Learning Multi-Domain Dialogues

Paweł Budzianowski
Spoken Dialogue Systems

“Book me a cab to Russell Square“

speech  →  speech

Language Understanding

inform(service=taxi, dest=Russell Square)

Dialogue Management

_request(depart_time)

Response Generation

speech  →  text

semantics

text  →  text

Speech Synthesis

"When do you want to leave?"
Reinforcement learning
Problems of Deep RL in dialogue policy optimisation

1. No uncertainty estimates,
Uncertainty estimates

ICASSP18, with Christopher Tegho (Calipsa)
Problems of Deep RL in dialogue policy optimisation

1. No uncertainty estimates,
2. Sample efficiency in large action space,
Large-action space

TASLP18, with Gellert Weisz (DeepMind)
Problems of Deep RL in dialogue policy optimisation

1. No uncertainty estimates,
2. Sample efficiency in large action space,
3. Multi-domain policy learning,
Hierarchical Reinforcement Learning

SIGDIAL18
Problems of Deep RL in dialogue policy optimisation

1. No uncertainty estimates,
2. Sample efficiency in large action space,
3. Multi-domain policy learning,
Problems?
Problems?
Does RL work?

WHENEVER SOMEONE ASKS ME IF RL WORKS, I TELL THEM IT DOESN'T

AND 70% OF THE TIME, I'M RIGHT

(Credit: https://www.alexirpan.com/2018/02/14/rl-hard.html)
Cold start problem

The most natural way (and realistic from industrial point of view) is to bootstrap the model on demonstration data and fine tune it in direct interactions with real users.
MultiWOZ - Large-Scale Multi-Domain Dataset for Task-Oriented Dialogue Modelling

EMNLP 2018, with Tsung-Hsien Wen (PolyAI)
Available corpora

Data size

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSTC2</td>
<td>1612</td>
</tr>
<tr>
<td>WOZ2.0</td>
<td>1200</td>
</tr>
<tr>
<td>KVRET</td>
<td>2,425</td>
</tr>
<tr>
<td>FRAMES</td>
<td>1369</td>
</tr>
<tr>
<td>M2M</td>
<td>1500</td>
</tr>
</tbody>
</table>
Available corpora vs industry

Data size

<table>
<thead>
<tr>
<th></th>
<th>DSTC2</th>
<th>WOZ2.0</th>
<th>KVRET</th>
<th>FRAMES</th>
<th>M2M</th>
<th>SIRI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>1612</td>
<td>1200</td>
<td>2,425</td>
<td>1369</td>
<td>1500</td>
<td>1,000,000,000</td>
</tr>
</tbody>
</table>
Human-to-Machine
Let's Go Bus Information System - Raux et al. 2005
Dialogue State Tracking Challenge - Williams et al. 2013

Machine-to-Machine
Bordes et al. 2017, Shah et al. 2018

Human-to-Human
Motivation

1. Fully labeled, MTurk-based data collection set-up,
2. Large-scale, complex and multi-domain dataset,
Data Collection Set-up

We followed the Wizard-of-Oz set-up (Kelley, 1984) - corpora of annotated dialogues can be gathered at relatively low costs and with a small time effort.
Wizard of Oz Setup

USER

TASK

SYSTEM

GUI
Wizard of Oz Setup

USER

TASK

SYSTEM

GUI
Data Collection Set-up

We followed the Wizard-of-Oz set-up (Kelley, 1984) - corpora of annotated dialogues can be gathered at relatively low costs and with a small time effort.

Such WOZ set-up has been successfully validated by Wen et al. (2017).

However, we aim at more complex, longer, fully annotated and multi-domain dialogues that can be gathered at large-scale.
Use-side

Help Desk: Hello, welcome to the TownInfo centre. I can help you find a restaurant or hotel, look for tourist information, book a train or taxi. How may I help you?
Customer: I want a place to stay in the east.
Help Desk: I have 6 guesthouses and 1 hotel on the east side. What's your price range?
Customer: Doesn't matter too much. I'd like a 4 star property, though, and would prefer one of the guesthouses.
Help Desk: I'd recommend 517a coldham lane. Would you like me to book a room?
Customer: Could you give me their phone number? I would like to verify that they have free parking.
Help Desk: Allentell does have parking and the phone is 01223210353. Can I help with anything else?

Next turn

Customer: (Your response)

Please try to chat about the following topic:

Task MUL002:

- You are traveling to and looking forward to try local restaurants.
- You are looking for a place to stay. The hotel should be in the east and should include free parking.
- You are also looking at the hotel, which should be in the moderate price range.
- If there is no such hotel, you seek a restaurant that serves turkish food.
- Once you find the hotel, you decide to book a room for 4 people at 17:45 on friday.
- Make sure you get the reference number

--- The End ---
You are traveling to Cambridge and looking forward to try local restaurants.

Your role: the hotel should be in the type of hotel and should be in the centre.
The hotel should include free wifi and should have a star of 4.

Once you find the hotel you want to book it for 3 people and 5 nights starting from Monday.

Make sure:

You are also looking for a restaurant. The restaurant should serve australasian food and should be in the moderate price range.
The restaurant should be in the same area as the hotel.

If there is no such restaurant, how about one that serves british food?
Once a table for the same group of people at 18:30 on the same day.
Make sure you get the reference number
User-side

Hello, welcome to the Towninfo centre. I can help you find a restaurant or hotel, look for tourist information, book a train or taxi. How may I help you?

Customer: I want a place to stay in the east.

Help Desk: I have 6 guesthouses and 5 hotels on the east side. What's your price range?

Customer: Doesn't matter too much, and would prefer one of the guesthouses.

Help Desk: I'd recommend Allentown. Could you give me the parking.

Customer: Allenbell does have parking and the phone is 0123201353. Can I help with anything else?

Next turn

You need to go through the dialogue first by clicking the 'Next turn' button.

What topics were mentioned in this turn:

General: Booking: Restaurant: Hotels: Police: Train: Taxi: Bus:

User answer

Please try to chat about the following topic:

Task MUL0002:

- You are traveling to and looking forward to try local restaurants.
- You are looking for a place to stay. The hotel should be in the east and should include free parking.
- The hotel should be in a type of guesthouse.
- Make sure you get the reference number.
- You are also looking for some place to eat. Restaurant should be in the moderate price range and serves turkish food.
- If there is no such place, please ask for a reservation for 4 people at 17:45 on Friday.
- Make sure you get the reference number.

--- The End ---
Belief state

**User:** I need to find a luxury hotel please.
**Belief state:** inform(domain=hotel, price=expensive)

**System:** Pinehouse will be a good place.

**User:** Book me a room there, please.
**Belief state:** book(hotel)

**System:** Sure, what day and for how long?
System side

Dialogue context

Database GUI

System response
Annotation of system turns

How to acquire high-quality labels for a very specific and challenging task even for NLP practitioners?

We perform two step-approach - turkers were asked to annotate an illustrative, long dialogue which covered many problematic examples.

The chosen subset of well-performing turkers were given more detailed instructions and required to go through the test again.
### Annotation of system turns

<table>
<thead>
<tr>
<th>Turn 1</th>
<th>John</th>
<th>Hi Anna, how can I help you?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anna</td>
<td>I want to arrive by 11:30 at Cambridge.</td>
<td></td>
</tr>
</tbody>
</table>

### Dialogue context

There are three trains, one arrives at 6:43, one at 8:43, and one at 10:43, which would you like to book?

### Dialogue acts

- informing that no trains are available
- informing that train was booked

### Slot-value pairs

- **Train domain**
  - John is requesting: day | departure place | destination place | leave after | arrive by | people
  - informing about
  - people | possible choices 3 | destination place | leave after | arrival by 6:43, 8:43, 10:43
  - offering to book a train
  - people | possible choices | departure place | destination place | leave after | arrival by | day | reference | trainID | ticket price | travel time

### Domains

- Booking
- Restaurant
- Hotel
- Attraction
- Taxi
- Train
- Hospital
- Police

- More help
- Greetings
- Goodbye
- You're welcome
- Not sure
MultiWOZ corpus

The dataset consists of natural conversations between a tourist and a clerk from an information center in Cambridge.

The corpus consists of 7 domains including: Attraction, Hospital, Police, Hotel, Restaurant, Taxi and Train.

There are various possible dialogue scenarios ranging from finding a train, a suitable hotel and booking a taxi between both places.

To make the dialogues more complex and realistic the initial goal for the user was sometimes impossible to accomplish.
Data structure

<table>
<thead>
<tr>
<th></th>
<th>Single</th>
<th>Multi</th>
</tr>
</thead>
<tbody>
<tr>
<td># of dialogues</td>
<td>3,406</td>
<td>7,032</td>
</tr>
<tr>
<td># of domains</td>
<td>1-2</td>
<td>2-6</td>
</tr>
</tbody>
</table>

Each dialogue consists of the user goal, the task description presented to MTurkers, multiple user and system utterances along with annotations for both sides of the conversation.
Data analysis
Benchmarks

Dialogue State Tracking

Dialogue-Context-to-Text Generation

Natural Language Generation
Dialogue State Tracking

**User:** I need to find a luxury hotel please.

**Belief state:** inform(domain=hotel, price=expensive)

**System:** Pinehouse will be a good place.

**User:** Book me a room there, please.

**Belief state:** book(hotel)

**System:** Sure, what day and for how long?
## Dialogue State Tracking

<table>
<thead>
<tr>
<th>Slot</th>
<th>WOZ 2.0</th>
<th>MultiWOZ (restaurant)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint goals</td>
<td>85.5</td>
<td>80.9</td>
</tr>
</tbody>
</table>
## Dialogue State Tracking

<table>
<thead>
<tr>
<th>Slot</th>
<th>WOZ 2.0</th>
<th>MultiWOZ (restaurant)</th>
<th>MultiWOZ (all domains)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Joint goals</td>
<td>85.5</td>
<td>80.9</td>
<td>25.8</td>
</tr>
</tbody>
</table>
Dialogue-Context-to-Text Generation

**User**: I need to find a luxury hotel please.
Belief state: inform(domain=hotel, price=expensive)

**System**: Pinehouse will be a good place.

**User**: Book me a room there, please.
Belief state: book(hotel)

**System**: Sure, what day and for how long?
Dialogue-Context-to-Text Generation

**DATABASE POINTER**

- Restaurant: Seven stars, Italian melody, Green world
- Hotel: Your world, Imperial Hotel, Royal Resort & Spa

**ORACLE BELIEF STATE**

- Domain: Price:
  - Restaurant: Expensive
  - Hotel: Not mentioned
  - Attraction: None
  - Train: None

- Sure what about sth in the center?

- Find me some really luxurious hotel alright?
## Dialogue-Context-to-Text Generation

<table>
<thead>
<tr>
<th></th>
<th>Cam676</th>
<th>MultiWOZ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>w/o attention</td>
<td>w/ attention</td>
</tr>
<tr>
<td>Inform(%)</td>
<td>99.17</td>
<td>99.58</td>
</tr>
<tr>
<td>BLEU</td>
<td>0.219</td>
<td>0.204</td>
</tr>
</tbody>
</table>
Natural Language Generation

User: I need to find a luxury hotel please.
Belief state: inform(domain=hotel, price=expensive)

System: Pinehouse will be a good place.
Dialogue act: inform(name=Pinehouse)

User: Book me a room there, please.
Belief state: book(hotel)

System: Sure, what day and for how long?
Dialogue act: request(day, stay)
Natural Language Generation

<table>
<thead>
<tr>
<th>Metric</th>
<th>SFX</th>
<th>MultiWOZ (restaurant)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SER (%)</td>
<td>0.46</td>
<td>4.378</td>
</tr>
<tr>
<td>BLEU</td>
<td>0.731</td>
<td>0.616</td>
</tr>
</tbody>
</table>
Future work

1. Introducing noise in the conversations (?),
2. Making it multi-lingual
You can find the corpus and benchmarks at:

http://dialogue.mi.eng.cam.ac.uk/index.php/corpus/
Acknowledgments

We would like to thank Google for providing a generous support for this data collection.
Managing Concurrent Actions
I have found 4 hotels satisfying your criteria. Do you have any preference for the area?

**Action-slot-value representation** -
Inform(domain=hotel, price=moderate, entities=4)
Request(domain=hotel, area)
Concurrent actions
### Action list in MultiWOZ

<table>
<thead>
<tr>
<th>Dialogue act</th>
<th>System-specific?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Request</td>
<td>✗</td>
</tr>
<tr>
<td>Inform</td>
<td>✗</td>
</tr>
<tr>
<td>Request more</td>
<td>✓</td>
</tr>
<tr>
<td>Book</td>
<td>✗</td>
</tr>
<tr>
<td>Recommend</td>
<td>✗</td>
</tr>
<tr>
<td>Goodbye</td>
<td>✓</td>
</tr>
<tr>
<td>You are welcome</td>
<td>✓</td>
</tr>
<tr>
<td>Train was booked</td>
<td>✗</td>
</tr>
<tr>
<td>Train booking intent</td>
<td>✗</td>
</tr>
<tr>
<td>No entities available</td>
<td>✗</td>
</tr>
<tr>
<td>Booking not possible</td>
<td>✗</td>
</tr>
<tr>
<td>Select</td>
<td>✗</td>
</tr>
<tr>
<td>Greet</td>
<td>✓</td>
</tr>
<tr>
<td>No annotations</td>
<td>✗</td>
</tr>
</tbody>
</table>
Main architecture

**Encoded State**

**Encoder**
Tourist:
- Hotel is reserved right? Can you also book a taxi after 12:00?

**Oracle Dialogue State**
- Hotel (name=Sunny Vale, area=center, price=cheap)
- Taxi (departure=Sunny Vale, arrival_place=, time=12:00)

**Database Pointer**
- Hotel:
  - Sunny Vale 0
  - Holiday Resort 1
  - Hetford’s House 0
- Taxi:
  - White Suzuki 0
  - Black Nissan 1

**Policy**
- Inform
- Recommend
- Goodbye
- Request
- Book
- Deny booking
- Greet
- Offer booking

**Loss 1:** Binary cross-entropy

**Loss 2:** Log-likelihood

**Loss 3:** REINFORCE

**Target:**
- Reward = 1

**Generated:**
- Yes, I booked you a room. What is the departure place?
- Reward = 0

**Reward:**
- Reward = 0 - 1 + BLEU

**SL Training**

**Decoder**
Target:
- Yes, your stay is booked. Sure, where do you want to go?

**Generated:**
- Yes, I booked you a room. What is the departure place?

**RL Training**
Input to the network

Input - a sequence of input tokens $\mathbf{w}_t = (w^0_t, w^1_t, \ldots, w^L_t)$ encoded through $\text{RNN}_\theta(\mathbf{u}_t)$ from which the last one $\mathbf{u}_L$ is used as an encoding of the user intent:

$$e_t = u^L_t.$$

We model also prediction of dialogue state:

$$b_t = \bigoplus_{d} \bigoplus_{s} b_{s\cdot d, t}.$$

And list of entities satisfying current constraint:

$$kb_t = \bigoplus_{d} kb_{d, t}.$$
Policy modelling

This serves as an input:

\[ x_t = e_t \oplus b_t \oplus kb_t \]

to predict probabilities over action set:

\[ \pi(a_t | x_t) = MLP(x_t). \]
Sigmoid vs Softmax

Actions are sampled from the derived probabilities:

\[ a_t \sim \pi(a_t | x_t). \]

The set of all possible actions \( \mathcal{A} = \{a_1, a_2, \ldots, a_N\} \) consists of \( N \) individual actions from which we can choose a subset.

Standard reinforcement learning approaches restrict a choice to one action per time step through stochastic policy \( \pi : \mathcal{X} \rightarrow \mathcal{A} \).

One-hot encoding leads to \( 2^N \) possible outputs. Even for a small action space that is considered here (14), we arrive at 16384 values.

The sigmoid output does not suffer from that as it scales linearly.
Training with supervised loss

Cross-entropy over output words:

\[ L_1(\theta) = \sum_d \sum_t \sum_j y_j^t \log p_j^t. \]

Cross-entropy over actions predictions:

\[ L_2(\theta) = \sum_d \sum_t (1 - a_t)(1 - \log p_a^t) + a_t \log p_a^t. \]
Main architecture

ENCODER
Tourist:
Hotel is reserved right?
Can you also book a taxi after 12:00?

ORACLE DIALOGUE STATE
Hotel(name=Sunny Vale, area=center, price=cheap)
Taxi(departure=Sunny Vale, arrival_place=?, time=12:00)

DATABASE POINTER
Hotel:
Sunny Vale 0
Holiday Resort 1
Hertford’s House 0

Taxi:
White Suzuki 0
Black Nissan 1

POLICY
Loss 1: Binary cross-entropy
Inform
Recommend
Goodbye
Request
Book
Deny booking
Greet
Offer booking

DECODER
Target:
Yes, your stay is booked.
Sure, where do you want to go?

Generated:
Yes, I booked you a room.
What is the departure place?

Loss 2: Log-likelihood

RL TRAINING
Sampling operation
σ
Concatenation
⊗
Multiplcation

Loss 3: REINFORCE
Target:
Reward = 1

Generated:
What is the departure place?
Reward = 0

Reward = 0 - 1 + BLEU
Reinforcement learning fine-tuning phase

Ideally, the system would be getting better through autonomous learning with direct interactions with real users.

\[ L_3 = \frac{1}{T} \sum_{t}^{T} \nabla \log \pi_{\theta}(a_t| x_t) r_t. \]

Employing a standard RL framework here is not possible as it requires softmax probabilities.
Multi-action reinforcement learning

We followed initial work on concurrent actions by Harmer et al (2018) where each action is conditionally independent given the state $\mathbf{x}$, i.e.

$$
\pi_\theta(a_t|x_t) = \prod_{n=1}^{N} \pi(a^n_t|x_t).
$$

This assumption allows to treat each action as a Bernoulli random variable leading to:

$$
\pi(a_t|x_t) = \prod_{n=1}^{N} (a^n_t z^n_t + (1-a^n_t)(1-z^n_t)).
$$
New RL loss

By putting it back to the original RL loss we get:

\[
L_3 = -\frac{1}{T} \sum_t^T \nabla_\theta \log \left( \prod_{n=1}^N a_t^n \log(z_t^n) + (1 - a_t^n) \log(1 - z_t^n) \right) r_t.
\]
Main architecture
Experiments
Metrics

0.5 * Inform + 0.5 * Success + BLEU
Experiments - SL phase

Constrained set:

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Sigmoid</th>
<th>Softmax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inform (%)</td>
<td>78.71</td>
<td>82.49</td>
<td>82.19</td>
</tr>
<tr>
<td>Success (%)</td>
<td>65.21</td>
<td>66.95</td>
<td>68.11</td>
</tr>
<tr>
<td>BLEU (%)</td>
<td>17.7</td>
<td>18.8</td>
<td>18.79</td>
</tr>
</tbody>
</table>

Full set:

<table>
<thead>
<tr>
<th></th>
<th>Sigmoid</th>
<th>Softmax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inform (%)</td>
<td>84.15,</td>
<td>83.0</td>
</tr>
<tr>
<td>Success (%)</td>
<td>67.42</td>
<td>34.339</td>
</tr>
<tr>
<td>BLEU (%)</td>
<td>17.76</td>
<td>9.8</td>
</tr>
</tbody>
</table>
## Experiments - RL phase

<table>
<thead>
<tr>
<th></th>
<th>Sigmoid (subset)</th>
<th>Sigmoid (full set)</th>
<th>Softmax (subset)</th>
<th>Softmax (full set)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inform (%)</td>
<td>84.519</td>
<td>85.789</td>
<td>82.1</td>
<td>83.08</td>
</tr>
<tr>
<td>Success (%)</td>
<td>67.57</td>
<td>67.63</td>
<td>48.18</td>
<td>34.34</td>
</tr>
<tr>
<td>BLEU (%)</td>
<td>17.36</td>
<td>15.6</td>
<td>14.8</td>
<td>9.8</td>
</tr>
</tbody>
</table>
Multi-Action and Slot Dialogue Agent
Additional loss

Cross-entropy over slots predictions:

\[ L_4(\theta) = \sum_d \sum_t (1 - s_t)(1 - \log p_s^t) + s_t \log p_s^t. \]

Final loss consists of:

\[ L = L_1(\text{words}) + L_2(\text{actions}) + L_4(\text{slots}) + L_3(\text{RL}). \]
## Full model evaluation

<table>
<thead>
<tr>
<th></th>
<th>Sigmoid SL</th>
<th>Sigmoid RL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inform (%)</td>
<td>83.97,</td>
<td>88.34</td>
</tr>
<tr>
<td>Success (%)</td>
<td>71.44</td>
<td>75.41</td>
</tr>
<tr>
<td>BLEU (%)</td>
<td>16.78</td>
<td>15.95</td>
</tr>
</tbody>
</table>
Future work

1. Bridging unsupervised latent action space discovery with our approach,

2. Building a fully end-to-end model with dialogue state predictions obtained from the model rather than oracle ones,

3. Combining unsupervised pre-training with well-specified domain.
Any questions?