Welcome
2017 Faculty Summit Attendees

Faculty Summit 2017
microsoftfacultysummit.com

Microsoft Research
Microsoft.com/research

Facebook
@microsoftresearch

Twitter
@MSFTResearch
#FacSumm
#EdgeofAI
End-to-End Trainable Neural Network Models for Task Oriented Dialog

Ian Lane
Carnegie Mellon University
Trainable End-to-End Models for Dialog

- Can we learn to perform task-oriented dialog by modeling human conversations?
Trainable End-to-End Models for Dialog

Can we learn to perform task-oriented dialog by modeling human conversations?

User

I’m looking for an Italian Restaurant for Dinner.

Agent
Trainable End-to-End Models for Dialog

➢ Can we learn to perform task-oriented dialog by modeling human conversations?

User

I’m looking for an Italian Restaurant for Dinner.

Agent

Sure I can help you with that. Whereabouts are you thinking?
Trainable End-to-End Models for Dialog

Can we learn to perform task-oriented dialog by modeling human conversations?

User

I’m looking for an Italian Restaurant for Dinner.

Agent

Sure I can help you with that. Whereabouts are you thinking?

Somewhere in Menlo Park or Palo Alto.
Trainable End-to-End Models for Dialog

Can we learn to perform task-oriented dialog by modeling human conversations?

User

I’m looking for an Italian Restaurant for Dinner.

Somewhere in Menlo Park or Palo Alto.

Agent

Sure I can help you with that. Whereabouts are you thinking?

Well there is a couple of good places ...
Trainable End-to-End Models for Dialog

Can we learn to perform task-oriented dialog by modeling human conversations?

**User**

I’m looking for an Italian Restaurant for Dinner.

Somewhere in Menlo Park or Palo Alto.

**Agent Model**

Agent

Knowledge base

Output 1

Output 2

Carnegie Mellon University
End-to-End Models for Dialog

A Neural Conversational Model

Human: hello!
Machine: hello!
Human: how are you?
Machine: i’m good.
Human: what’s your name?
Machine: i’m julia.
Human: when were you born?
Machine: july 20th.
Human: what year were you born?
Human: where are you from?
End-to-End Models for Dialog

A Neural Conversational Model

Machine: hi
Human: hi
Machine: could you please let me know what is the operating system you are using?
Human: linux
Machine: ok this is the issue with the system performance, nasser, tools.
Human: what is that?
Trainable Task-Oriented Models for Dialog

**Agent**

- Understand user’s input in context
- Take **Action** to take given user input, dialog and task history
  - Respond to user?
  - Perform API call / database look up?
  - Perform other action?
  - Or combination of above.
- Update dialog state

**User**

- Estimate of **user** goal / goal change
An Neural Network Model for Dialog Agents

- **Agent Model**
  - Perform de-lexicalization on input (i.e. Named Entity Recognition)
  - Update dialog state (via. LSTM model)
  - Estimate beliefs state
  - **Action**: Issue API Call and process results (if required)
  - **Action**: Update reference to results (if required)
  - **Action**: Response generation
    - Estimate de-lexicalized response that best matches current state from all possible responses
End-to-End Neural Network Models for Dialog
Training - Approach

- **Manual Annotation of Training Data Not Required**
  - Assumes NLU for delexicalization and named entity detection
  - Train on dialogs from human-to-human chat interaction
  - Assumes turn-based interaction → can be extend to also model timing

- **Supervised Training**
  - Treat as a classification task per turn → Select best output given history

- **Reinforcement Learning**
  - Optimize based on end-of-dialog reward
Training – Results and Challenges

- **Prediction Accuracy**

<table>
<thead>
<tr>
<th>Model</th>
<th>Entity Pointer</th>
<th>Joint Goal</th>
<th>De-lex Res</th>
<th>Final Res</th>
</tr>
</thead>
<tbody>
<tr>
<td>BoW Emb Encoder</td>
<td>93.5</td>
<td>72.6</td>
<td>55.4</td>
<td>51.2</td>
</tr>
<tr>
<td>+ word2vec</td>
<td>93.6</td>
<td>74.3</td>
<td>55.9</td>
<td>51.5</td>
</tr>
<tr>
<td>Bi-LSTM Encoder</td>
<td>93.8</td>
<td><strong>77.2</strong></td>
<td>55.8</td>
<td>52.6</td>
</tr>
<tr>
<td>+ word2vec</td>
<td><strong>94.4</strong></td>
<td>76.6</td>
<td><strong>56.6</strong></td>
<td><strong>52.8</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Per-res Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory Networks [17]</td>
<td>41.1</td>
</tr>
<tr>
<td>Gated Memory Networks [29]</td>
<td>48.7</td>
</tr>
<tr>
<td>Sequence-to-Sequence [22]</td>
<td>48.0</td>
</tr>
<tr>
<td>Query-Reduction Networks [23]</td>
<td>51.1</td>
</tr>
<tr>
<td>Prediction Accuracy</td>
<td><strong>52.8</strong></td>
</tr>
</tbody>
</table>

- **Issues**

  - Even if action taken by agent at specific turn does not match training “training data” it may still be *appropriate*
  
  - “training data” is fixed, User does not respond based on output of Model

  ➔ Can we simulate complete USER-AGENT interactions to improve dialog?
I’m looking for an Italian Restaurant for Dinner. Somewhere in Menlo Park or Palo Alto.

Sure I can help you with that. Whereabouts are you thinking?

Well there is a couple of good places ...
Joint Modeling of Users and Agents

Can we jointly model task-oriented dialog to simulate human conversations?

User Model

Agent Model

I’m looking for an Italian Restaurant.

Sure I can help you with that. Whereabouts are you thinking?

Somewhere in Menlo Park or Palo Alto.

Well there is a couple of good places …
Joint Modeling of Users and Agents

User Model

Agent Model

Prezzo is a nice place in the west of town serving tasty Italian food

Ranked KB query results

1. La_margherita ...
2. Prezzo ...
5. Caffe_uno ...

User goal encoding
Italian, west, address

User input encoding
Request alternative

Agent input encoding
Offer entity No. 2

Agent output utterance

Agent action output
make_offer

LSTM
Joint Modeling of Users and Agents

User Model

Agent Model
Joint Modeling of Users and Agents

**User Model**

**Agent Model**

NLG Network

Agent output

Prezzo is a nice place in the west of town serving tasty Italian food

Ranked KB query results

1. La_margherita ... 
2. Prezzo ... 
5. Caffe_uno ...

Agent action output

LSTM

User output

may I have its address?

NLG Network

User action output

req_addr

User goal encoding

Italian, west, address

User input encoding

Request alternative

Offer entity No. 2

User input encoding

Italian, west, address

Agent input encoding

Offer entity No. 1

Agent input encoding

Request alternative

Offer entity No. 2

KB indicator (0/1)

1 (yes)
Joint Modeling of Users and Agents
Results (DSTC 2 Task)

Dialog Success Rate

- Significantly higher dialog success rate using jointly optimized models (65%) than optimizing AGENT only (50%), or baseline trained via supervised learning (35%)
Agent: Welcome to the restaurant booking system. How may I help you?
User: I would like expensive Portuguese food.
Agent: What part of town do you have in mind?
User: I don’t care.
Agent: (issue-api-call portuguese dontcare expensive)
User: (expect-api-results)
Agent: Sorry there is no Portuguese restaurant in the expensive price range.
User: I would like Italian food.
Agent: (issue-api-call italian dontcare expensive)
User: (expect-api-results)
Agent: Frankie and Bennys is a great restaurant serving Italian food and it is in the expensive price range.
User: May I have its address and phone number?
Agent: Frankie and Bennys is on Cambridge Leisure Park Clifton Way Cherry Hinton. The phone number of Frankie and Bennys is 01223 412430.
User: Thank you. Goodbye.
Conclusions and Challenges

• **Training task oriented dialog systems directly from human conversations seems plausible**
  • End-To-End trainable give logs of dialog + API calls
  • Well suited for task-oriented dialogs that include social, task-orientated dialog and external knowledge base look-up or actions
  • i.e. Product and service support (Call-Centers / Messaging Interfaces)

• **Ability to model both Agent and User in a conversation**
  • With data can model personas or individual Agents or Users

• **Exploring boot-strapping dialog systems ”on-the-fly”**
  • Generate suggested response ... see if Agent follows suggestion or not
  • Update models for next utterance / dialog
Thank you