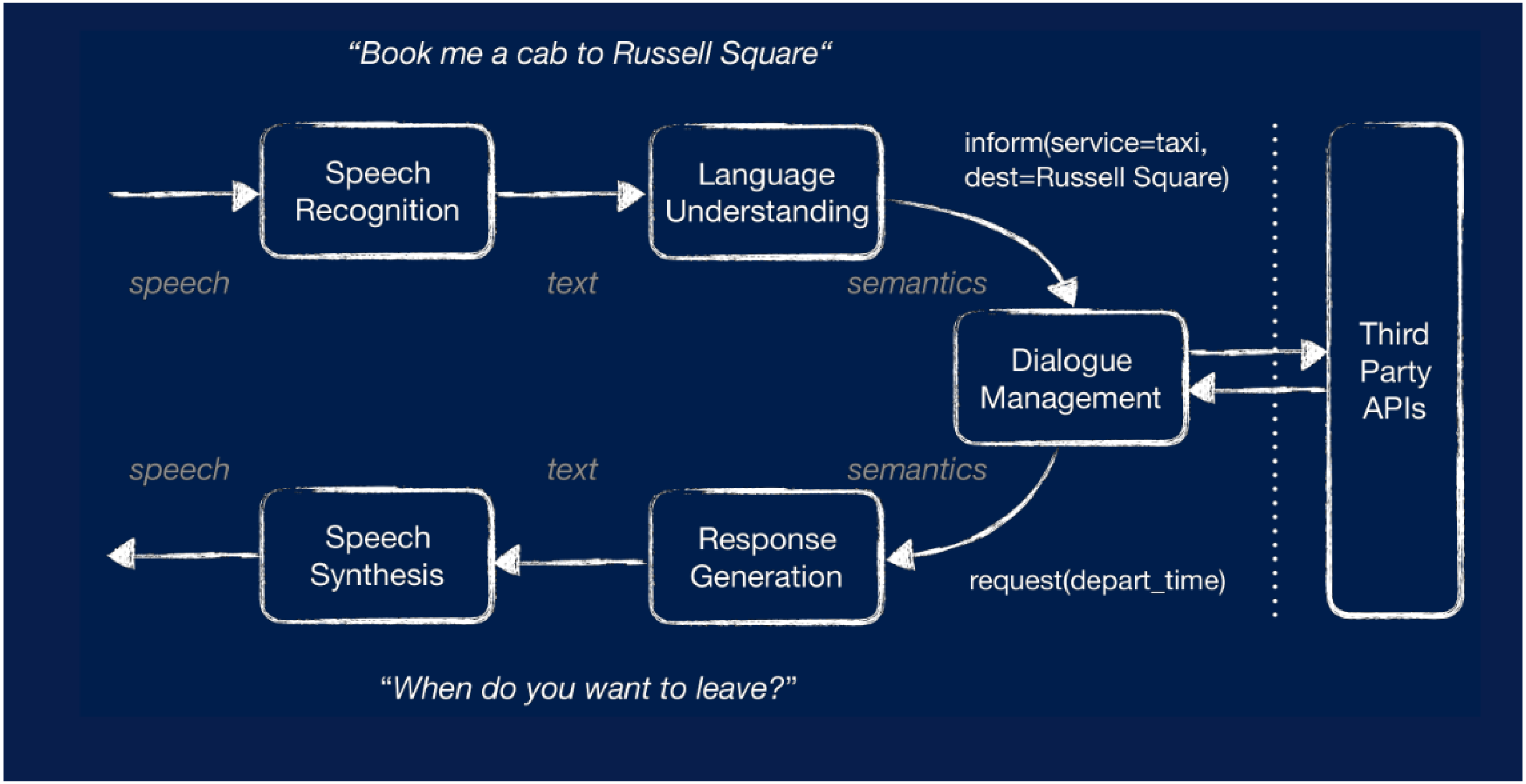
Learning Multi-Domain Dialogues

Paweł Budzianowski

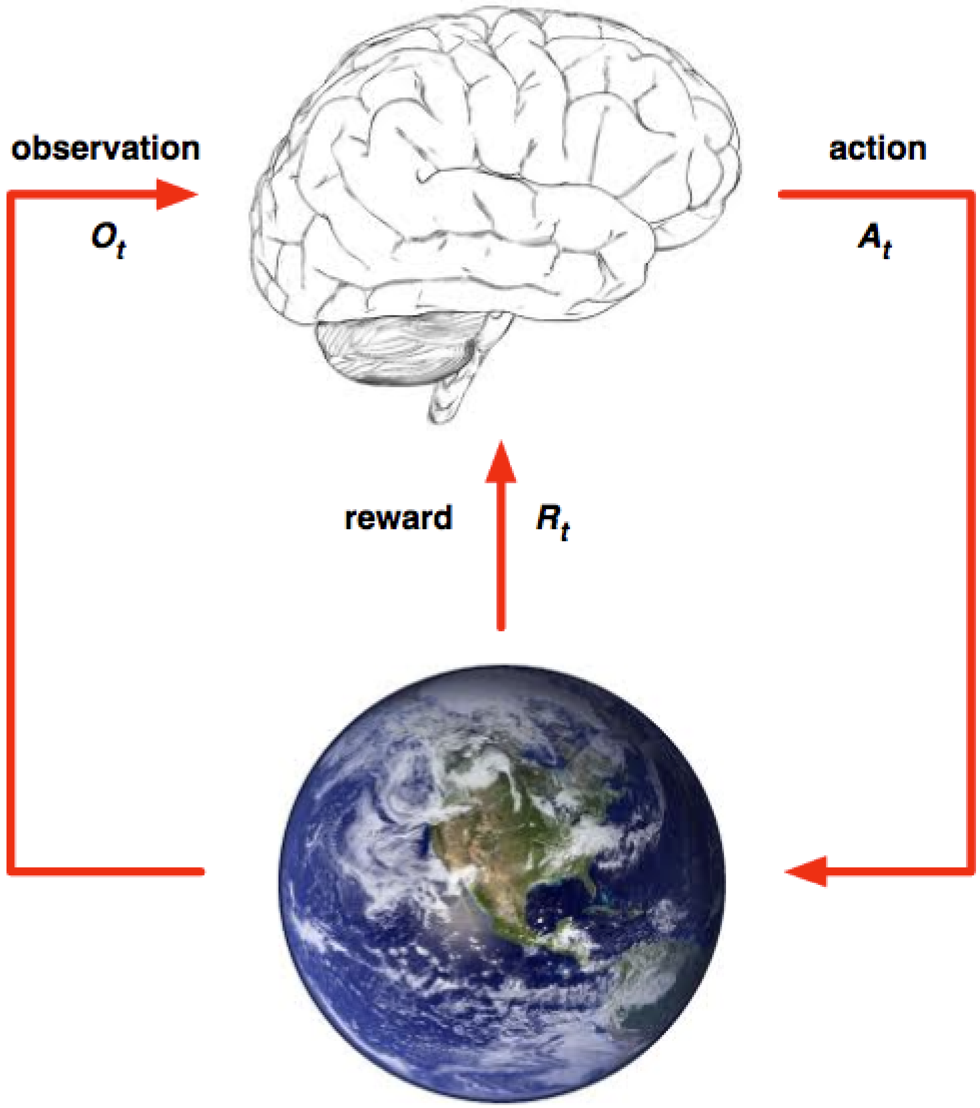
UNIVERSITY OF CAMBRIDGE

PolyAI

Spoken Dialogue Systems



Reinforcement learning

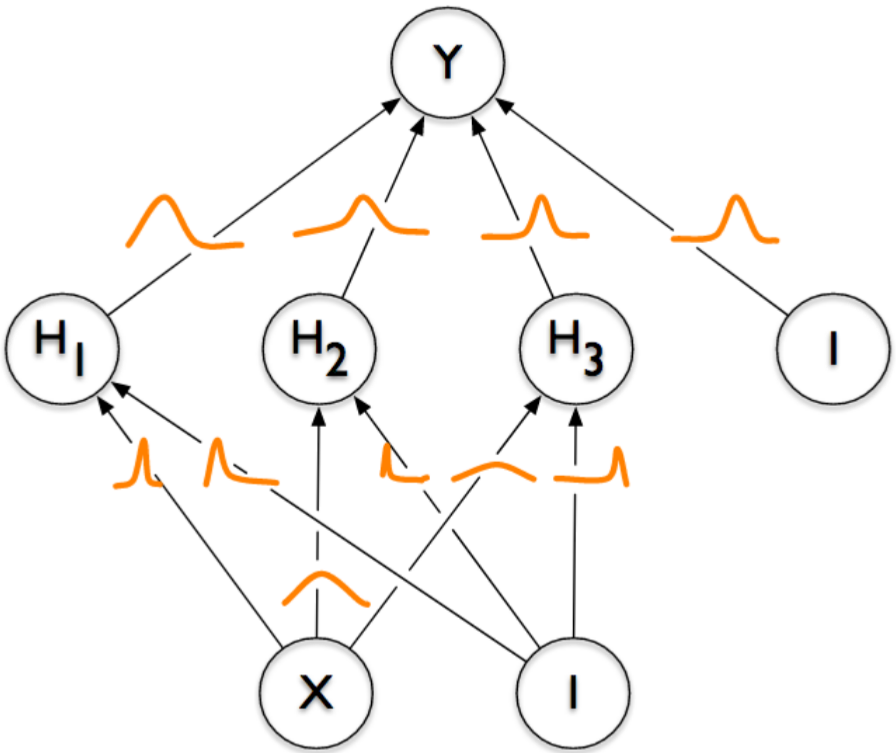


Problems of Deep RL in dialogue policy optimisation

1. No uncertainty estimates,

Uncertainty estimates

ICASSP18, with Christopher Tegho (Calipsa)

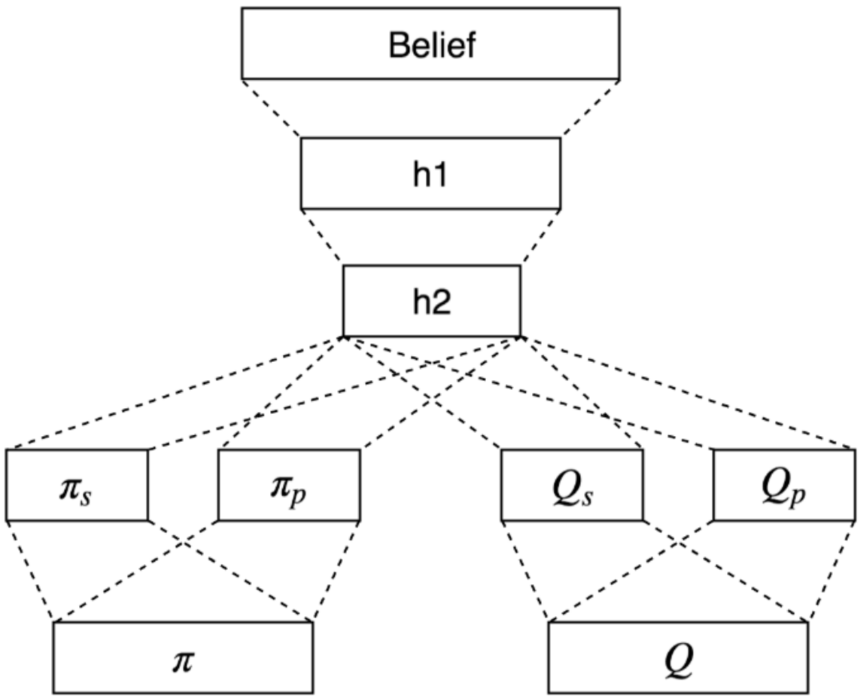


Problems of Deep RL in dialogue policy optimisation

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

Large-action space

TASLP18, with Gellert Weisz (DeepMind)

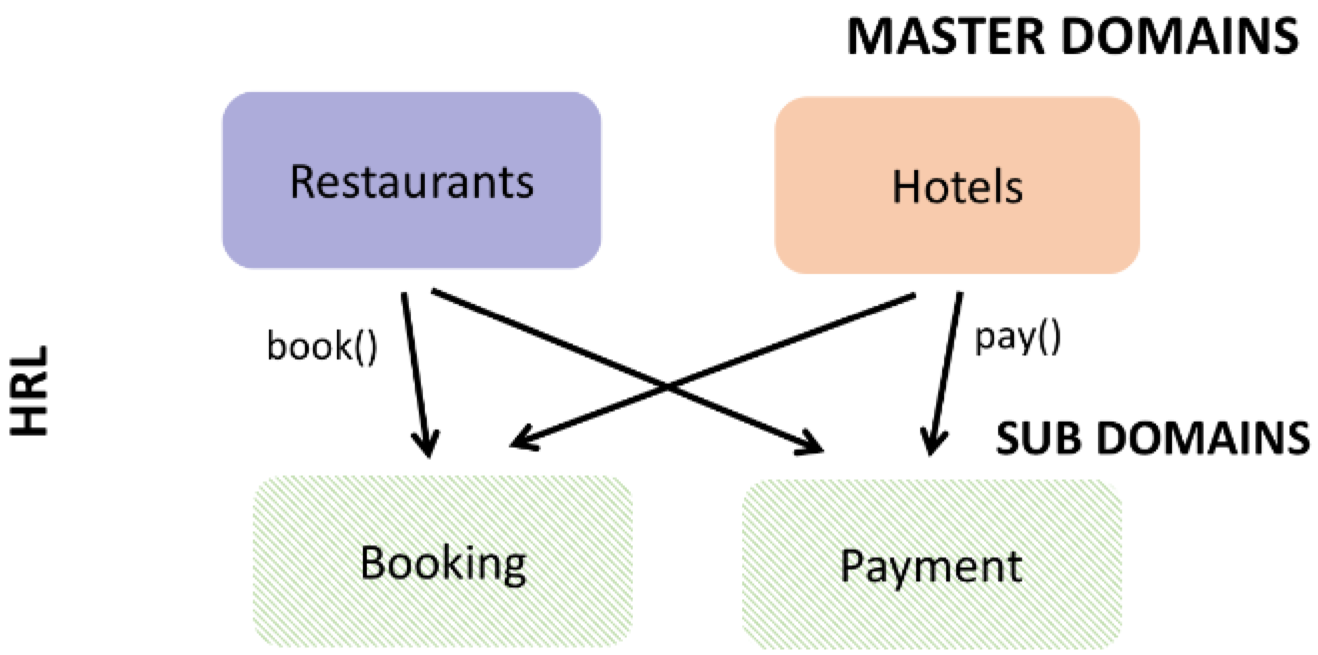


Problems of Deep RL in dialogue policy optimisation

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

Hierarchical Reinforcement Learning

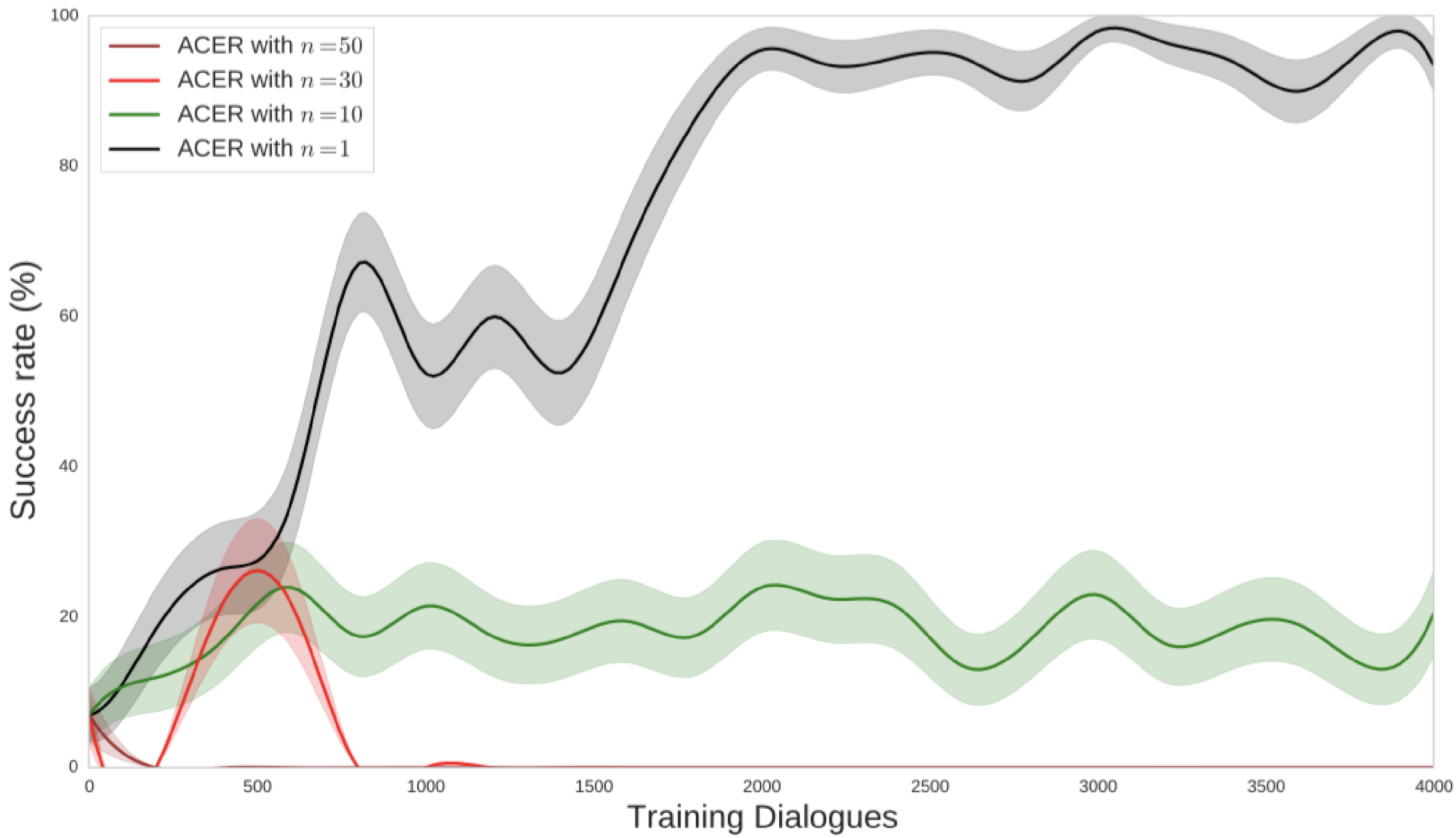
SIGDIAL18



Problems of Deep RL in dialogue policy optimisation

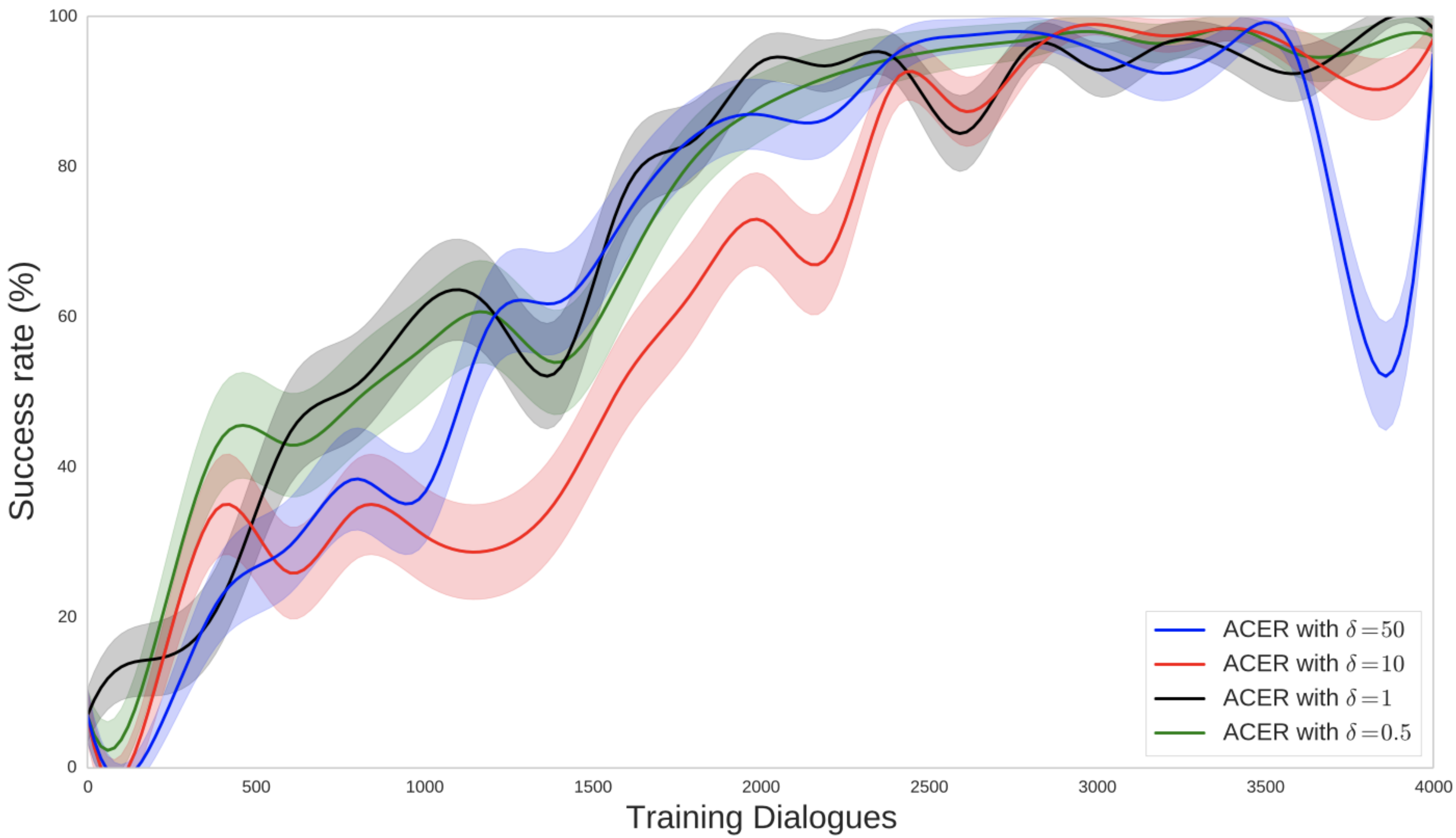
1. No uncertainty estimates,
2. Sample efficiency in large action space,
3. Multi-domain policy learning,
4. Cold start problem.

Problems?



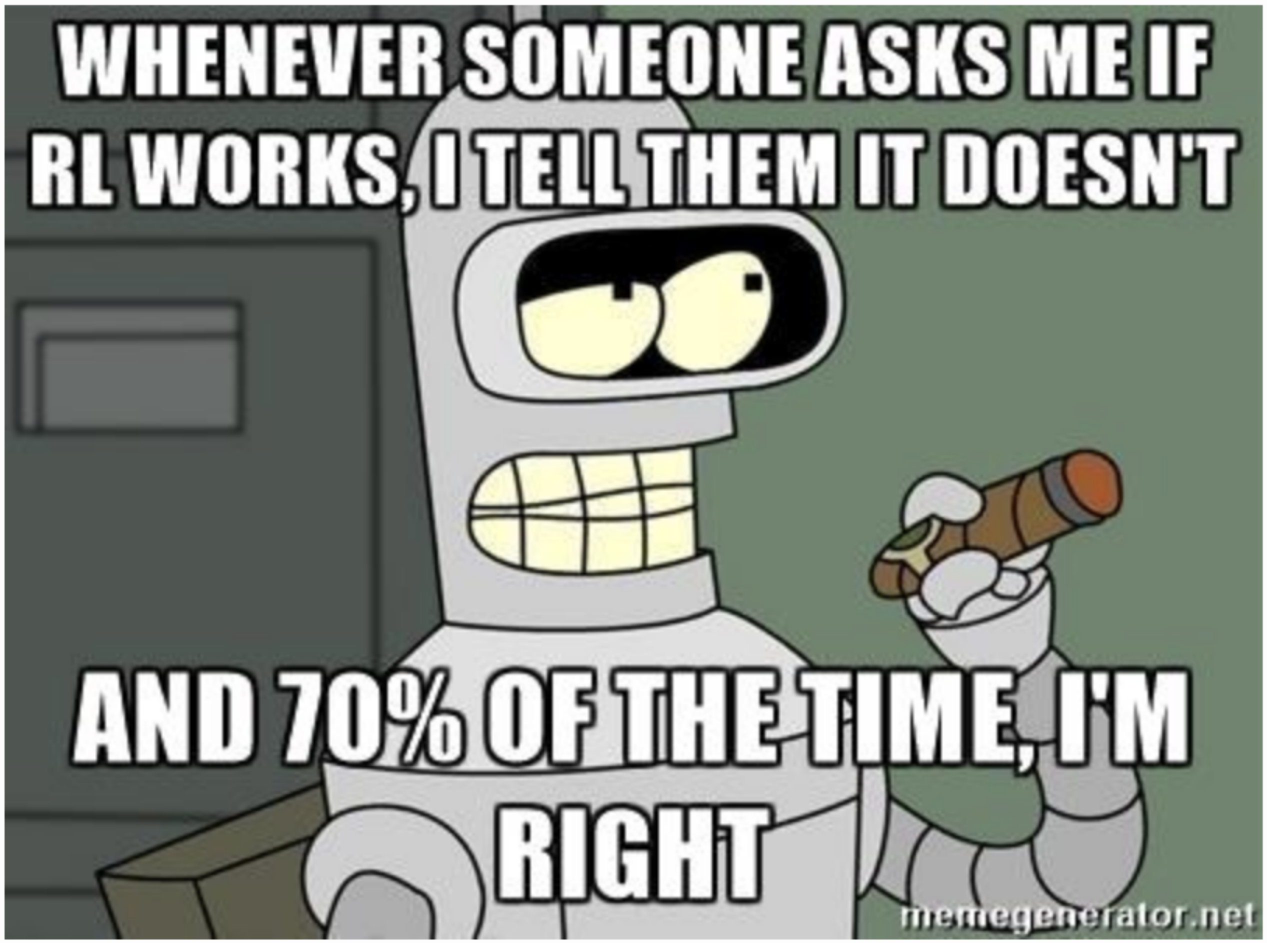
ACER with $n = 50$
ACER with $n = 30$
ACER with $n = 10$
ACER with $n = 1$
Success rate (%)
Training Dialogues
0
20
40
60
80
100
500
1000
1500
2000
2500
3000
3500
4000

Problems?



Success rate (%)
Training Dialogues
0
20
40
60
80
100
0
500
1000
1500
2000
2500
3000
3500
4000
ACER with $\delta = 50$
ACER with $\delta = 10$
ACER with $\delta = 1$
ACER with $\delta = 0.5$

Does RL work?



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

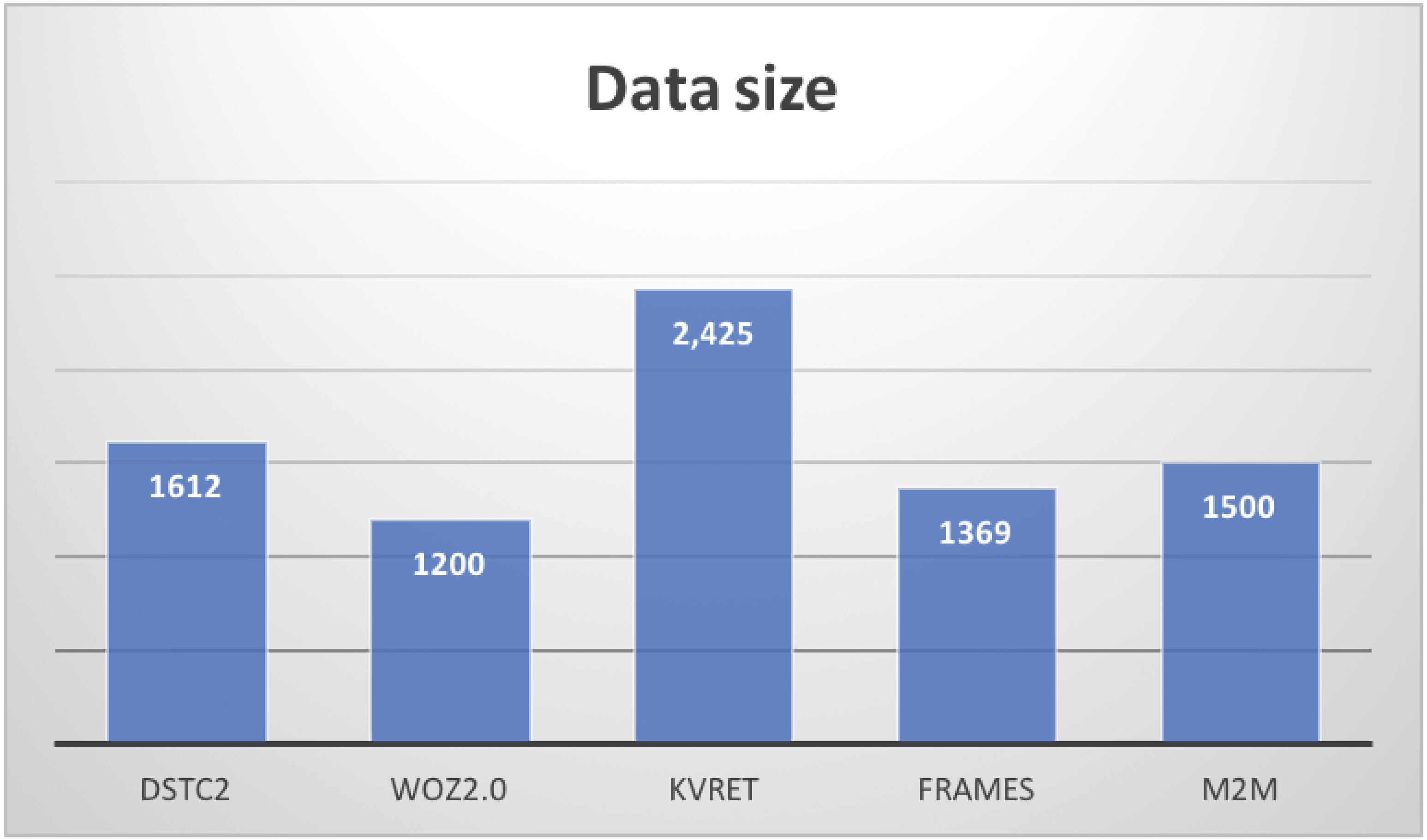
Cold start problem

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.

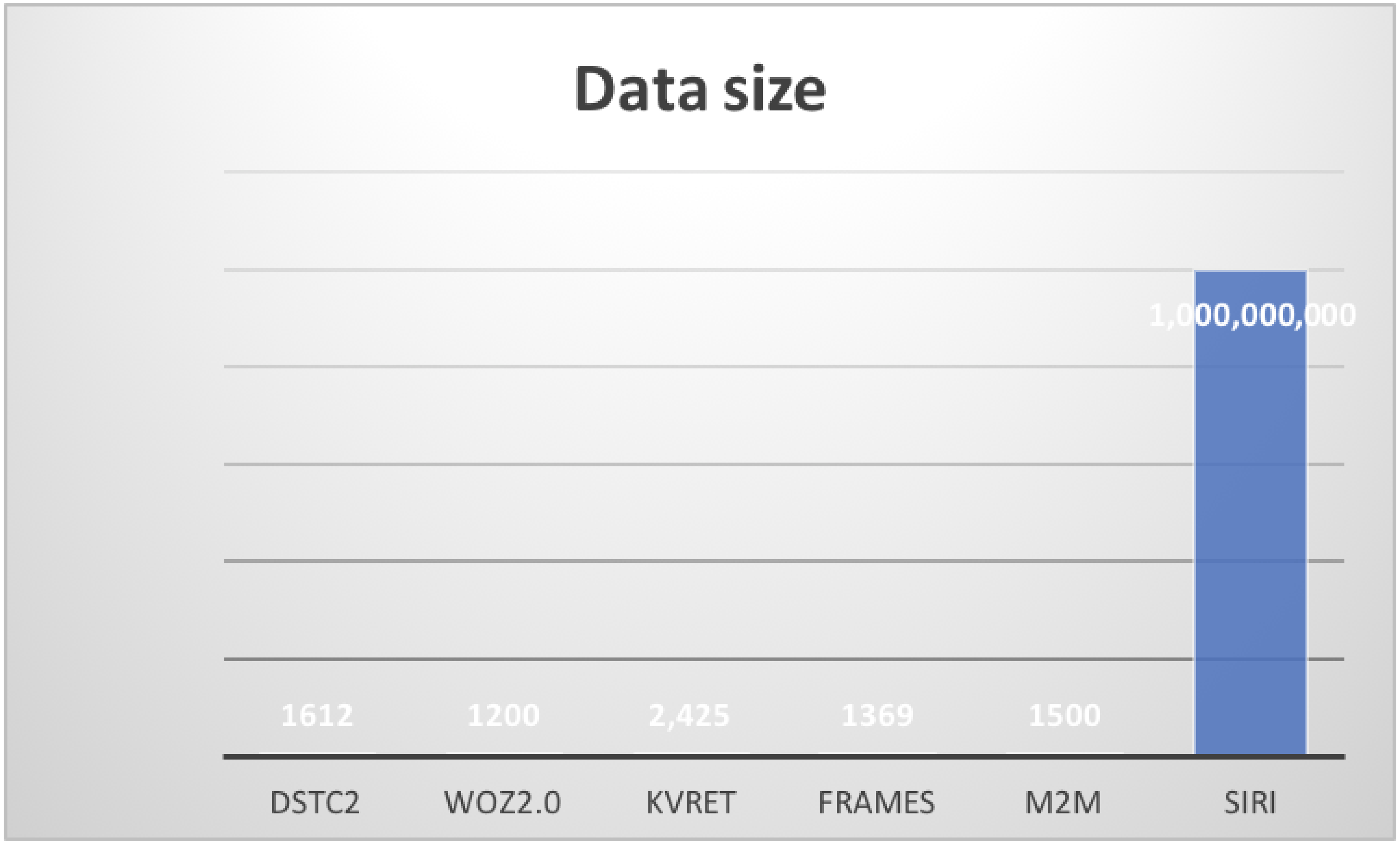
MultiWOZ - Large-Scale Multi-Domain Dataset for Task-Oriented Dialogue Modelling

EMNLP 2018, with Tsung-Hsien Wen (PolyAI)

Available corpora



Available corpora vs industry



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

Wen et al. 2017, Asri et al. 2017 and Eric et al. 2017

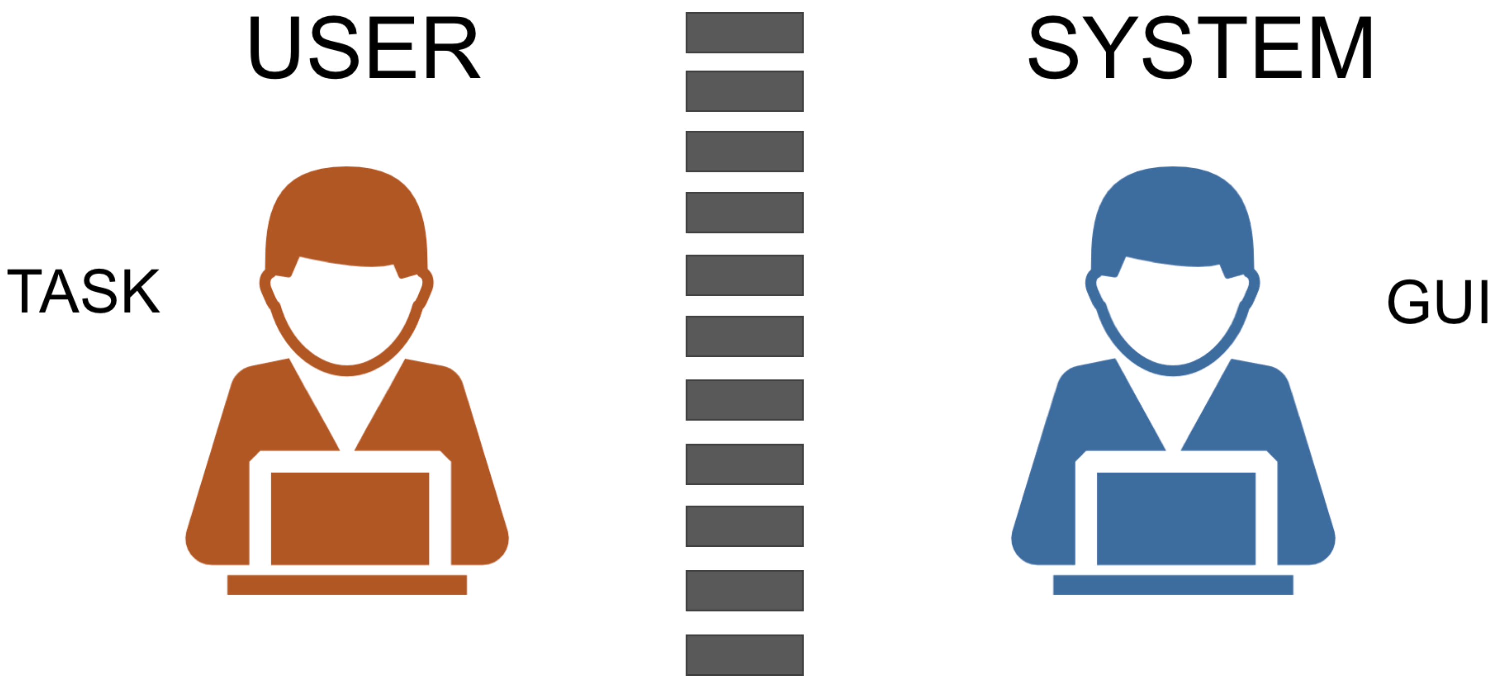
Motivation

1. Fully labeled, MTurk-based data collection set-up,
2. Large-scale, complex and multi-domain dataset,
3. Benchmarking platform.

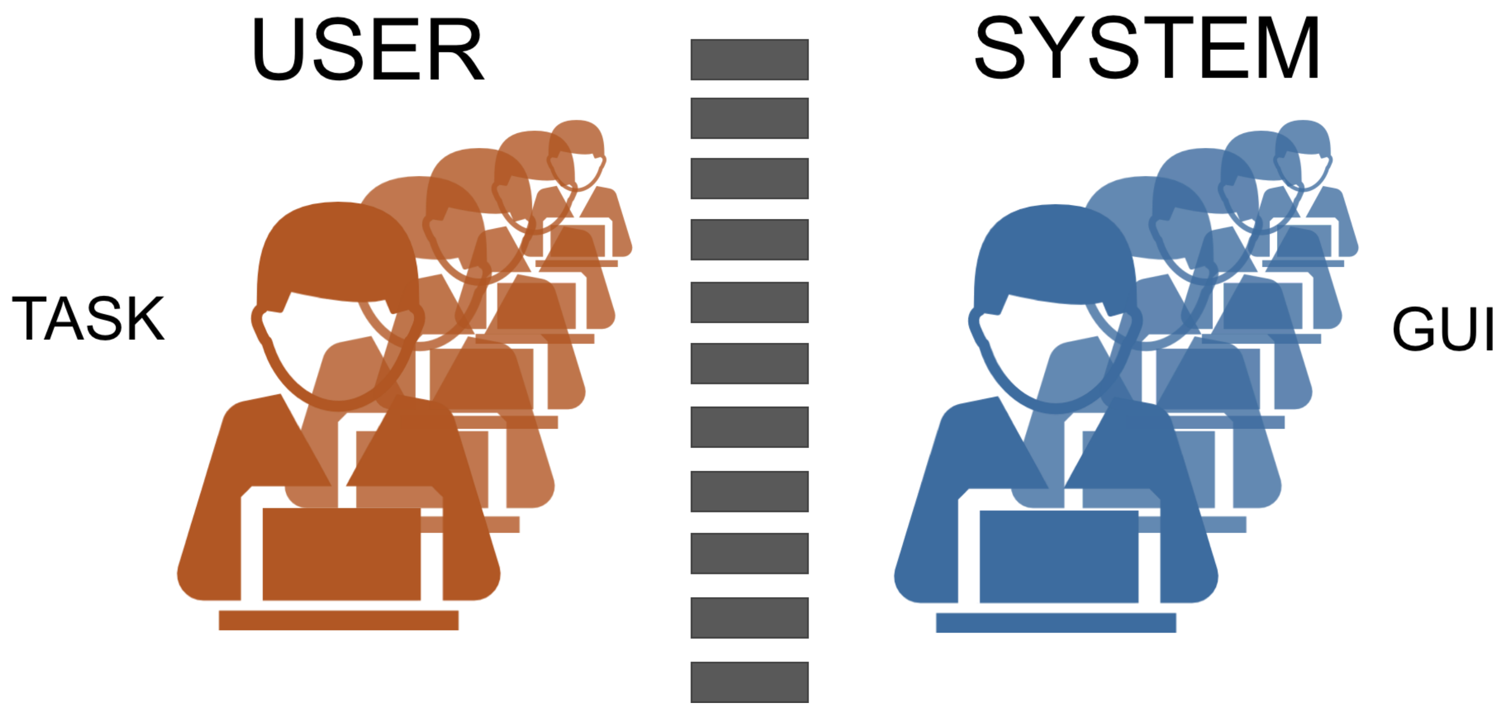
Data Collection Set-up

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



Wizard of Oz Setup



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.

User-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 :	Allenbell does have parking and the phone is 01223210353. Can I help with anything else?

Next turn

Customer : (**Your response**)

you need to go through the dialogue first by clicking the 'next turn' button

30

What topics were mentioned in **this turn**:

General: Booking: Restaurant: Hotels: Attraction:

Hospital: Police: Train: Taxi: Bus:

Submit the HIT

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 e type of **guesthouse**.
- Make sure you ge
- You are also look int should be in the **moderate** price r od.
- If there is no such es **turkish** food.
- Once you find the e for **4 people** at **17:45** on **friday**.
- Make sure you get the **reference number**

--- The End ---

User Task

Task

- You are traveling to Cambridge and looking forward to try local restaur

Your role

- You are looking 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**

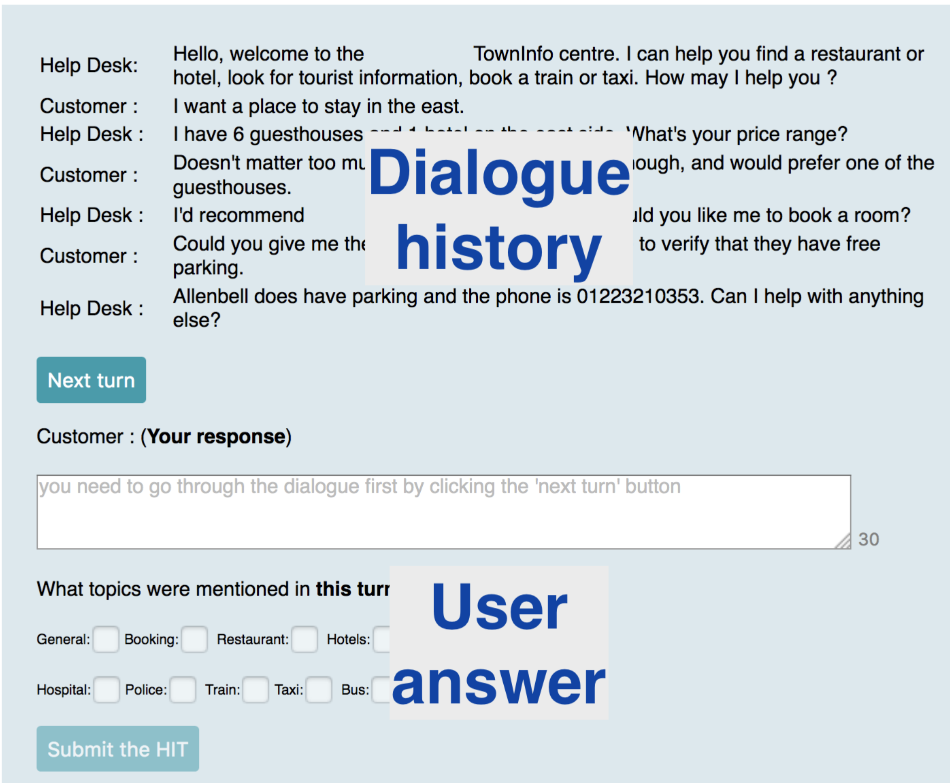
First Domain

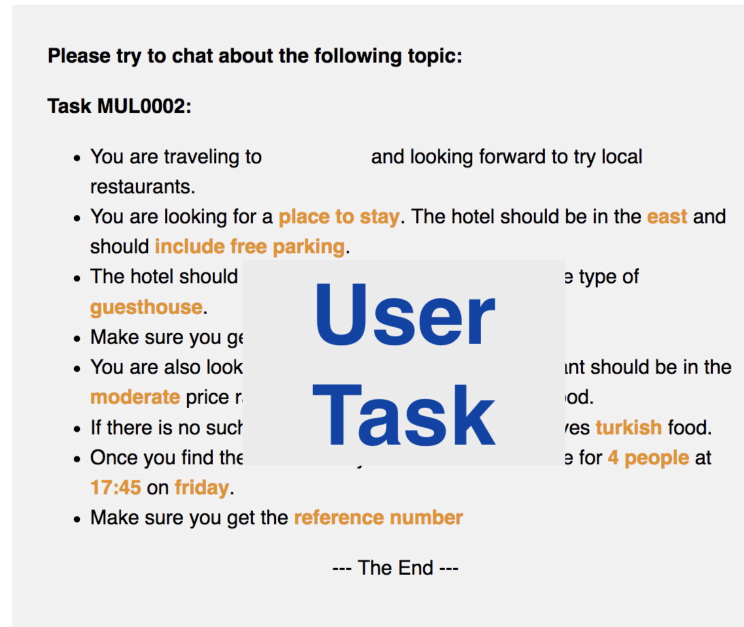
- Make sure
- You are al estaurant should serve **australasian** food and should be in the **moderate** price range.
- The restaurant should be **in the same area as the hotel**.
- If there is no such restaurant, how about one that serves **britis**

Second Domain

- 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


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
What's your price range?
Customer : Doesn't matter too mu
ough, and would prefer one of the guesthouses.
Help Desk : I'd recommend
ld you like me to book a room?
Customer : Could you give me the
to verify that they have free parking.
Help Desk : Allenbell does have parking and the phone is 01223210353. Can I help with anything else?
Dialogue history
Next turn
Customer : (Your response)
you need to go through the dialogue first by clicking the 'next turn' button
30
What topics were mentioned in this turn
General: Booking: Restaurant: Hotels:
Hospital: Police: Train: Taxi: Bus:
User answer
Submit the HIT


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
e type of guesthouse.
Make sure you ge
You are also look
nt should be in the moderate price r
od.
If there is no such
es turkish food.
Once you find the
e for 4 people at 17:45 on friday.
Make sure you get the reference number
--- The End ---
User Task

Belief state

User: I need to find a luxury hotel please.

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

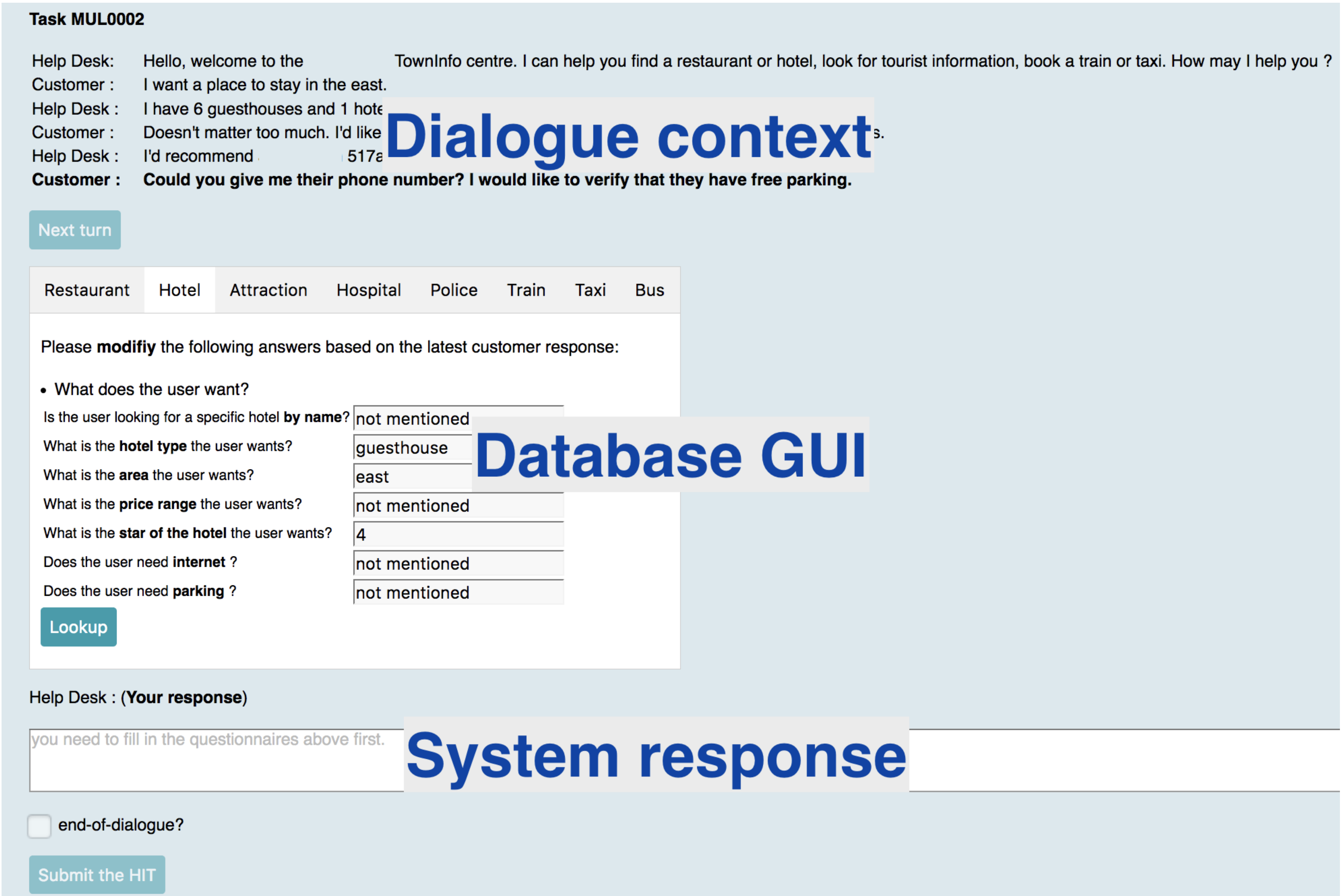
System: Pinehouse will be a good place.

User: Book me a room there, please.

```
Belief state: book(hotel)
```

System: Sure, what day and for how long?

System side



Task MUL0002
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 hote
Customer : Doesn't matter too much. I'd like
Help Desk : I'd recommend 517a
Customer : Could you give me their phone number? I would like to verify that they have free parking.
Dialogue context
Next turn
Restaurant
Hotel
Attraction
Hospital
Police
Train
Taxi
Bus
Please modifiy the following answers based on the latest customer response:
What does the user want?
Is the user looking for a specific hotel by name? not mentioned
What is the hotel type the user wants? guesthouse
What is the area the user wants? east
What is the price range the user wants? not mentioned
What is the star of the hotel the user wants? 4
Does the user need internet ? not mentioned
Does the user need parking ? not mentioned
Lookup
Database GUI
Help Desk : (Your response)
you need to fill in the questionnaires above first.
System response
end-of-dialogue?
Submit the HIT

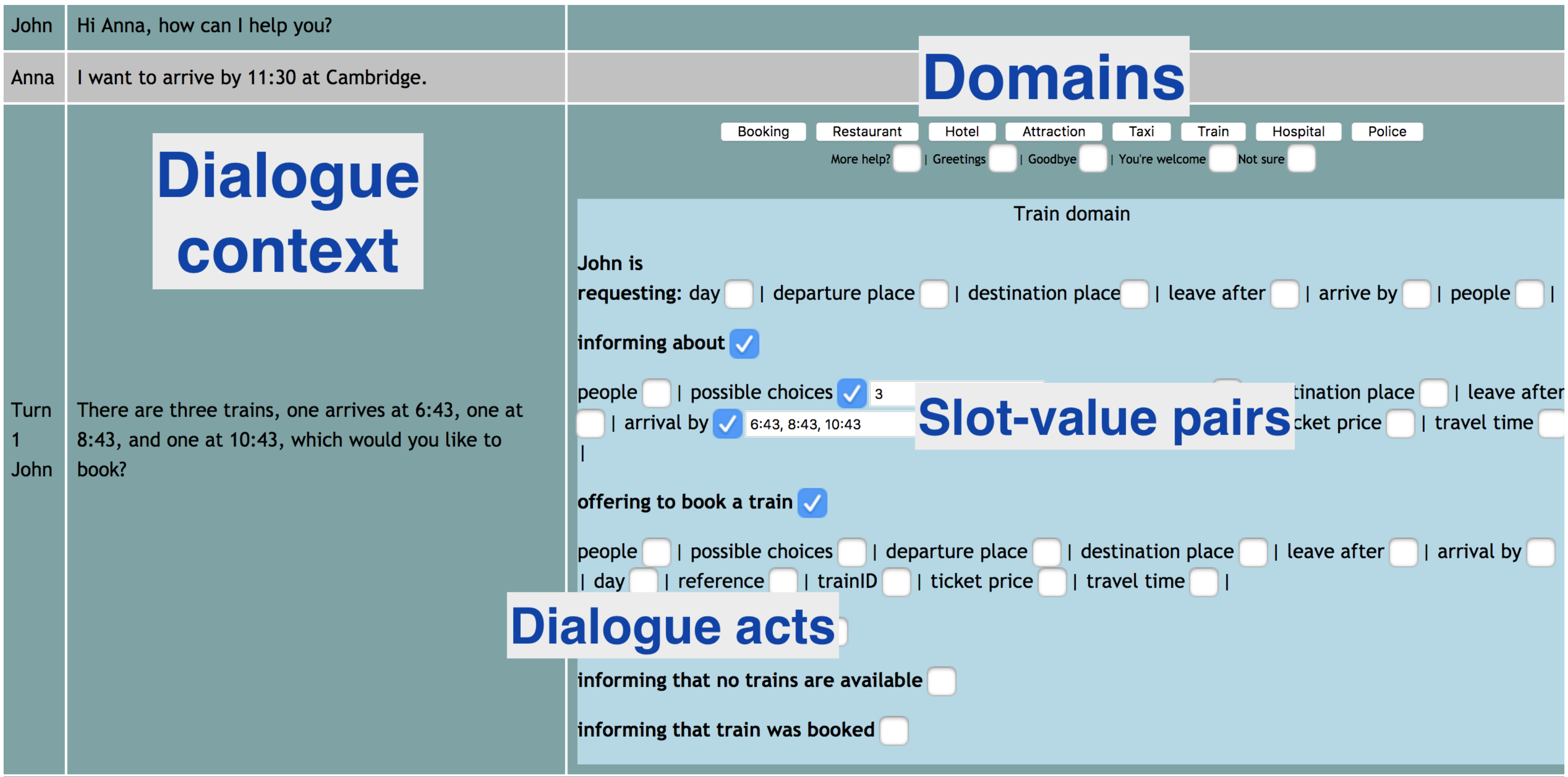
Annotation of system turns

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 where given more detailed instructions and required to go through the test again.

Annotation of system turns



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.

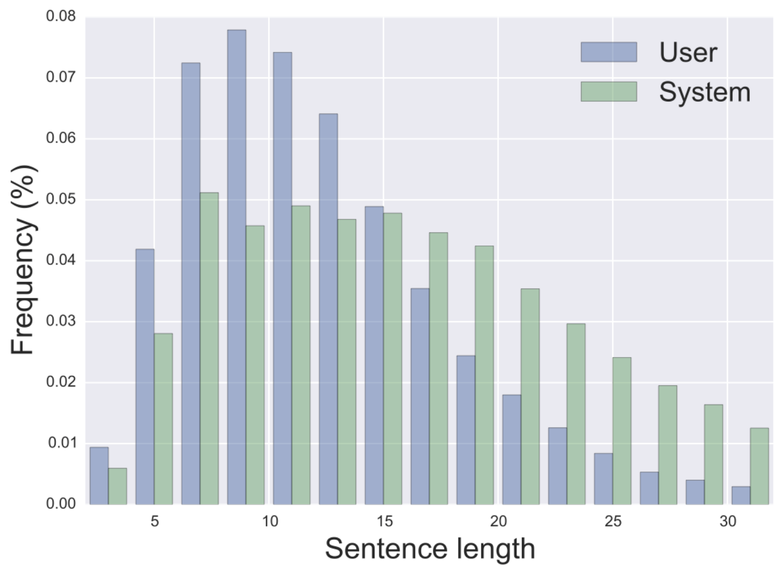
To make the dialogues more **complex and realistic** the initial goal for the user was sometimes impossible to accomplish.

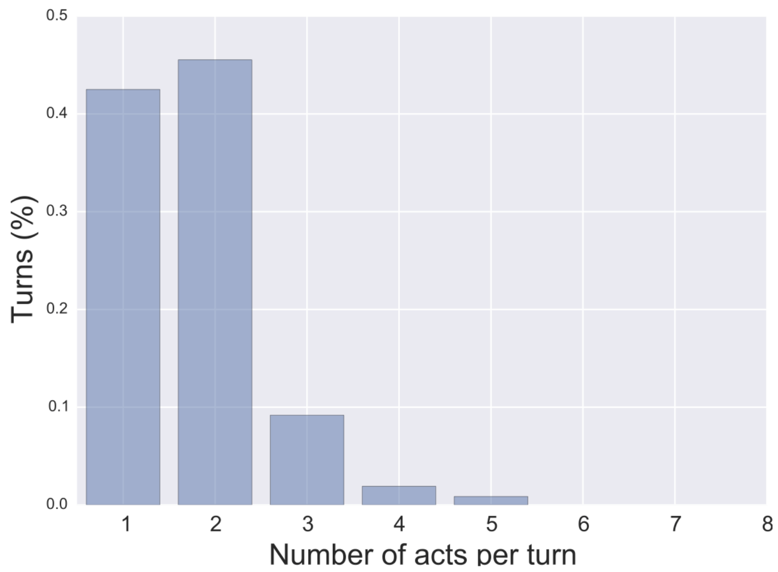
Data structure

	Single	Multi
# of dialogues	3,406	7,032
# of domains	1-2	2-6

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

Slot	WOZ 2.0	MultiWOZ (restaurant)
Joint goals	85.5	80.9

Dialogue State Tracking

Slot	WOZ 2.0	MultiWOZ (restaurant)	MultiWOZ (all domains)
Joint goals	85.5	80.9	25.8

Dialogue-Context-to-Text Generation

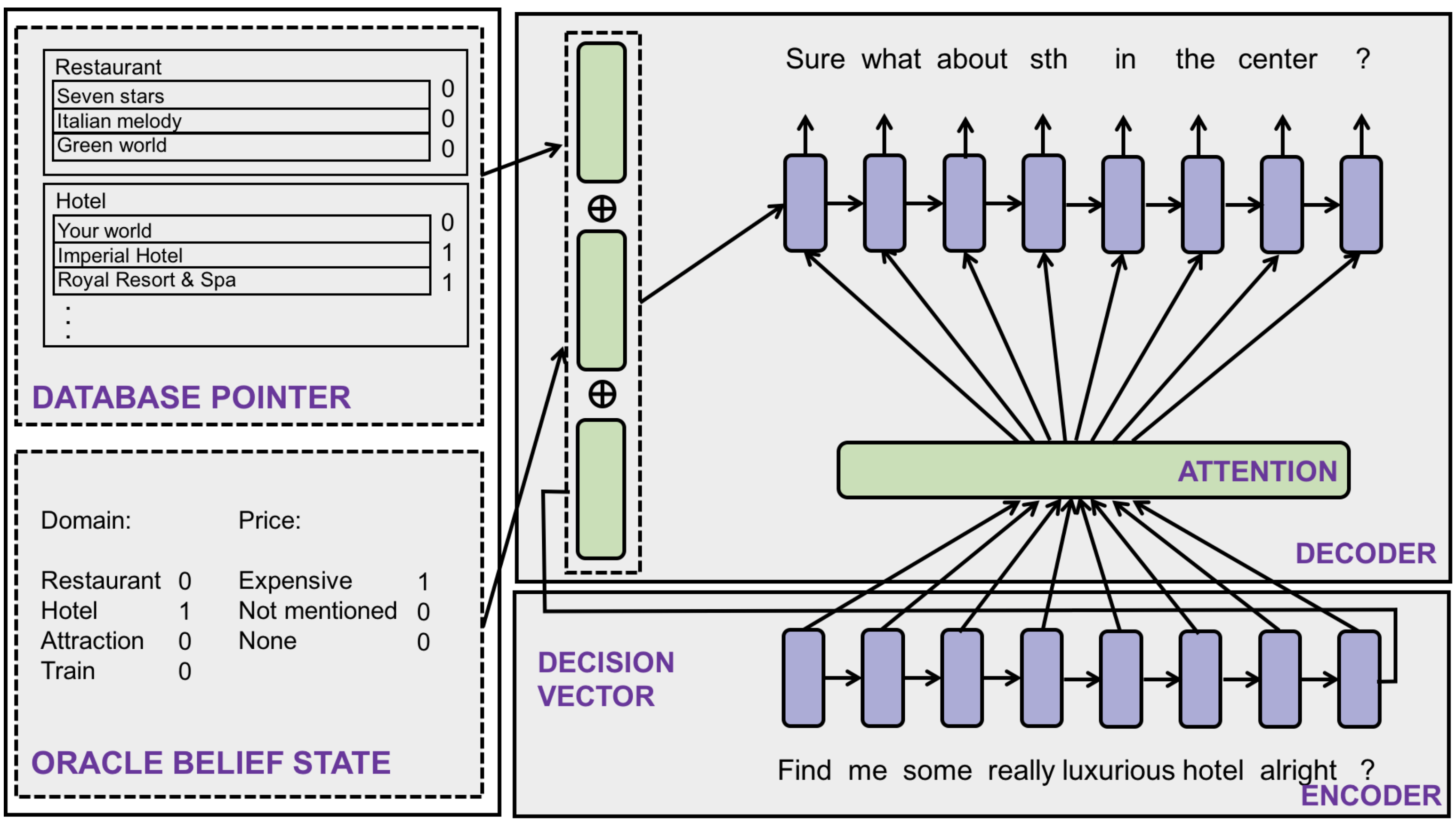
User: I need to find a luxury hotel please.
`Belief state``: inform(domain=hotel,price=expensive)`

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



Dialogue-Context-to-Text Generation

	Cam676		MultiWOZ	
	w/o attention	w/ attention	w/o attention	w/ attention
Inform (%)	99.17	99.58	71.29	71.33
BLEU	0.219	0.204	0.188	0.189

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

Metric	SFX	MultiWOZ (restaurant)
SER (%)	0.46	4.378
BLEU	0.731	0.616

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

Dialogue-act representation

Natural representation -

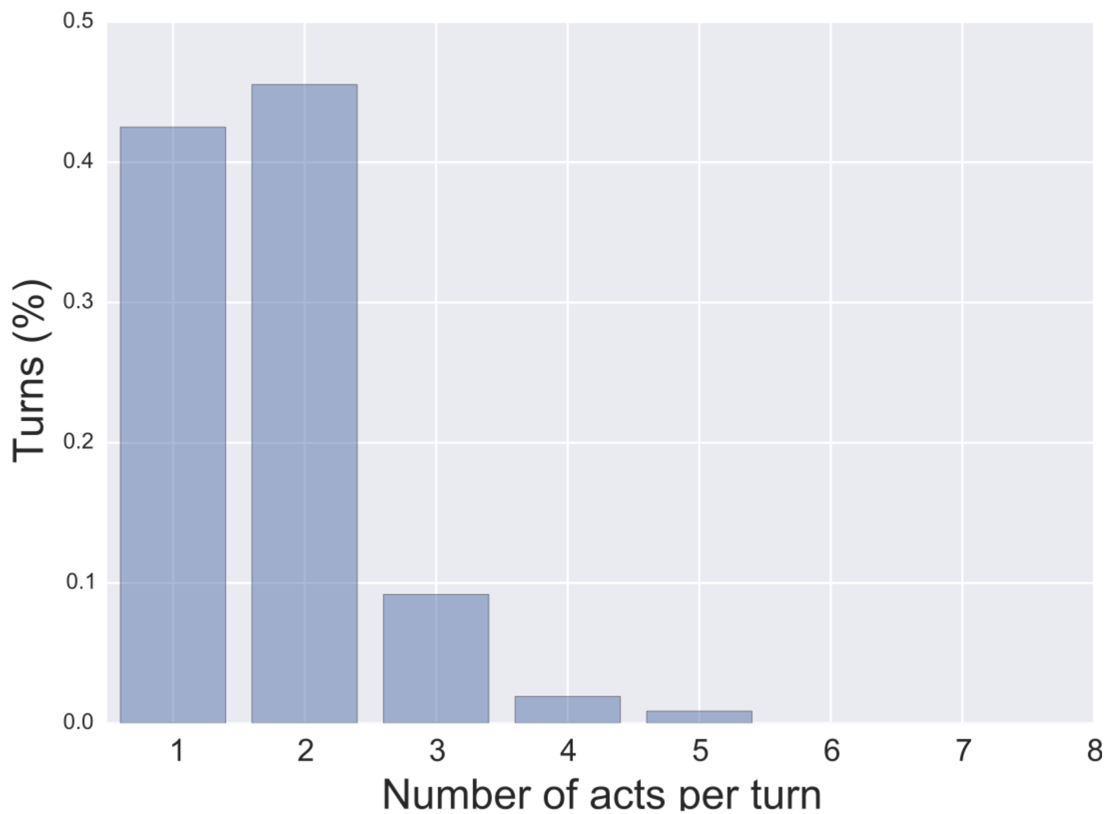
I have found 4 hotels satisfying your criteria. Do you have any preference for the area?

Action-slot-value representation -

Inform(domain=hotel, price=moderate, entities=4)

Request(domain=hotel, area)

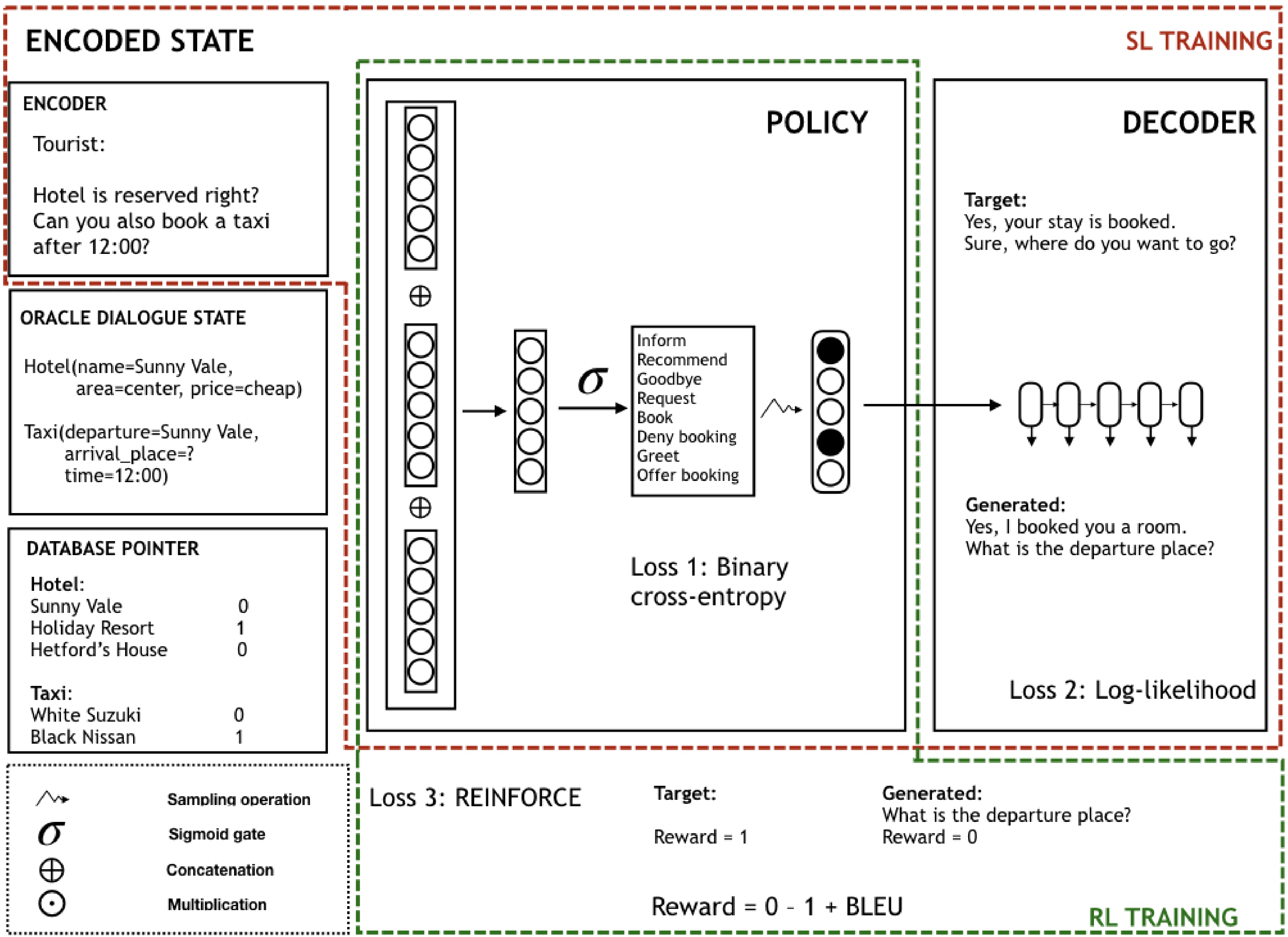
Concurrent actions



Action list in MultiWOZ

Dialogue act	System-specific?
Request	✗
Inform	✗
Request more	✓
Book	✗
Recommend	✗
Goodbye	✓
You are welcome	✓
Train was booked	✗
Train booking intent	✗
No entities available	✗
Booking not possible	✗
Select	✗
Greet	✓
No annotations	✗

Main architecture



Input to the network

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

$$\mathbf{e}_t = \mathbf{u}_t^L.$$

We model also prediction of dialogue state:

$$\mathbf{b}_t = \bigoplus_d \bigoplus_s \mathbf{b}_{s_d,t}.$$

And list of entities satisfying current constraint:

$$\mathbf{kb}_t = \bigoplus_d \mathbf{kb}_{d,t}.$$

Policy modelling

This serves as an input:

$$\mathbf{x}_t \ = \ \mathbf{e}_t \ \oplus \ \mathbf{b}_t \ \oplus \ \mathbf{kb}_t$$

to predict probabilities over action set:

$$\pi(\mathbf{a}_t|\mathbf{x}_t) = \mathrm{MLP}(\mathbf{x}_t).$$

Sigmoid vs Softmax

Actions are sampled from the derived probabilities:

$$\mathbf{a}_t \; \sim \; \pi(\mathbf{a}_t | \mathbf{x}_t).$$

The set of all possible actions $\mathcal{A} = \{a_1, a_2, ..., a_N\}$ 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 $\pi : \mathcal{X} \to \mathcal{A}$.

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

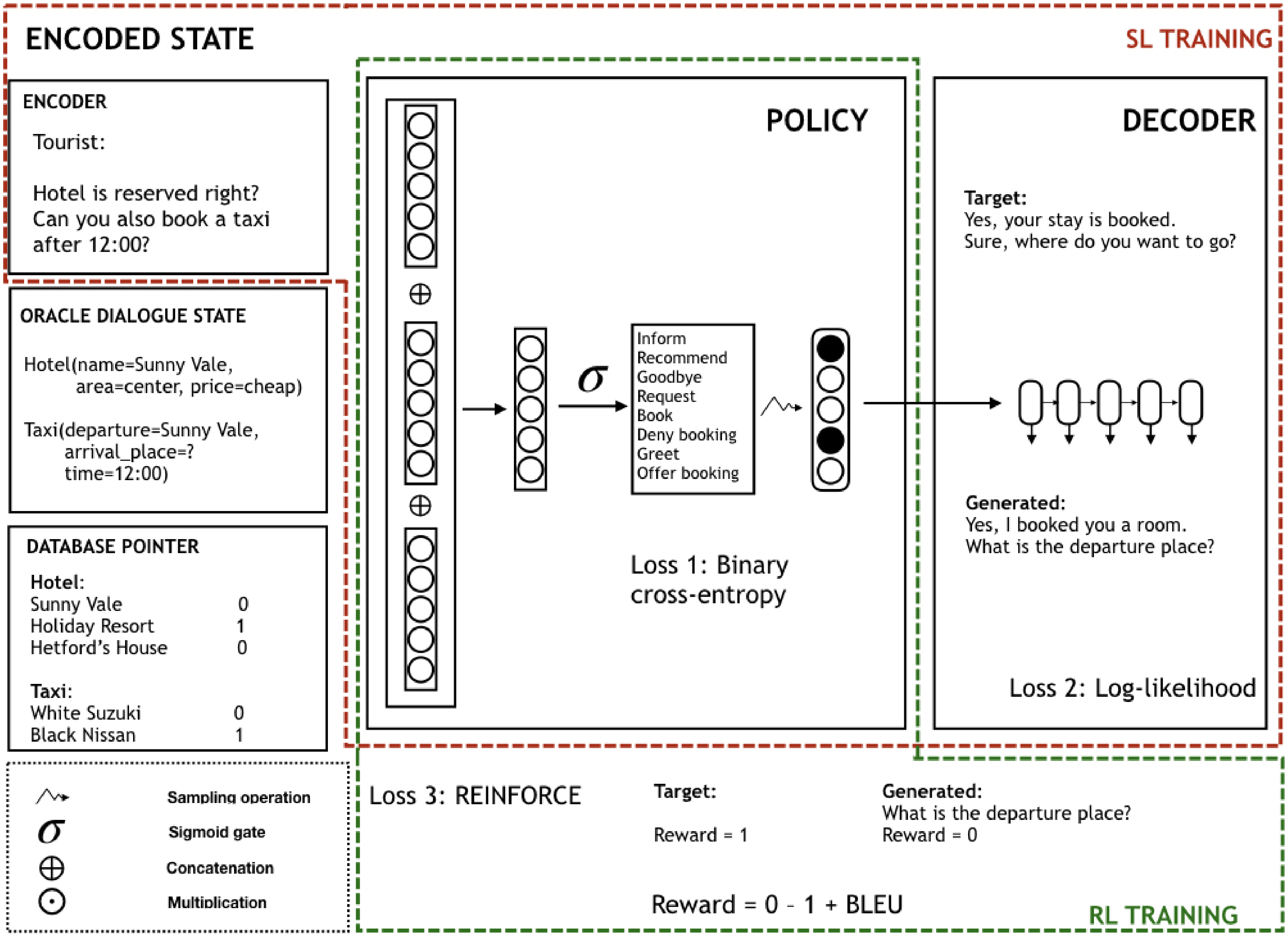
Cross-entropy over output words:

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

Cross-entropy over actions predictions:

$$L_2(\theta) = \sum_d \sum_t (\mathbf{1} - \mathbf{a}_t)(\mathbf{1} - \log \mathbf{p}_a^t) + \mathbf{a}_t \log \mathbf{p}_a^t.$$

Main architecture



Reinforcement learning fine-tuning phase

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

$$L_3 = \frac{1}{T}\sum_{t}^{T} \nabla \log \pi_\theta(\mathbf{a}_t|\mathbf{x}_t) r_t.$$

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.

$$\pi_\theta(\mathbf{a}_t|\mathbf{x}_t) = \prod_{n=1}^{N} \pi(a_t^n|\mathbf{x}_t).$$

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

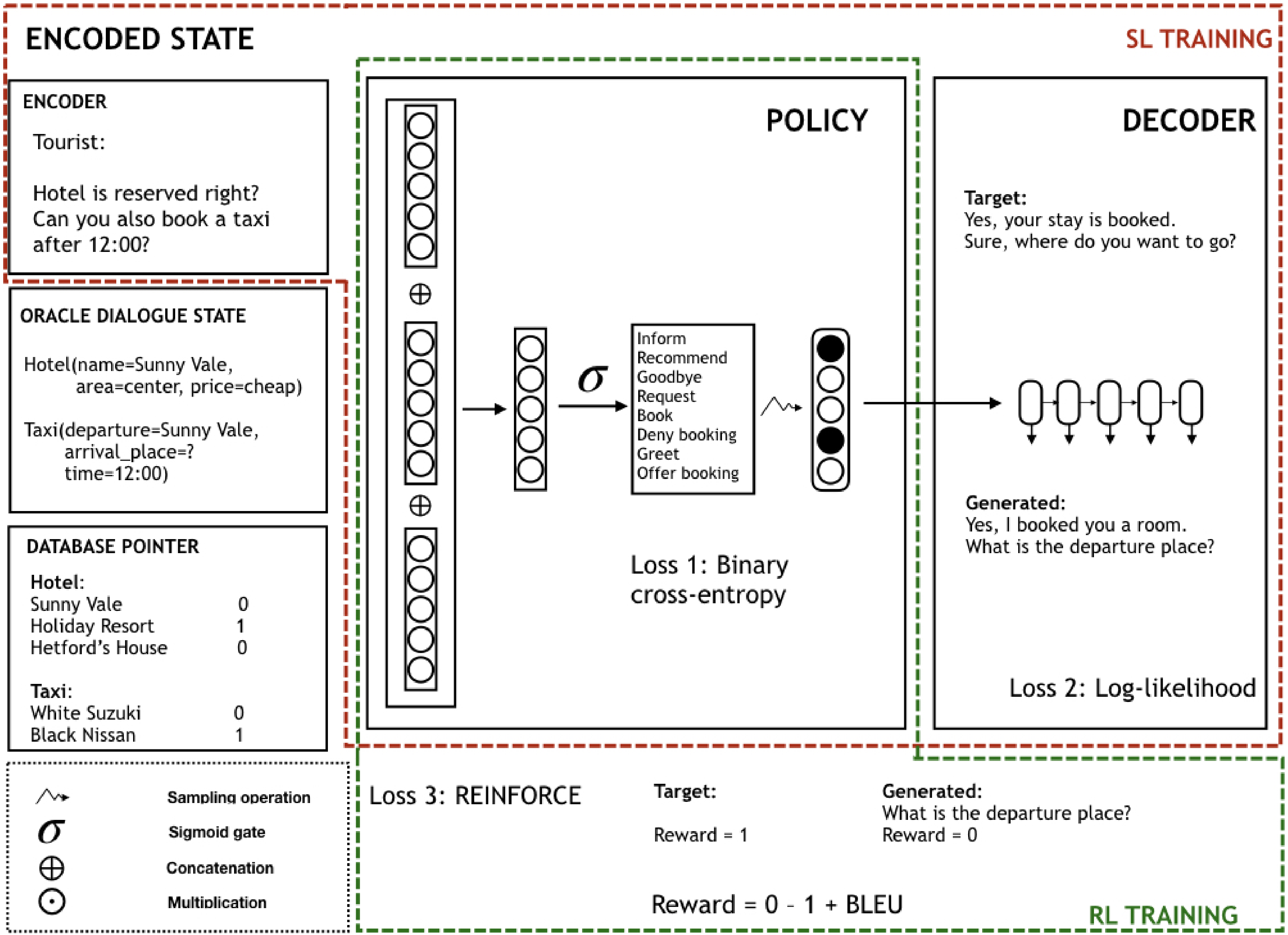
$$\pi(\mathbf{a}_t|\mathbf{x}_t) = \prod_{n=1}^{N} \left(a_t^n z_t^n + (\mathbf{1} - a_t^n)(\mathbf{1} - z_t^n)\right).$$

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

$$0.5 * \text{Inform} + 0.5 * \text{Success} + \text{BLEU}$$

Experiments - SL phase

Constrained set:

	Baseline	Sigmoid	Softmax
Inform (%)	78.71	82.49	82.19
Success (%)	65.21	66.95	68.11
BLEU (%)	17.7	18.8	18.79

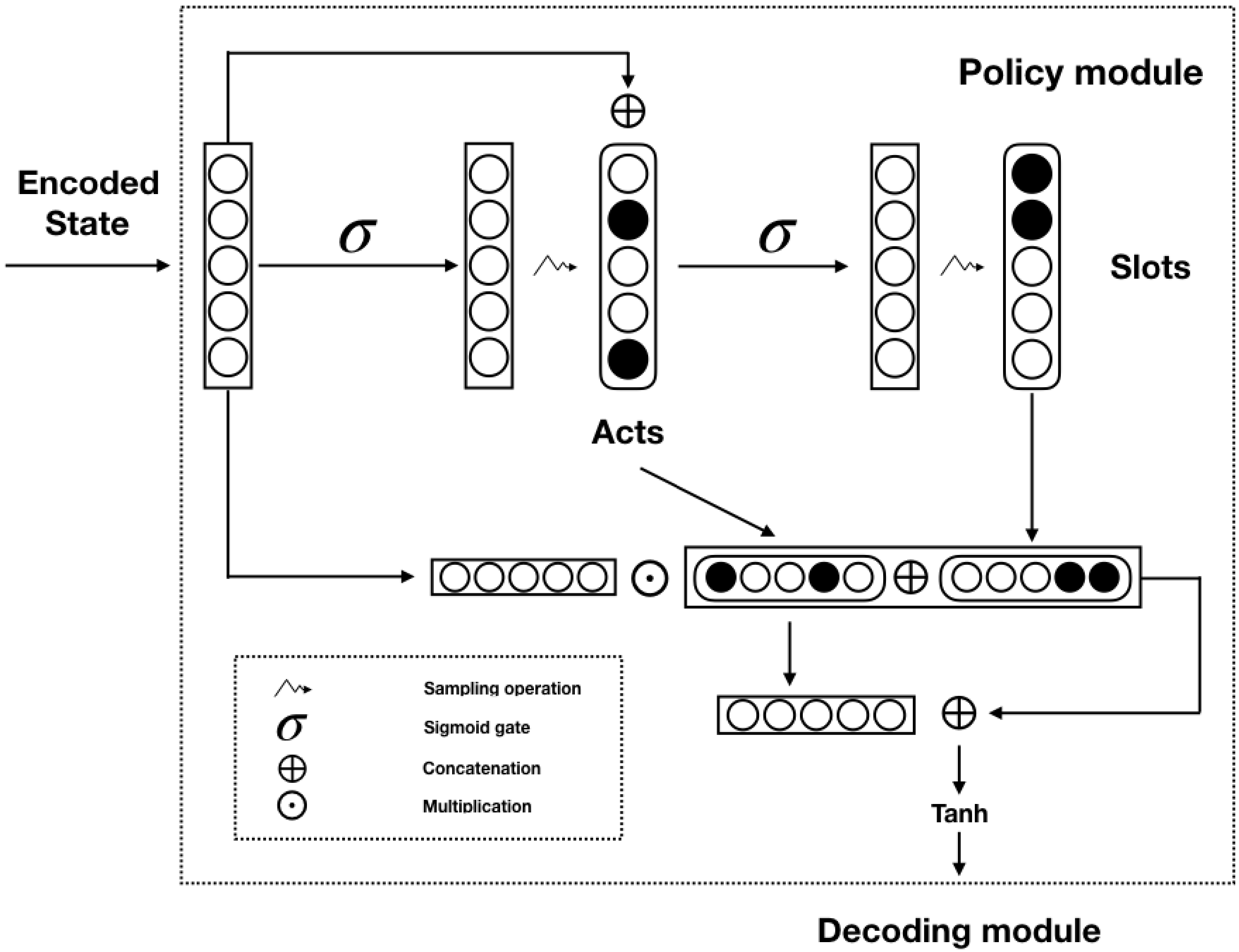
Full set:

	Sigmoid	Softmax
Inform (%)	84.15,	83.0
Success (%)	67.42	34.339
BLEU (%)	17.76	9.8

Experiments - RL phase

	Sigmoid (subset)	Sigmoid (full set)	Softmax (subset)	Softmax (full set)
Inform (%)	84.519	**85.789**	82.1	83.08
Success (%)	67.57	**67.63**	48.18	34.34
BLEU (%)	**17.36**	15.6	14.8	9.8

Multi-Action and Slot Dialogue Agent



Additional loss

Cross-entropy over slots predictions:

$$L_4(\theta) = \sum_d \sum_t (\mathbf{1} - \mathbf{s}_t)(\mathbf{1} - \log \mathbf{p}_s^t) + \mathbf{s}_t \log \mathbf{p}_s^t.$$

Final loss consists of:

$$L = L1(words) + L2(actions) + L4(slots) + L3(RL).$$

Full model evaluation

	Sigmoid SL	Sigmoid RL
Inform (%)	83.97,	**88.34**
Success (%)	71.44	**75.41**
BLEU (%)	16.78	15.95

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?