen a 2015 position paper from Ward and De Vault detailing ten key challenges (and opportunities) they see for future interactive dialog systems. The emphasis here is on spoken dialog, but there's something for everyone. The challenges are broken down into three main areas: improving power and robustness; reducing development costs; and deepening understanding.

### Improving power and robustness

The opening challenge certainly seems like a stretch on the surface: it is to not just match human performance in spoken dialog, but to *exceed* it! At first glance that doesn't seem possible – don't humans *define* what it means to converse naturally? And it certainly seems quite some way away given the current state-of-the-art we've been looking at. But the authors make a compelling case for "superhuman" dialog:



This possibility can be appreciated by listening to recordings of people in live conversation (especially yourself) and noting how inefficient and sometimes ineffective they are. While some disfluencies and awkwardnesses can be functional, to convey nuances or adjust the pace for the sake of the interlocutor, many are just regrettable. This is obvious in retrospect, when one can replay the recording to glean every detail of the interlocutors' behavior and can take the time to think about what should have been said and how, at each point in

time. Future dialog systems, not subject to human cognitive limitations, might be able to do this in real time: to sense better, consider more factors, plan dialog steps further ahead, and so on, to attain superhuman charm and produce superhuman dialog.

The second challenge is to layer different personalities or behaviour styles on top of the same basic functionality, and the third challenge is to enrich interaction behaviour. A good example of this is [backchanneling](#) (feedback given while someone else is talking “ok”, “yeah”, “uh-huh” and so-on).



Consider for example backchanneling behavior. This is a prototypical interactive behavior, probably the best studied one, and one already used successfully in several systems which backchannel at appropriate times in response to user speech....

In human-human dialog a 2012 study by Ward et al. uncovered 12 different activity types strongly affecting the probability of a back-channel (rambling, expressing sympathy, etc.). This 12 types all had independent manifestations.



In general, there is a lot more going on in human interaction than we are modeling today.

Challenge number four is to “synthesize multifunctional behaviours” – by which the authors mean that instead of making independent decisions about (for example), whether to back-channel, what back-channel word to use, and which [prosodic](#) form to use, it may be better to make these decisions jointly and optimize them together.

The next challenge is to integrate both learned and designed behaviours.



Today even the most interactive systems have a fixed skeleton specifying the overall dialog flow. Within this a few decision points may be left underspecified, for subsequent filling in with more data-driven decision rules. While ultimately dialog system behaviors might be entirely learned from data, for the foreseeable future interactive systems will include both learned and designed behaviors... We see an opportunity to explore new ways of integrating learned and designed behaviors, and in particular to develop architectures which give a larger role to behaviors learned from corpora.

The final challenge in this section applies to spoken dialog systems, and concerns the evolution of dialog state tracking from a discrete turn-based process to a *continuous* process. "People continuously track the current state of the dialog, not only when the other is speaking, but when speaking themselves... while implementing continuous state tracking won't be easy, the potential value is significant." This tracking can include elements such as gaze and gestures. You can capture some of the importance of this in the difference between a conference call and an in-person conversation.

## Reducing development costs

There are two challenges presented in this section. Firstly, the ability to compose behaviour specifications (and hence reuse them).



For example, imagining that we have developed a general policy for choosing the next interview question, a general policy for showing empathy, and a general policy for supportive turn taking, we could imagine that these could be composed to produce a system capable of effective, natural, and warm first-encounter dialogs.

The second challenge is to figure out ways to bypass the dearth of labelled training data and increase our ability to use completely unsupervised methods.

## Deepening understanding

The final two challenges are to develop evaluation methods that give more actionable feedback on the performance of different parts of the system and its conversational style, and to find a way to engage social scientists and their findings in the development of dialog systems:



The behaviors in today's dialog systems are seldom based on the findings of social scientists, and conversely, the results of dialog systems research are rarely noticed by them. One reason is that the most interactive aspects of dialog systems are often not fully understandable: they may work, but it is hard to know why. There is a need for more comprehensible models. Ways to achieve this might include deeper analysis of what a learned model really has learned, more use of modeling techniques which are intrinsically more understandable, and more use of declarative representations of behaviors rather than decision algorithms.

POSTED IN [UNCATEGORIZED](#)

[MACHINE LEARNING](#)

---

[< PREVIOUS](#)

*Multi-domain dialog state tracking using recurrent neural networks*

[NEXT >](#)

*Realtime data processing at Facebook*

---



Start a conversation ...



[Terms of use](#) - [Privacy](#) - [Report a bug](#)

powered by [GraphComment](#)

---