Task-oriented Neural Dialogue Systems

TTIC Young Researcher Seminar, 09/11/2016
Tsung-Hsien (Shawn) Wen
Outline

- Intro
- Neural Dialogue System
- Wizard-of-Oz Data Collection
- Experiments
- Conclusion
Outline

- **Intro**
- Neural Dialogue System
- Wizard-of-Oz Data Collection
- Experiments
- Conclusion
The two paradigms

M: Hello, welcome to the Cambridge dialogue system, What kind of food would you like?
H: Yeah I want to find a restaurant that serves European food.
M: Hotel du Vin and Bistro is a nice place. It serves European food.
H: Uh what is the address and phone number?
M: Hotel du Vin and Bistro is a nice place. Their address is XXX and their phone is YYY.
H: Thank you goodbye.
M: Thank you for using this system.

A: How old are you?
B: I’m 16. Why are you asking?
A: I thought you were 12.
B: What made you think so?
A: I don’t know what you are talking about.
B: You don’t know what you are saying.

[Su et al, 2016]

[Li et al, 2016]
The two paradigms

- **Task-Oriented Dialogue Systems**
  - Goal-oriented
  - Require precise understanding, hard to collect data.
  - Modular, highly handcrafted, restricted ability, but meaningful/useful systems.

- **Chat-based Conversational Agents**
  - Chit-chat (non-goal).
  - Vast amount of data (but probably not helpful).
  - End-to-end, highly data-driven, but meaningless/inappropriate responses, unreliable systems.

- Can we train a useful (complete tasks) dialogue system directly from data?
- How can we collect the data to train this model?
Outline

- Intro
- Neural Dialogue System
- Wizard-of-Oz Data Collection
- Experiments
- Conclusion & Discussion
Traditional Dialogue Systems

- Speech Recognition
- Language Understanding
- Language Generation
- Dialogue Manager
- KB
- Web

Text flow:
- Speech Synthesis
- Dialogue System
Neural Dialogue Systems

[Diagram showing a speech recognition system, a neural dialogue system, and connections to KB and Web]

Speech Recognition

Speech Synthesis

Neural Dialogue System

text

KB

Web
Can I have Korean. 

Little Seoul serves great Korean.
Can I have <v.food> <v.name> serves great <v.food>.
Intent Network

Can I have <v.food>

Generation Network

<v.name> serves great <v.food> .

 Seq2Seq

Can I have Korean

Belief Tracker

Intent Network

Can have <v.food>

Generation Network

<v.name> serves great <v.food>

Language Grounding

Can I have Korean

Belief Tracker

MySQL query: "Select * where food=Korean"

Can I have Korean

Belief Tracker

Policy Network

Decision Making

Intent Network

Generation Network

MySQL query: “Select * where food=Korean”

Database Operator

Can I have Korean food?

MySQL query: "Select * where food=Korean"

Outline

- Intro
- Neural Dialogue System
- Wizard-of-Oz Data Collection
- Experiments
- Conclusion
Hi, I want a cheap Korean restaurant.

What area are you looking for?
Hi, I want a **cheap Korean restaurant**.

**What area are you looking for?**

**Somewhere in the north.**

**Little Seoul** is nice one in the north.

---

**What user wants?**

<table>
<thead>
<tr>
<th>Food</th>
<th>Korean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>Cheap</td>
</tr>
<tr>
<td>Area</td>
<td>North</td>
</tr>
</tbody>
</table>

**Search Table**

| Little Seoul | ... |

---

**Task:** Find a **restaurant**, **cheap**, **Korean**, **North**

Ask **phone number**
Wizard of Oz Data Collection

Hi, I want a cheap Korean restaurant.

What area are you looking for?

Somewhere in the north.

Little Seoul is nice one in the north.

What is the phone number?

Its phone number is 01223456789.

What user wants?

<table>
<thead>
<tr>
<th>Food</th>
<th>Korean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>Cheap</td>
</tr>
<tr>
<td>Area</td>
<td>North</td>
</tr>
</tbody>
</table>

Search Table

<table>
<thead>
<tr>
<th>Little Seoul</th>
<th>...</th>
</tr>
</thead>
</table>
Hi, I want a cheap Korean restaurant.

What area are you looking for?

Somewhere in the north.

Little Seoul is nice one in the north.

What is the phone number?

Its phone number is 0123456789.

Thank you very much, good bye.

Thank you for using the system.
Hi, I want a cheap Korean restaurant.

What area are you looking for?

Somewhere in the north.

Little Seoul is nice one in the north.

What is the phone number?

Its phone number is 01223456789.

Thank you very much, good bye.

Thank you for using the system.

What user wants?

<table>
<thead>
<tr>
<th>Food</th>
<th>Korean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>Cheap</td>
</tr>
<tr>
<td>Area</td>
<td>North</td>
</tr>
</tbody>
</table>
Wizard of Oz Data Collection

- Online parallel version of WOZ on MTurk
  - Randomly hire a worker to be user/wizard.
  - Task: Enter an appropriate response for one turn.
  - Repeat the process until all dialogues are finished.
- Example user page

**Task 02004:** You are looking for and it should serve gastropub food. You don’t care about the price range. You want to know the address.

Info Desk: Hello, welcome to the Cambridge restaurant system. You can ask for restaurants by area, price range or food type. How may I help you?
Customer: I want a gastropub food
Info Desk: There are 4 restaurants serving gastropub food, what price range do you want?

Next turn!

Customer: (Your response)

I don't care about the price range, just give me the address please.
Wizard of Oz Data Collection

Example wizard page

Info Desk: Hello, welcome to the Cambridge restaurant system. You can ask for restaurants by area, price range or food type. How may I help you?
Customer: I want a gastropub food
Info Desk: There are 4 restaurants serving gastropub food, what price range do you want?
Customer: I don’t care

Please modify the following answers based on the latest customer response:

- What does user want?
  What is the food type the user wants?  gastropub
  What is the area the user wants?  not mentioned

- What does user ask?
  Is the user asking for food type of an offered venue?  No
  Is the user asking for price range of an offered venue?  No
  Is the user asking for area of an offered venue?  No
  Is the user asking for postcode of an offered venue?  No
  Is the user asking for phone number of an offered venue?  No
  Is the user asking for address of an offered venue?  No
  Is the user mentioning any restaurant names?  No

Info Desk: (Your response)
I would recommend backstreet bistro, a great gastropub restaurant in the centre. do you want their phone number?

<table>
<thead>
<tr>
<th>Name</th>
<th>Food</th>
<th>Area</th>
<th>Price Range</th>
<th>Phone</th>
<th>Address</th>
<th>Postcode</th>
</tr>
</thead>
<tbody>
<tr>
<td>backstreet bistro</td>
<td>gastropub</td>
<td>centre</td>
<td>expensive</td>
<td>01223 308305</td>
<td>2 Storton Street City Centre</td>
<td>CB 1, 2 QA</td>
</tr>
<tr>
<td>royal standard</td>
<td>gastropub</td>
<td>east</td>
<td>expensive</td>
<td>01223 247877</td>
<td>260 Mill Road City Centre</td>
<td>CB 1, 3 NL</td>
</tr>
<tr>
<td>the cow pizza kitchen</td>
<td>gastropub</td>
<td>centre</td>
<td>moderate</td>
<td>01223 308871</td>
<td>Corn Exchange Street</td>
<td>CB 2, 3 QF</td>
</tr>
<tr>
<td>the slug and lettuce</td>
<td>gastropub</td>
<td>centre</td>
<td>expensive</td>
<td>--</td>
<td>34 - 35 Green Street</td>
<td>CB 2, 3 JU</td>
</tr>
</tbody>
</table>

Showing 1 to 4 of 4 entries (filtered from 110 total entries)
CamRest676 dataset

- **Ontology:**
  - Cambridge restaurant domain, 99 venues.
  - 3 informable slots: area, price range, food type
  - 3 requestable slots: address, phone, postcode

- **Dataset**
  - 676 dialogues, ~2750 turns
  - 3000 HITS, takes 3 days, costs ~400 USD
  - Data cleaning takes 2-3 days for one person

Link: [https://www.repository.cam.ac.uk/handle/1810/260970](https://www.repository.cam.ac.uk/handle/1810/260970)
Outline

- Intro
- Neural Dialogue System
- Wizard-of-Oz Data Collection
- Experiments
- Conclusion
Experiments

- Experimental details
  - Train/valid/test: 3/1/1
  - SGD, l2 regularisation, early stopping, gradient clip=1
  - Hidden size = 50, Vocab size: ~500
- Two stage training:
  - Training trackers with label cross entropy
  - Training other parts with response cross entropy
- Decoding
  - Beam search w/ beam width 10
  - Decode with average word likelihood
# Response Generation Task

<table>
<thead>
<tr>
<th>Model</th>
<th>Match (%)</th>
<th>Success (%)</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq2Seq [Sutskever et al, 2014]</td>
<td>-</td>
<td>-</td>
<td>0.1718</td>
</tr>
<tr>
<td>HRED [Serban et al, 2015]</td>
<td>-</td>
<td>-</td>
<td>0.1861</td>
</tr>
<tr>
<td>Our model w/o req. trackers</td>
<td>89.70</td>
<td>30.60</td>
<td>0.1799</td>
</tr>
<tr>
<td>Our full model</td>
<td>86.34</td>
<td>75.16</td>
<td>0.2313</td>
</tr>
<tr>
<td>Our full model + attention</td>
<td>90.88</td>
<td>80.02</td>
<td>0.2388</td>
</tr>
</tbody>
</table>
# Human evaluation

## Quality assessment

<table>
<thead>
<tr>
<th>Metric</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success</td>
<td>98%</td>
</tr>
<tr>
<td>Comprehension</td>
<td>4.11</td>
</tr>
<tr>
<td>Naturalness</td>
<td>4.05</td>
</tr>
<tr>
<td># of dialogues</td>
<td>245</td>
</tr>
</tbody>
</table>

## System Comparison

<table>
<thead>
<tr>
<th>Metric</th>
<th>NN</th>
<th>HDC</th>
<th>Tie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subj. Success</td>
<td>96.95%</td>
<td>95.12%</td>
<td>-</td>
</tr>
<tr>
<td>Avg. # of Turn</td>
<td>3.95</td>
<td>4.54</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Comparisons (%)</th>
<th>NN</th>
<th>HDC</th>
<th>Tie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naturalness</td>
<td>46.95*</td>
<td>25.61</td>
<td>27.44</td>
</tr>
<tr>
<td>Comprehension</td>
<td>45.12*</td>
<td>21.95</td>
<td>32.93</td>
</tr>
<tr>
<td>Preference</td>
<td>50.00*</td>
<td>24.39</td>
<td>25.61</td>
</tr>
<tr>
<td>Performance</td>
<td>43.90*</td>
<td>25.61</td>
<td>30.49</td>
</tr>
</tbody>
</table>

* p <0.005,  # of comparisons: 164
# Example dialogues

<table>
<thead>
<tr>
<th>Human</th>
<th>Restaurant in any area that serves Vietnamese food</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine</td>
<td>thanh binh serves vietnamese food type . is there anything else i can help you with ?</td>
</tr>
<tr>
<td>Human</td>
<td>What is the phone number and post code</td>
</tr>
<tr>
<td>Machine</td>
<td>the number is 01223 362456 and the area code is c.b 3, 0 a.f . is there anything else i can do for you ?</td>
</tr>
<tr>
<td>Human</td>
<td>Good bye</td>
</tr>
<tr>
<td>Machine</td>
<td>thank you for using the cambridge restaurant system .</td>
</tr>
</tbody>
</table>

| Human | Hi, I’m hungry for some Indonesian food |
| Machine | i am sorry , there are no indonesian restaurants . would you like a different food ? |
| Human | Oh no. How about Indian? |
| Machine | there are 9 restaurants serving indian types of food . do you have a preference on side of town ? |
| Human | East side please |
| Machine | sitar tandoori is a indian restaurant in the east side of town . would you like their information ? |
| Human | Yes, the address and phone number please |
| Machine | sitar tandoori is located at 43 high street cherry hinton cherry hinton and their phone number is 01223 249955 . |
| Human | Okay, thank you |
| Machine | thank you , goodbye . |
## Example dialogues

<table>
<thead>
<tr>
<th>Human</th>
<th>Machine</th>
</tr>
</thead>
<tbody>
<tr>
<td>hello</td>
<td>what cuisine would you like?</td>
</tr>
<tr>
<td>afghan</td>
<td>i am sorry, there are no afghan restaurants. would you like something else?</td>
</tr>
<tr>
<td>korean</td>
<td>little seoul serves korean food, it is located at 108 regent street city centre.</td>
</tr>
<tr>
<td></td>
<td>is it expensive?</td>
</tr>
<tr>
<td></td>
<td>little seoul serves korean food and is in the expensive price range.</td>
</tr>
<tr>
<td></td>
<td>what area is that</td>
</tr>
<tr>
<td></td>
<td><strong>little seoul is located in the centre area of town. would you like their number?</strong></td>
</tr>
<tr>
<td>yes</td>
<td>the telephone number is 01223 308681. is there anything else i can help you with today?</td>
</tr>
<tr>
<td>no goodbye</td>
<td></td>
</tr>
<tr>
<td></td>
<td>thank you for using the cambridge restaurant system, goodbye.</td>
</tr>
</tbody>
</table>
Visualising action embedding
Outline

- Intro
- Neural Dialogue System
- Wizard-of-Oz Data Collection
- Experiments
- Conclusion
Conclusion

- An end-to-end trainable task-oriented dialogue system architecture is proposed.
- A complementary WOZ data collection is also proposed (no latency, parallel, cheap).
- Results show that it can learn from human-human conversations and help users to complete tasks.
- Explicit language grounding is crucial, but what is the best way to represent semantics?
The paper


References

- M. Henderson, B. Thomson and S. Young. Word-Based Dialog State Tracking with Recurrent Neural Networks, SigDial 2014.
Thank you! Questions?

Tsung-Hsien Wen is supported by a studentship funded by Toshiba Research Europe Ltd, Cambridge Research Laboratory