End-to-end conversational agents: what's missing?

Marco Baroni
Facebook Artificial Intelligence Research

Let's Discuss Workshop at NIPS 2016
With many thanks to...

• Angeliki Lazaridou, Alex Peysakhovich
• Gemma Boleda
• Jason Weston, Marc'Aurelio Ranzato, Sumit Chopra, Antoine Bordes
• Tomas Mikolov, Armand Joulin, Germán Kruszewski
• Raquel Fernandez
• Aurelie Herbelot
• ...

Why is conversation so easy?

Simon Garrod¹ and Martin J. Pickering²

¹University of Glasgow, Department of Psychology, Glasgow, UK, G12 8QT
²University of Edinburgh, Department of Psychology, Edinburgh, UK, EH8 9JZ
The conversational agent pipeline

1. Input speech
2. Spoken Language Understanding (SLU)
3. Natural Language Generation (NLG)
4. State Estimator
5. Policy
6. System Response

User dialogue act at each turn is determined by a second stochastic function represented by a suitable stochastic model, called a policy. The system regards the output of the SLU as a noisy observation reflecting the uncertainty in the interpretation of user utterances; instead, at each turn, the system regards the output of the SLU as a Markov process, i.e., starting in some initial state $s_0$, the dialogue state $s_t$ is updated and a deterministic decision rule is used to select the dialogue act $a_t$. The dialogue and the attribute values (often called slots) that represent states and actions, and arcs representing user inputs, are determined by belief monitoring, policy optimisation, and reinforcement learning.
Anthes 2010

Sutskever et al. 2014
Training end-to-end machine translation

Io sono il dottore di cui in questa novella si parla talvolta con parole poco lusinghiere.

I am the doctor occasionally mentioned in this story, in unflattering terms.
More recently, work by Sordoni et al. (2014) shows that recurrent neural networks are rather effective models for natural language. Other tasks such as parsing (Jurafsky & Martin, 2004), image captioning (Vinyals et al., 2014b), and speech recognition (Huang et al., 2012) have been demonstrated. It has also been used for retrieval (Wang et al., 2011), machine translation (Sutskever et al., 2014b), and even for predicting chemical properties from molecular structures (Gilmer et al., 2017). However, most of these systems require a rather complicated processing pipeline of many steps, each of which is itself a supervised learning task. Therefore, fully unsupervised learning is out of the scope of this paper to provide an exhaustive comparison with any other system.


to maximize the cross entropy of the correct sequence given its context. During inference, given that the true output sequence is provided, the model is trained using backpropagation. The model is trained on a dataset of conversations, with each conversation consisting of multiple turns. At each turn, the input sequence is the context (previous sentences) and the output sequence is the reply (next sentence). The lack of a model to ensure consistency between turns is another obvious limitation. Unlike easier tasks like translation, however, a model like sequence-to-sequence will not be able to successfully "solve" the problem of modeling dialogue due to several reasons. The strength of this model lies in its simplicity and generality. We can use this model for machine translation, question answering, and conversations without major changes in the architecture. Applying this technique to conversational modeling is also straightforward: the input sequence can be the concatenation of what has been conversed so far (the context), and the output sequence is the reply.

The hidden state of the model when it receives the end of sequence symbol "<eos>" above. The hidden state of the model when it receives the end of sequence symbol "<eos>" above.

*End-to-end conversational agents*
Training end-to-end conversational agents

A: Shall we call Alan Turner?
B: Alan Turner?
A: Yeah.
B: I haven't got his number have you got it?

http://bncweb.lancs.ac.uk/
Training end-to-end conversational agents

A: Shall we call Alan Turner?

B: Alan Turner?

A: Yeah.

B: I haven't got his number have you got it?
End-to-end conversational agents

H: i am seeing an error related to vpn

M: what is the error message that you are getting when connecting to vpn using network connect?

H: connection refused or something like that

M: try this, goto <URL> page and enter your name and password

H: done

M: try logging in with <NAME> and check if it's working fine now

H: yes, now it works!

Vinyals and Le 2015
End-to-end conversational agents

H: what do you think about messi?
M: he’s a great player.

H: what do you think about cleopatra?
M: oh, she’s very regal.

H: what do you think about england during the reign of elizabeth?
M: it was a great place.

Vinyals and Le 2015
End-to-end conversational agents

H: what is your job?
M: i’m a lawyer.
H: what do you do?
M: i’m a doctor.
End-to-end conversational agents

H: what are you doing?
M: nothing.

H: what is your name?
M: i don’t know, sir.

H: how old are you?
M: i’m not sure.
Conversation as a joint project
Clark 1996

• **(Extra-linguistic) context:** a conversation will take place in a certain state of the world

• **Common ground:** the participants must coordinate on a (dynamic) set of "mutual knowledge, mutual beliefs, and mutual assumptions"

• **Purpose:** we engage in a conversation in order to achieve something
Context

A: Shall we call Alan Turner?

B: Alan Turner?

A: Yeah.

[B knows Alan Turner number]

B: I haven't got his number have you got it?
Context

A: Shall we call Alan Turner?

B: Alan Turner?

A: Yeah.

[Alan Turner's number is 333 454443]

B: His number is 911
Accessing extra-linguistic context

• Teach machine to query a data-base (Bordes and Weston 2016):

  A: Shall we call Alan Turner?
  B: Alan Turner?
  A: Yeah.
  B: \texttt{db_search(Alan\_Turner, Has\_Number, \_\_\_\_)}
  DB: 333 454443
  B: Alan's number is 333 454443
Accessing extra-linguistic context

Das et al. 2016
Accessing extra-linguistic context: problems and solutions

• Subject-elicited data: small
• Scripted interactions: limited, artificial

• Mix large-scale purely-text-based training and small/artificial real-world-based training (how?)
• Clever ways to collect grounded data (games?)
• Grounded data in the wild (movies?)
How to use contextual information

Contextual vs generic

- What's the color of cows?
- What's the color of this cow?

Precise vs vague

- Where does Alan Turner live?
- 34 Washington Street / 35 Lincoln Avenue / In the US
- In the US / In the United States / In Winesburg, Ohio / 34 Washington Street
Conversation as a joint project
Clark 1996

• (Extra-linguistic) context: a conversation will take place in a certain state of the world

• Common ground: the participants must coordinate on a (dynamic) set of "mutual knowledge, mutual beliefs, and mutual assumptions"

• Purpose: we engage in a conversation in order to achieve something
A: Shall we call Alan Turner?
B: Alan Turner?
A: Yeah.
B: I haven't got his number have you got it?
A: Shall we call Alan Turner?

B: Alan Turner?

A: Yeah.

B: I haven't got his number have you got it?
Common ground

Leslie: Gotta see them next time.
Leslie: See, I was a bit shocked, you know
Leslie: Mike said he went to Chessington, was it?
Rob: Yeah.
Leslie: Ha ha!
Leslie: Why didn't he do that when we were here?

http://bncweb.lancs.ac.uk/
Common ground in end2end systems

H: i am seeing an error related to vpn

M: what is the error message that you are getting when connecting to vpn using network connect?

H: connection refused or something like that

M: try this, goto <URL> page and enter your name and password

H: done

M: try logging in with <NAME> and check if it's working fine now

H: yes, now it works!
Common ground in end2end systems

H: what is your job?
M: i’m a lawyer.
H: what do you do?
M: i’m a doctor.

Vinyals and Le 2015
Common ground and conversation partners

Rob: Let me introduce you to my friend Mike...
Machine: Nice to meet you!

... 

Machine: Mike said he went to Chessington, was it?
Rob: Yeah.

Machine: I know this person, Mike, who went to Chessington.
Alice: Yeah
Personas
(Li et al. 2016b)
Negotiating common ground

Alan: In fact it was the Society's policy now that we won't put animals to sleep unless there is an extreme cause for that...

James: Why is it that in the north of England, particularly in this region of the north of England, we seem to be worse than anywhere else?

Alan: I don't think you are. You know, since I've been down in Horsham, I have found that the northeast is not alone. There are many other areas in the country where animals are seriously abused. The northeast isn't the only place.

http://bncweb.lancs.ac.uk/
Updating common ground

A: Shall we call Alan Turner?
B: Alan Turner?
A: Yeah.
B: I haven't got his number have you got it?

... 

A: Shall we call our... his name isn't Alan Turner.
B: Well what's his name?
A: His name is... Richard... Thorpe, which is why am I looking under T in my book...
Conversation as a joint project
Clark 1996

• **(Extra-linguistic) context:** a conversation will take place in a certain state of the world

• **Common ground:** the participants must coordinate on a (dynamic) set of "mutual knowledge, mutual beliefs, and mutual assumptions"

• **Purpose:** we engage in a conversation in order to achieve something
The purpose of conversations

Task-oriented conversations
A: I'd like to book a restaurant for tonight.
B: What kind of food?
A: Indian, perhaps?
B: How about Shahi Tandoor on George Street?
A: That sounds great, thank you!

Chitchat?
A: Shall we call Alan Turner?
B: Alan Turner?
A: Yeah.
B: I haven't got his number have you got it?
The purpose of conversations

A: Shall we call Alan Turner?
B: Alan Turner?
A: Yeah.
B: I haven't got his number have you got it?
The purpose of conversations

B: Before the end of the show, we've got ten minutes left...

A: Shall we call Alan Turner?
B: Alan Turner?
A: Yeah.
B: I haven't got his number have you got it?

A: No... hey, have you seen they're filming in Harewood this week?
Conversational success is inherently rewarding

Brain-to-brain coupling: a mechanism for creating and sharing a social world

Uri Hasson\textsuperscript{1,2}, Asif A. Ghazanfar\textsuperscript{1,2}, Bruno Galantucci\textsuperscript{3,4}, Simon Garrod\textsuperscript{5,6} and Christian Keysers\textsuperscript{7,8}
Learning and transferring linguistic cues of success
See also Li et al. 2016c

<table>
<thead>
<tr>
<th>Explicit reward</th>
<th>Implicit reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: I'd like to book a restaurant for tonight.</td>
<td>A: Shall we call Alan Turner?</td>
</tr>
<tr>
<td>B: What kind of food?</td>
<td>B: Alan Turner?</td>
</tr>
<tr>
<td>B: How about Shahi Tandoor on George Street?</td>
<td>B: Here's his number: 333 454443</td>
</tr>
<tr>
<td>A: That sounds great, thank you!</td>
<td>A: Great, thank you!</td>
</tr>
</tbody>
</table>

+1
Conversation as a joint project
Clark 1996

- **(Extra-linguistic) context:** a conversation will take place in a certain state of the world

- **Common ground:** the participants must coordinate on a (dynamic) set of "mutual knowledge, mutual beliefs, and mutual assumptions"

- **Purpose:** we engage in a conversation in order to achieve something
A: Shall we call Alan Turner?

B: Well, this is the last slide, we ran out of time again!
• Anthes. 2010. Automated translation of Indian languages. Comm ACM.
• Clark. 1996. Using language. CUP.
• Garrod and Pickering. 2004. Why is conversation so easy. TICS.
• Sutskever et al. 2014. Sequence to sequence learning with neural networks. NIPS.
• Vinyals and Le. 2015. A neural conversational model. ICML DLW.
• Young et al. 2013. POMDP-based statistical spoken dialogue systems: A review. Proc IEEE.