Dialogue State Tracking as Question Answering

Hung-yi Lee 李宏毅
Dialogue State Tracking (DST)

• It is the core of task-oriented dialogue.

It was called *Dialog State Tracking Challenge (DSTC)* before 2017.

• State-of-the-art approaches consider DST as QA.
Task-oriented Dialogue

Input: several sequences, Output: a sequence
Dialogue State Tracking

User: I will be traveling to Cambridge and cannot wait to try the restaurants, but could you help me with finding a hotel?
Agent: Would you prefer a guesthouse? the Acorn guest house is in the north and moderately priced. 4 starts with free internet and parking.
User: I would prefer a hotel type hotel, it just needs to be in the moderate price range and does not need to have free parking, but needs wifi.
Agent: Sure! I found Ashley hotel. It is located at 74 Chesterton road. Would that work for you?
User: Yes, that sounds like a good place. I appreciate the information.
Agent: You are welcome. Would you like a room, or perhaps I could help you with something else?
User: Yes, can you tell me.
Agent: It has 2 stars.
User: Thanks, I also need a room.
Agent: How about the Gal?
User: Sounds good. I would.
Agent: Okay I was able to
User: Could you also book
Agent: Where will you be
User: From the hotel. I wo
Agent: I was able to book t
User: No, that will be all. T
Dialogue State Tracking

User: I will be traveling to Cambridge and cannot wait to try the restaurants, but could you help me with finding a hotel?
Agent: Would you prefer a guesthouse? the Acorn guest house is in the north and moderately priced. 4 starts with free internet and parking.
User: I would prefer a hotel type hotel, it just needs to be in the moderate price range and does not need to have free parking, but needs wifi.
Agent: Sure! I found Ashley hotel. It is located at 74 Chesterton road. Would that work for you?
User: Yes, that sounds like a good place. I appreciate the information.
Agent: You are welcome. Would you like a room, or perhaps I could help you with something else?
User: Yes, can you tell me the star rating of that hotel?
Agent: It has 2 stars.
User: Thanks, I also need a restaurant in the centre that serves European food. It needs to be in the same price range as hotel.
Agent: How about the Galleria?
User: Sounds good. I would like to make a reservation for Friday at 17:15 for 4 people.
Agent: Okay I was able to make you a reservation and your reference number is DWMSURSO.
User: Could you also book me a taxi that arrives at the restaurant by the time of my reservation?
Agent: Where will you be departing from?
User: From the hotel. I would like to get there before something changes.
Agent: I was able to book that taxi for you. It will be a yellow Honda. Can I assist you with anything else today?
User: No, that will be all. Thank you, good bye.
Dialogue State Tracking (DST)

**Dialogue State**
A set of (key, value)

- **key**
  - hotel price range:
    - **value**
      - moderate
      - cheap
      - expensive
      - don’t care
      - null

Key: Defined before developing your system
Value: choose from a set of predefined options
Dialogue State Tracking (DST)

**Dialogue State**
A set of (key, value)

```
hotel type: hotel
hotel price range: moderate
hotel Internet: yes
hotel name: Ashley hotel
restaurant area: centre
restaurant food: European
restaurant price range: moderate
restaurant name: Galleria
restaurant book day: Friday
restaurant book time: 17:15
restaurant book people: 4
taxi departure: Ashley hotel
taxi destination: Galleria
taxi arrive by: 17:15
```

**Key**
- hotel price range:
- restaurant price range:
- hotel internet:

**Value**
- moderate
- yes / no

Key: Defined before developing your system
Value: choose from a set of options
Dialogue State Tracking (DST)

What the user and agent have said

$S$: agent, $U$: user

Key $k$ include domain and slot

$k_1$: \[ \{v_1^1, v_2^1, \ldots, v_{n_1}^1\} \]

$k_2$: \[ \{v_1^2, v_2^2, \ldots, v_{n_2}^2\} \]

$k_3$: \[ \{v_1^3, v_2^3, \ldots, v_{n_3}^3\} \]

Select the correct value for each key
MultiWOZ 2.0

[Budzianowski, et al., EMNLP’18]

- Multi-domain

There are 7 domains actually.

<table>
<thead>
<tr>
<th>Slots</th>
<th>Hotel</th>
<th>Train</th>
<th>Attraction</th>
<th>Restaurant</th>
<th>Taxi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>price, type, parking, stay, day, people, area, stars, internet, name</td>
<td>destination, departure, day, arrive by, leave at, people</td>
<td>area, name, type</td>
<td>food, price, area, name, time, day, people</td>
<td>destination, departure, arrive by, leave by</td>
</tr>
<tr>
<td>Train Valid Test</td>
<td>3381</td>
<td>3103</td>
<td>2717</td>
<td>3813</td>
<td>1654</td>
</tr>
<tr>
<td></td>
<td>416</td>
<td>484</td>
<td>401</td>
<td>438</td>
<td>207</td>
</tr>
<tr>
<td></td>
<td>394</td>
<td>494</td>
<td>395</td>
<td>437</td>
<td>195</td>
</tr>
</tbody>
</table>
What is WOZ?
What is WOZ?

• One way of system test / data collection

https://dl.acm.org/doi/pdf/10.1145/2163.358100

Artificial AI (AAI)
### Schema-Guided Dialogue (SGD)

[Rastogi, et al., AAAI’20]

<table>
<thead>
<tr>
<th>Metric</th>
<th>DSTC2</th>
<th>WOZ2.0</th>
<th>FRAMES</th>
<th>M2M</th>
<th>MultiWOZ</th>
<th>SGD</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of domains</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td>No. of dialogues</td>
<td>1,612</td>
<td>600</td>
<td>1,369</td>
<td>1,500</td>
<td>8,438</td>
<td>16,142</td>
</tr>
<tr>
<td>Total no. of turns</td>
<td>23,354</td>
<td>4,472</td>
<td>19,986</td>
<td>14,796</td>
<td>113,556</td>
<td>329,964</td>
</tr>
<tr>
<td>Avg. turns per dialogue</td>
<td>14.49</td>
<td>7.45</td>
<td>14.60</td>
<td>9.86</td>
<td>13.46</td>
<td>20.44</td>
</tr>
<tr>
<td>Avg. tokens per turn</td>
<td>8.54</td>
<td>11.24</td>
<td>12.60</td>
<td>8.24</td>
<td>13.13</td>
<td>9.75</td>
</tr>
<tr>
<td>Total unique tokens</td>
<td>986</td>
<td>2,142</td>
<td>12,043</td>
<td>1,008</td>
<td>23,689</td>
<td>30,352</td>
</tr>
<tr>
<td>No. of slots</td>
<td>8</td>
<td>4</td>
<td>61</td>
<td>13</td>
<td>24</td>
<td>214</td>
</tr>
<tr>
<td>No. of slot values</td>
<td>212</td>
<td>99</td>
<td>3,871</td>
<td>138</td>
<td>4,510</td>
<td>14,139</td>
</tr>
</tbody>
</table>

**simulator**

REQUEST(location)
REQUEST(cuisine)
INFORM(location="los angeles")
INFORM(date="2019-03-02")
INFORM(cuisine="greek")

**template**

Which city are you in? What kind of food?
I want to eat in LA. I want to go there tomorrow. I want to eat greek food.
Opa is a nice restaurant.

**human**

Where are you located and what is your preferred food?
I am looking for a greek restaurant in LA for tomorrow. How does Opa sound?

[Image of dialogue simulation]
# CrossWOZ

Zhu, et al., TACL’20

<table>
<thead>
<tr>
<th>Type</th>
<th>Single-domain goal</th>
<th>Multi-domain goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset</td>
<td>DSTC2</td>
<td>WOZ 2.0</td>
</tr>
<tr>
<td>Language</td>
<td>EN</td>
<td>EN</td>
</tr>
<tr>
<td>Speakers</td>
<td>H2M</td>
<td>H2H</td>
</tr>
<tr>
<td># Domains</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td># Dialogues</td>
<td>1,612</td>
<td>600</td>
</tr>
<tr>
<td># Turns</td>
<td>23,354</td>
<td>4,472</td>
</tr>
<tr>
<td>Avg. domains</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Avg. turns</td>
<td>14.5</td>
<td>7.5</td>
</tr>
<tr>
<td># Slots</td>
<td></td>
<td></td>
</tr>
<tr>
<td># Values</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**CrossWOZ**

**usr:** 你好，可以帮我推荐一个评分是4.5分以上的景点吗？

**Hello, could you recommend an attraction with a rating of 4.5 or higher?**

**sys:** 天安门城楼，簋街小吃和北京欢乐谷都是很不错的地方呢。

**Tiananmen, Gui Street, and Beijing Happy Valley are very nice places.**

**usr:** 我喜欢北京欢乐谷，你知道这个景点周边的酒店都是什么吗？

**I like Beijing Happy Valley. What hotels are around this attraction?**

**sys:** 那可多了，有A酒店, B酒店, C酒店。

**There are many, such as hotel A, hotel B, and hotel C.**

**usr:** 太好了，我正打算在景点附近找个酒店住宿呢，知道哪家评分是4分以上，提供叫醒服务的不？

**Great! I am planning to find a hotel to stay near the attraction. Which one has a rating of 4 or higher and offers wake-up call service?**
Challenge - Infinite values

• Some slots have almost infinite possible values (e.g. leave time, phone number)

What is the [SLOT NAME]?

What is the phone number?

Dialogue History

$S_1, U_1, S_2, U_2, ..., S_T, U_T$

Module for Source

Module for Answer

Span in dialogue history

Module for Question

attention
## Challenge – New Service

### Having training data

<table>
<thead>
<tr>
<th>Slots</th>
<th>Hotel</th>
<th>Train</th>
<th>Attraction</th>
<th>Restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>price, type, parking, stay, day, people, area, stars, internet, name</td>
<td>destination, departure, day, arrive by, leave at, people</td>
<td>area, name, type</td>
<td>food, price, area, name, time, day, people</td>
</tr>
</tbody>
</table>

- **Question:** What is the **destination** of **train**?
- **Question:** What is the **destination** of **taxi**?
TRADE

TRAnsferrable Dialogue state generator

[Wu, et al., ACL’19]

(a) Utterance Encoder

Context Vector $c_{j0}$

Slot Gate $G_j$

- PTR
- DONTCARE
- NONE

(b) State Generator

$p_{j0}^{\text{final}}$

$\times (1 - p_{j0}^{\text{gen}})$

$\times (p_{j0}^{\text{gen}})$

$p_{j0}^{\text{history}}$

$h_{j0}^{\text{dec}}$

Hotel?

Ashley

(c) Utterances

Bot: Which area are you looking for the hotel?
User: There is one at east town called Ashley Hotel.

(Do not need a complete question)
<table>
<thead>
<tr>
<th>parking</th>
<th>internet</th>
<th>food</th>
<th>name</th>
<th>type</th>
<th>area</th>
<th>pricerange</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TRADE

- Zero Shot

<table>
<thead>
<tr>
<th>Slots</th>
<th>Hotel</th>
<th>Train</th>
<th>Attraction</th>
<th>Restaurant</th>
<th>Taxi</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>price, type, parking, stay, day, people, area, stars, internet, name</td>
<td>destination, departure, day, arrive by, leave at, people</td>
<td>area, name, type</td>
<td>food, price, area, name, time, day, people</td>
<td>destination, departure, arrive by, leave by</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Trained Single</th>
<th>Zero-Shot</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Joint</td>
<td>Slot</td>
</tr>
<tr>
<td>Hotel</td>
<td>55.52</td>
<td>92.66</td>
</tr>
<tr>
<td>Train</td>
<td>77.71</td>
<td>95.30</td>
</tr>
<tr>
<td>Attraction</td>
<td>71.64</td>
<td>88.97</td>
</tr>
<tr>
<td>Restaurant</td>
<td>65.35</td>
<td>93.28</td>
</tr>
<tr>
<td>Taxi</td>
<td>76.13</td>
<td>89.53</td>
</tr>
</tbody>
</table>
Embedding from Description

[Rastogi, et al., AAAI’20]
Slot Carryover Prediction

[Graphical representation of the prediction process]

User: I need to book a hotel in the east that has 4 stars. **area=east, stars=4**
Agent: I can help you with that. What is your price range?
User: That doesn’t matter if it has free wifi and parking. **parking=yes, internet=yes, price=don't care, stars=4, area=east**

(All other slots are NULL)

[Reference: Gao, et al., SIGDIAL’19]
The (domain, slot) pairs are not independent. If a user booked a restaurant, then the destination of the taxi is likely to be that restaurant.
**SimpleTOD**

[Hosseini-Asl, et al., arXiv’20]

Fine-tuned from GPT-2

**Dialogue Context**

<table>
<thead>
<tr>
<th>Context</th>
<th>[context] [user] user input [system] system response . . . [user] user input [endofcontext]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belief State</td>
<td>[belief] domain slot_name value, domain slot_name value, . . . [endofbelief]</td>
</tr>
<tr>
<td>DB Search</td>
<td>[db] #_matches, booking_status [endofdb]</td>
</tr>
<tr>
<td>Action</td>
<td>[action] domain action_type slot_name, domain action_type slot_name, . . . [endofaction]</td>
</tr>
<tr>
<td>Response</td>
<td>[response] system delexicalized response [endofresponse]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Decoder</th>
<th>Context Encoder</th>
<th>Extra Supervision</th>
<th>Joint Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRADE</td>
<td>Generative + Classifier</td>
<td>Bidirectional</td>
<td>-</td>
<td>45.6</td>
</tr>
<tr>
<td>DSTQA</td>
<td>Classifier</td>
<td>Bidirectional</td>
<td>knowledge graph</td>
<td>51.17</td>
</tr>
<tr>
<td>DST-Picklist</td>
<td>Classifier</td>
<td>Bidirectional</td>
<td>-</td>
<td>53.3</td>
</tr>
<tr>
<td>SST</td>
<td>Generative</td>
<td>Bidirectional</td>
<td>schema graph</td>
<td>55.23</td>
</tr>
<tr>
<td>SimpleTOD (ours)</td>
<td>Generative</td>
<td>Unidirectional</td>
<td>-</td>
<td><strong>55.72</strong></td>
</tr>
</tbody>
</table>

Table 2: Evaluation of Dialogue State Tracking (DST) on MultiWOZ 2.1 using joint accuracy metric.
<table>
<thead>
<tr>
<th>Context ('MUL1015', turn 10)</th>
<th>Target</th>
<th>SimpleTOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;context&gt; i am looking for a pool somewhere in the south of cambridge. &lt;system&gt; i am sorry, but it does not look like we have a pool that matches your criteria. &lt;user&gt; how about some type of theater? &lt;system&gt; the junction theatre is in the south area. would you like the address? &lt;user&gt; yes, please provide the address and postcode as well. &lt;system&gt; the junction theatre is located on clifton way at postcode cb17gx. is there anything else i can assist you with? &lt;user&gt; thanks. yes, i also need to find a guesthouse for my stay there. &lt;system&gt; are you needing the guesthouse in the south, also? do you have a preference in price range? &lt;user&gt; the area should be the centre. i need it to be cheap and have free parking. &lt;system&gt; i have 2 guesthouses that fit your criteria. 1 is the alexander bed and breakfast and the other is el shaddai. &lt;user&gt; what is the star rating -s for them? &lt;system&gt; the alexander bed and breakfast has a 4 star rating while the el shaddai has 0 stars. would you like to book a room with either of these? &lt;user&gt; ok great, thank you. i also need to book a taxi to get back and forth. &lt;system&gt; where did you want to leave from and go to? &lt;user&gt; i’ll be leaving the theatre at 14:45 and going to the guesthouse. &lt;system&gt; let’s back up a moment. which lodgings did you prefer, so we know where the taxi will be taking you? &lt;user&gt; let’s go with the alexander bed and breakfast. &lt;system&gt; your booking is complete, a grey volkswagen will be picking you up, contact number 07753429052. can i help with anything else? &lt;user&gt; cool beans! sorry i was so much trouble. thanks so much for all of your help, you need a raise dealing with customers like me! &lt;endofcontext&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;belief&gt; attraction area south, attraction name the junction, attraction type theatre, hotel area centre, hotel name alexander bed and breakfast, hotel parking yes, hotel pricerange cheap, hotel type guesthouse, taxi departure the junction, taxi destination alexander bed and breakfast, taxi leaveat 14:45 &lt;belief&gt; &lt;action&gt; general bye none &lt;endofaction&gt; &lt;response&gt; it was no trouble at all! i hope you have a fantastic trip and enjoy your stay! &lt;endofresponse&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;belief&gt; attraction area south, attraction name the junction, attraction type theatre, hotel area centre, hotel name alexander bed and breakfast, hotel parking yes, hotel pricerange cheap, hotel type guesthouse, taxi departure the junction, taxi destination alexander bed and breakfast, taxi leaveat 14:45 &lt;endofbelief&gt; &lt;action&gt; general bye none &lt;endofaction&gt; &lt;response&gt; thank you for using our service &lt;endofresponse&gt;</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Reference

- Shuyang Gao, Abhishek Sethi, Sanchit Agarwal, Tagyoung Chung, Dilek Hakkani-Tur, Dialog State Tracking: A Neural Reading Comprehension Approach, SIGDIAL, 2019
- Li Zhou, Kevin Small, Multi-domain Dialogue State Tracking as Dynamic Knowledge Graph Enhanced Question Answering, NeurIPS Workshop, 2019
Reference

• Qi Zhu, Kaili Huang, Zheng Zhang, Xiaoyan Zhu, Minlie Huang, CrossWOZ: A Large-Scale Chinese Cross-Domain Task-Oriented Dialogue Dataset, TACL, 2020

• Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, Milica Gašić, MultiWOZ -- A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling, EMNLP, 2018
One slide for this course

Audio BERT → Model → ASR

Model → Model → VC, Source Separation

Model → Model → Speaker Verification

Model → Model → class

Model → Model → Text “Style” Transfer

Model → Model → BERT, Multi-BERT

Model → Model → class

Vocoder
<table>
<thead>
<tr>
<th></th>
<th>One Sequence</th>
<th>Multiple Sequences</th>
</tr>
</thead>
</table>
| **One Class**           | Sentiment Classification  
Stance Detection  
Veracity Prediction  
Intent Classification  
Dialogue Policy         |                                          |
| **Class for each Token**| POS tagging  
Word segmentation  
Extractive Summarization  
Slotting Filling  
NER                      |                                          |
| **Copy from Input**     |                                                   | Extractive QA                            |
| **General Sequence**    | Abstractive Summarization  
Translation  
Grammar Correction  
NLG                     | General QA  
Chatbot  
State Tracker  
Task Oriented Dialogue  |
| **Other?**              | Parsing  
Coreference Resolution                          |                                          |