Do we really need a conversational interface? Some challenges in designing for conversational interaction

Michael McTear
Ulster University
Northern Ireland, UK
mf.mctear@ulster.ac.uk

Conversational Interaction Conference, San Jose, CA, USA, February 10-11, 2020
Conversational interfaces are all around us
There's a lot going on in the world. I can send you daily updates with top stories, or you can message me directly about a topic you're interested in, like "politics" or "tech," and I'll send you the latest.

Ok, let's get started. Here are some options:

- Daily update
- Breaking news

Type a message...
What is a conversational interface?

A conversational interface allows users to interact with services and devices in a **conversational style**

i.e. by taking turns in a dialog to accomplish a task or engage in a conversation

This provides a different (possibly easier) and more intuitive way to do things e.g. simply ask a query as opposed to navigating menus and drop-down boxes.
Conversational interfaces – the bad news!!
“... CUI is the worst thing that has happened to usability since MS-DOS”

Alex Mohebbi Why conversational interfaces are taking us back to the Dark Ages of usability
Intelligent Assistants Have Poor Usability

Usability testing finds that both voice-only and screen-based intelligent assistants work well only for very limited, simple queries that have fairly simple, short answers.

Raluca Budiu and Page Laubheimer, Nielsen Norman Group. July 22, 2018
Observing users struggle with the AI interfaces felt like a *return to the dark ages of the 1970s*: the need to memorize cryptic commands, oppressive modes, confusing content, inflexible interactions — basically an unpleasant user experience.
Part of the misfire with the conversational aspect of bots has to do with the fact that natural language processing and artificial intelligence are not yet accomplished at managing human-like conversations.
The technology of conversational interfaces
Modular architecture

- Speech Recognition
- Natural Language Understanding
- Text to Speech Synthesis
- Natural Language Generation
- Dialog Manager
  - Dialog Policy
  - Dialog State Tracking
Natural Language Understanding (NLU)

Current NLU engines e.g. Dialogflow, Alexa, etc support the Intent-Entity-Context-Response (IECR) paradigm

Combination of AI/Machine learning and hand-crafting

Can the system understand if the user’s input is unconstrained as in natural conversation?

Many systems avoid NLU problems by using pre-determined user responses (quick replies)
  – easy to interpret but do not give the user a full choice of response
Dialog Management / Conversation Flow

One-shot queries

- Similar to voice search
- User controls the dialog
- User has to include all required information in the query

User: What’s the weather forecast for Frankfurt for tomorrow
System: In Frankfurt tomorrow it will be mostly sunny with a high of 23 and a low of 9

User: What about Belfast
System: In Belfast tomorrow there will be scattered showers with a high of 17 and a low of 12
Slot-filling
The system asks questions to collect the values for a series of slots that are required in order to complete the task

Book_flight
To: unknown
From: unknown
Date: unknown
Time: unknown
Conversation flow diagram

The dialog follows a pre-determined path

What happens if the user or the system want to deviate from the path?
Deviating from the path

The system inserts a query about the user’s input because something is missing or ambiguous

User: I am looking for a restaurant nearby
System: What kind of food would you like?

User: What time does the bus leave for San Jose?
System: Do you mean the express bus or the regular one?
User requests repetition or clarification, changes or corrects something, ...

System: what kind of food would you like?
User: what are my choices?
The system learns a dialog policy – which action to take at each choice point in the dialog

Example - Deciding whether to confirm a user’s input or move on

The system learns an optimal policy based on feedback from multiple interactions

Problem that policy optimization quickly becomes intractable
Neural dialog – end-to-end, seq2seq

- Natural Language Understanding
- Response Generation
- Neural Model
Some issues with deep learning for conversational systems

Requires large amounts of training data but there is not much suitable training data available

Limited use of context – applies only to the immediate history of the current utterance

Lack of explicit long-term memory – leads to semantic inconsistency in responses

Tendency to produce bland and uninformative utterances that have higher probability in the training data e.g. ‘that’s great’, ‘I don’t know’
Symbolic AI – example of plan recognition for collaborative dialogue

(Scenario: it is 10pm, user wants to catch the next bus to Suwon)

User: Do you know when is the next bus to Suwon?

Yes

The next bus leaves tomorrow at 5am

Sorry, the last bus has left for the evening. You will have to take a bus to Seongnam and then transfer to the bus to Suwon. The bus leaves here from bay number 6 at 11pm.
To provide the third answer the system has to infer the user’s high level goal of wanting to be in Suwon that evening.

Technologies required for collaborative dialog:

- Representation of plans and intentions
- Reasoning over plans and intentions
- Ontologies to store background knowledge

See research on advanced dialog systems at Rochester (James Allen) and Monash (Phil Cohen)
Modelling the sequential mechanics of conversation

Natural Conversation Framework (NCF)
• a design framework for conversation

Library of generic conversational UX patterns
• based on natural patterns from human conversation as documented in the Conversation Analysis literature
Takeaways

There is a lot of activity in Conversational AI but there is a long way to go

“Conversational AI is one of the most challenging problems in the artificial intelligence field ... it is still Day One for Conversational AI”

Research is currently focused on deep learning approaches combined with deep reinforcement learning

Need to integrate neural and symbolic approaches (neuro-symbolic AI)
Takeaways

Most work on conversational agents focuses on issues of design and implementation (technology driven)

Need to focus also on co-creation and user involvement (UX driven)

We need to know how the systems are actually used e.g. if the interactions are successful, what users do when they are not, what did users think of the interactions
Thank you for listening