Dialogue State Tracking

Gulyaev Pavel

Neural Networks and Deep Learning Lab

Moscow Institute of Physics and Technology

Goal-oriented dialogue system

Book a table at Cascal for 6 pm today.

Sure. For how many people?

Task-Oriented Dialogue System

```
get_availability(
        restaurant="Cascal",
        time = "18:00:00",
        date = "2018-11-28",
)
```

API


```
{
  "restaurant" : "Cascal",
  "time" : "18:00:00",
  "date" : "2018-11-28",
  "seats" : 5
}
```

Goal-oriented dialogue system

Book a table at Cascal for 6 pm today.

Sure. For how many people?

Language Understanding and State Tracking

Policy

Response Generation

```
Intent: RESERVE_RESTAURANT
restaurant: "Cascal"
time: "6 pm"
date: "today"
```

API

```
[
    AFFIRM,
    REQUEST(SEATS),
]
```

Challenges from Real World APIs

- The number of APIs to support is large
- No uniform schema or entity names
  - Slot names
    - "origin" = "from", "departing from" = "where from", ……
  - Slot values
    - "San Francisco" = "SFO" = "SF", ……
- The set of supported entities may be large and dynamic
  - May not expose the list of all available entities
  - Robustness to OOV problem
- Don't allow arbitrary calls
  - find_restaurant(*city=”San Francisco”*, date=”today”)
  - find_restaurant(city, *date*="today")

Requirements of Virtual Assistants

- Facilitate dialogues across multiple Services/APIs
  - Handle large universe of services spanning multiple domains
  - Data efficient: Zero-shot or few-shot learning for tail services
- Can handle multi-domain conversations
  - Seamlessly switching domains
  - Carrying over relevant slots among APIs
- Robust to changes in schema
  - No retraining if the interface of an API changes
  - Robust to new/unseen slot values

Existing Dialogue Datasets

Metric	DSTC2	WOZ2.0	FRAMES	M2M	MultiWOZ
#Domains	1	1	3	2	**7**
#Dialogues	1 612	600	1 369	1 500	**8 438**
#Turns	23 354	4 472	19 986	14 796	**113 556**
#Slots	8	4	61	13	**24**
#Values	212	99	3 871	138	**4 510**

- Datasets not large enough for training generic virtual assistants
- Very few domains, slots and possible values
- Make simplified assumptions on underlying APIs/Services

The Schema-Guided Dialogue Dataset (SGD)

Metric	DSTC2	WOZ2.0	FRAMES	M2M	MultiWOZ	SGD
#Domains	1	1	3	2	7	**16**
#Dialogues	1 612	600	1 369	1 500	8 438	**16 142**
#Turns	23 354	4 472	19 986	14 796	113 556	**329 964**
#Slots	8	4	61	13	24	**214**
#Values	212	99	3 871	138	4 510	**14 139**

- **Largest** publicly available dataset for task-oriented dialogues
- Contains multi-domain conversations involving **34 services** over **16 domains**
- Test set contains few unseen services to test zero shot generalization

Example - single domain

Okay, it's family movie night. We want to see a good movie that's family oriented.

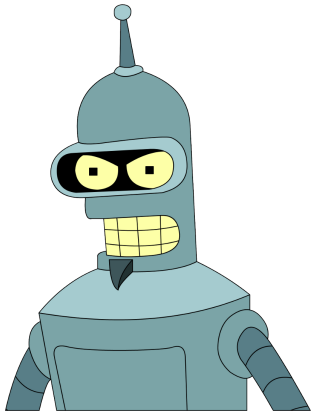
Well, my first choice is Captain Marvel.

Oh. No, I want us to see a movie with subtitles tonight. I'm thinking The Curse of La Llorona movie might be a good choice if it has subtitles.

Do you know the title of the movie you want to see?

Okay, Captain Marvel is available. There are no subtitles with this movie. Please confirm if this is the movie you want to see.

Yes, The Curse of La Llorona film is available with subtitles. If you want to see that movie, please confirm.



Movie Domain

Example - multi domain

Please tell me the weather forecast for March 8th.

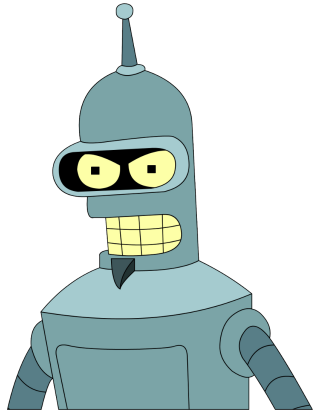
Please check the weather in **Menlo Park**.

Sounds good! Please find me a restaurant offering Latin American cuisine there.

Any particular city's weather forecast?

The expected average temperature for the day is 83 degrees Fahrenheit, with a 9 percent chance of rain.

Cafe Del Sol Restaurant, a nice restaurant **Menlo Park** is among the 3 I found.



Weather Domain → Restaurant Domain

Challenge Summary

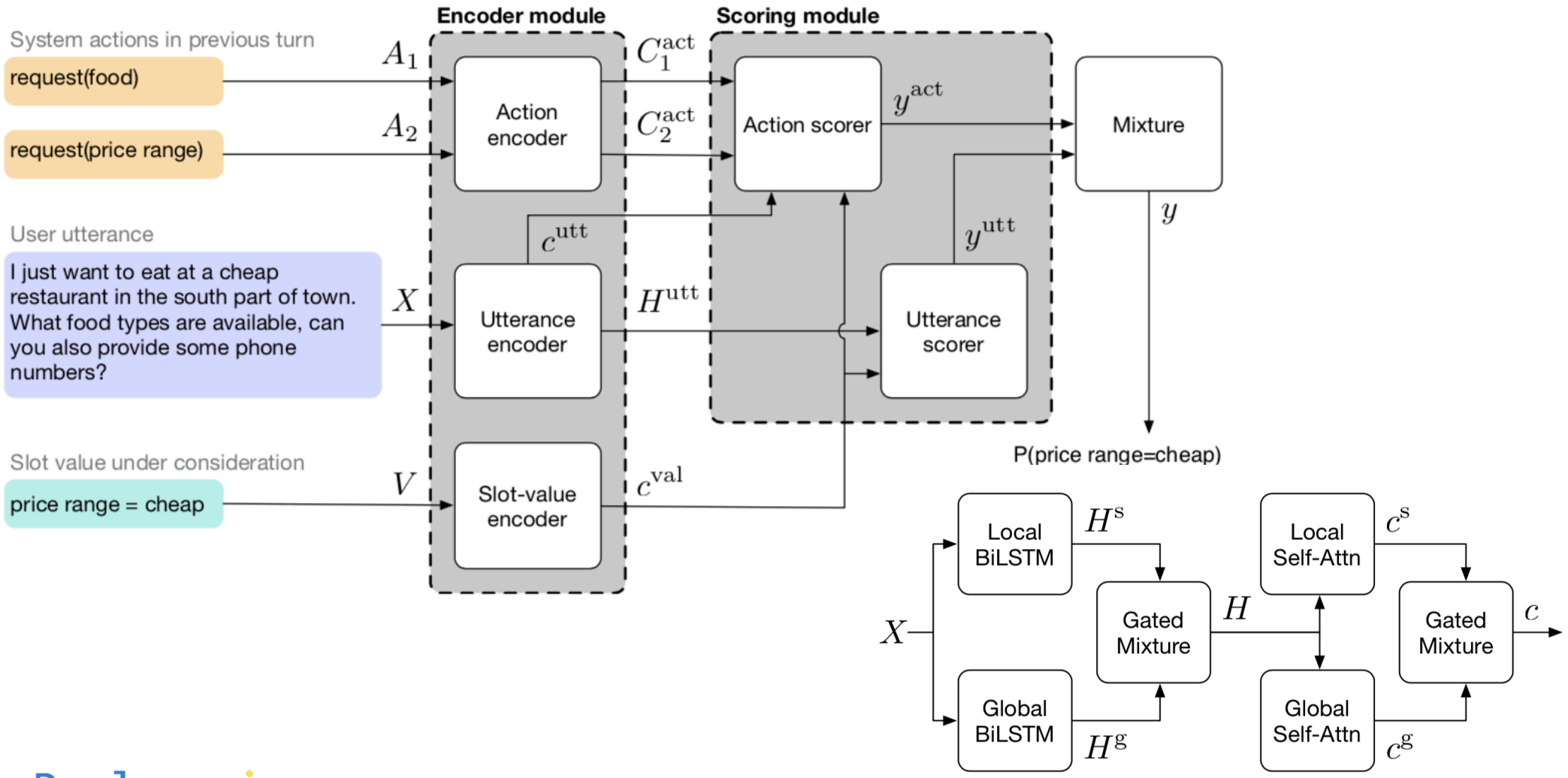
- Aim
  - Dialogue state tracking in a virtual assistant

- New Challenges
  - Zero-shot learning of unseen services
  - Handling OOV slot values
  - Transfer of slot values in multi-domain dialogues

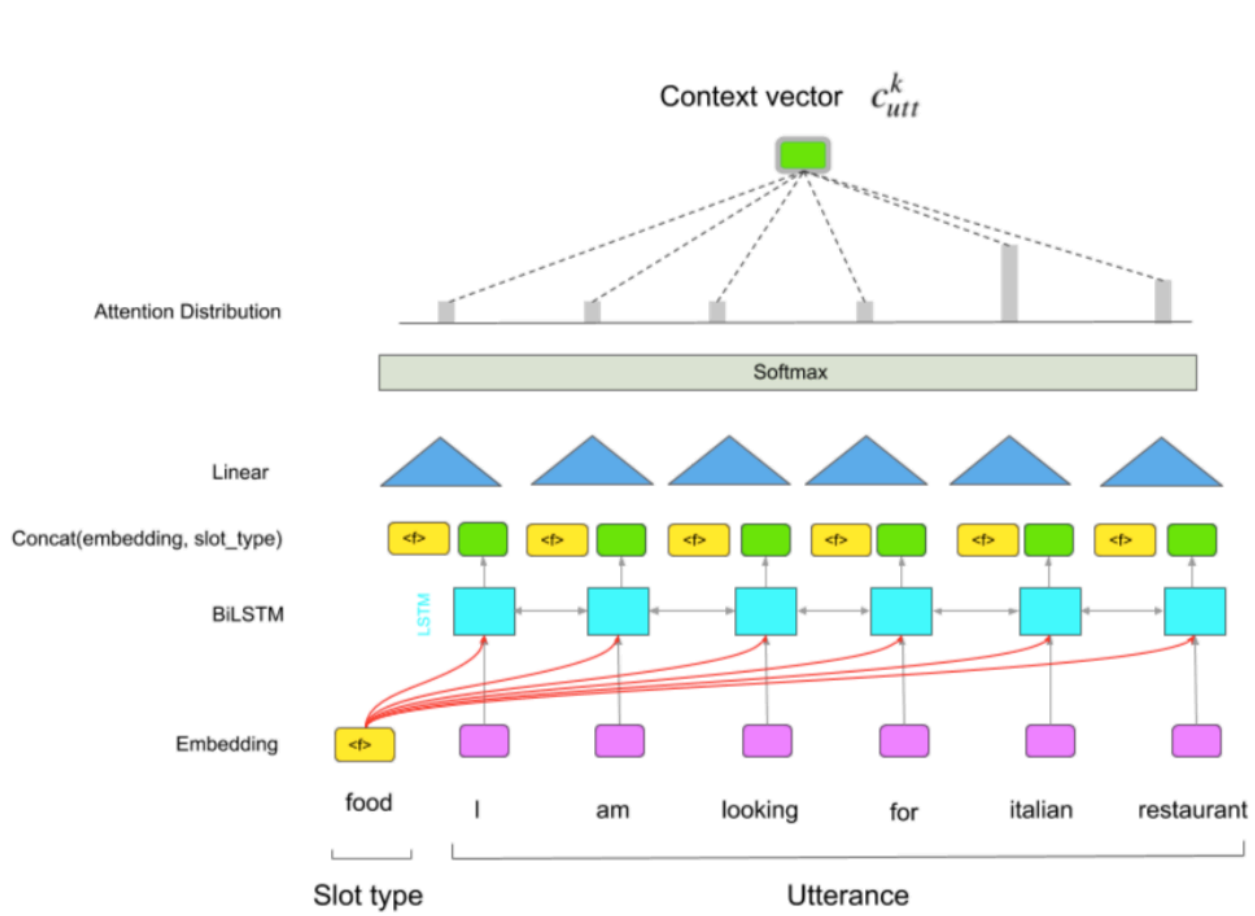
Metrics

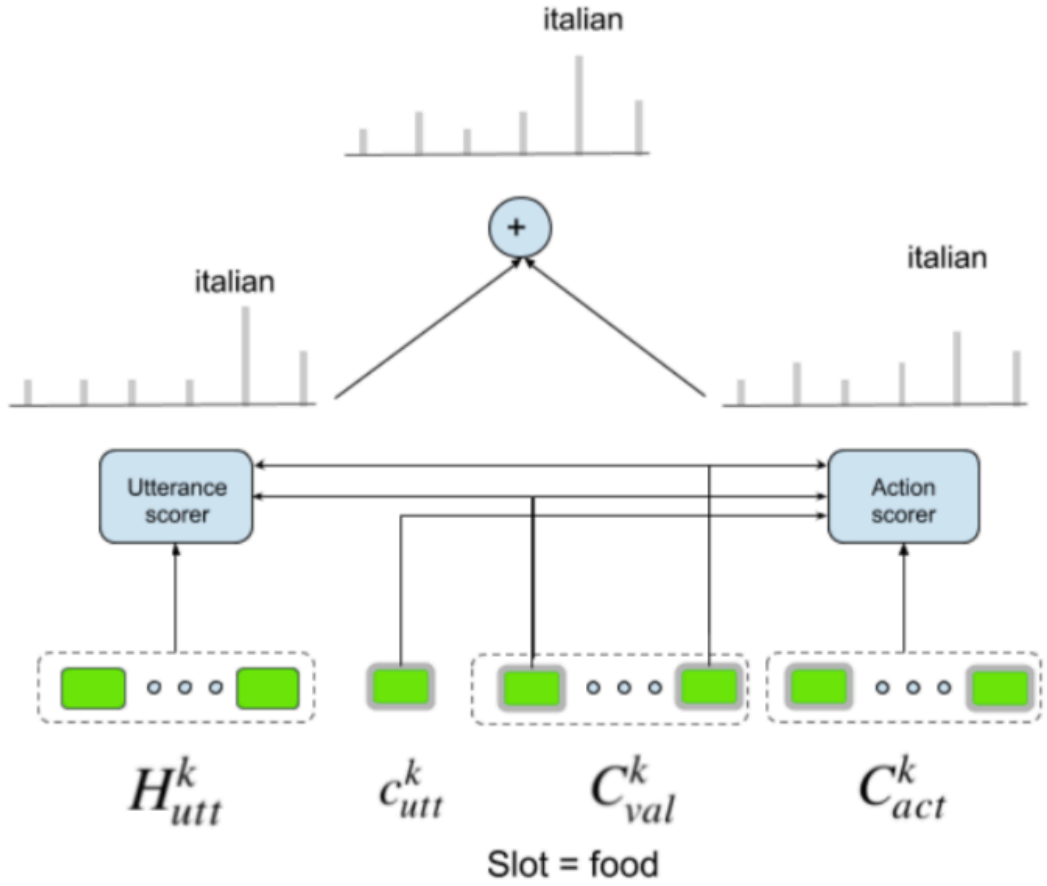
1. **Active intent accuracy** - The fraction of user turns for which the active intent has been correctly predicted.
2. **Requested slots F1** - The macro-averaged F1 score for requested slots over the turns.
3. **Average goal accuracy** - For each turn, we must predict a single value for each slot present in the dialogue state. This is the average accuracy of predicting the value of a slot correctly. A fuzzy matching based score is used for non-categorical slots.
4. **Joint goal accuracy** - This is the average accuracy of predicting all slot assignments for a turn correctly. This is the primary evaluation metric used for ranking submissions. For non-categorical slots a fuzzy matching score is used to reward partial matches with the ground truth.

The Global-Locally Self-Attentive Dialogue State Tracker

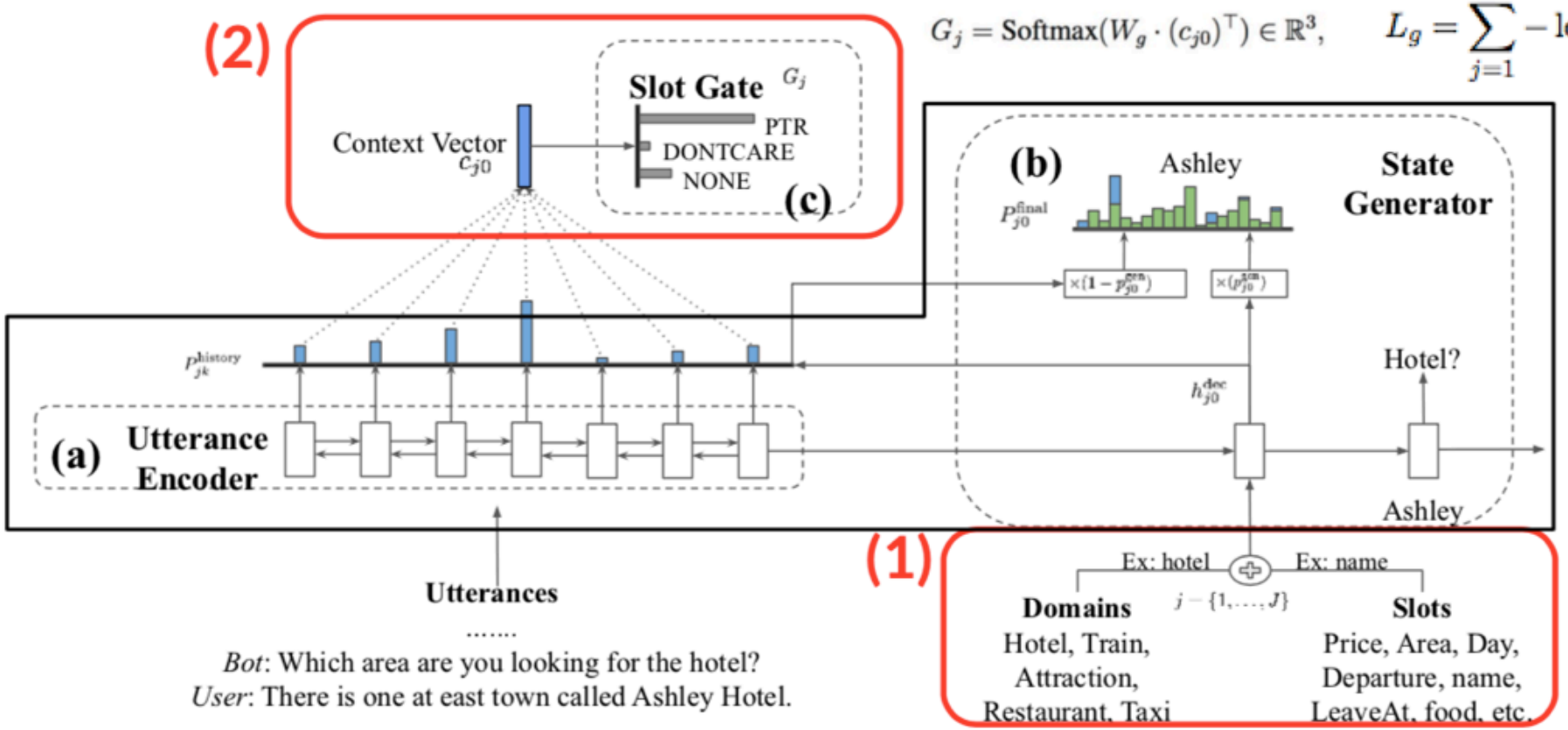


Globally-Conditioned Encoder





TRAnsferable **D**ialogue stat**E** generator



$$G_j = \text{Softmax}(W_g \cdot (c_{j0})^\top) \in \mathbb{R}^3, \qquad L_g = \sum_{j=1}^{J} -\log(G_j \cdot (y_j^{\text{gate}})^\top)$$

$$P_{jk}^{\text{vocab}} = \text{Softmax}(E \cdot (h_{jk}^{\text{dec}})^\top) \in \mathbb{R}^{|V|},$$
$$P_{jk}^{\text{history}} = \text{Softmax}(H_t \cdot (h_{jk}^{\text{dec}})^\top) \in \mathbb{R}^{|X_t|}.$$

$$P_{jk}^{\text{final}} = p_{jk}^{\text{gen}} \times P_{jk}^{\text{vocab}} + (1 - p_{jk}^{\text{gen}}) \times P_{jk}^{\text{history}} \in \mathbb{R}^{|V|}.$$

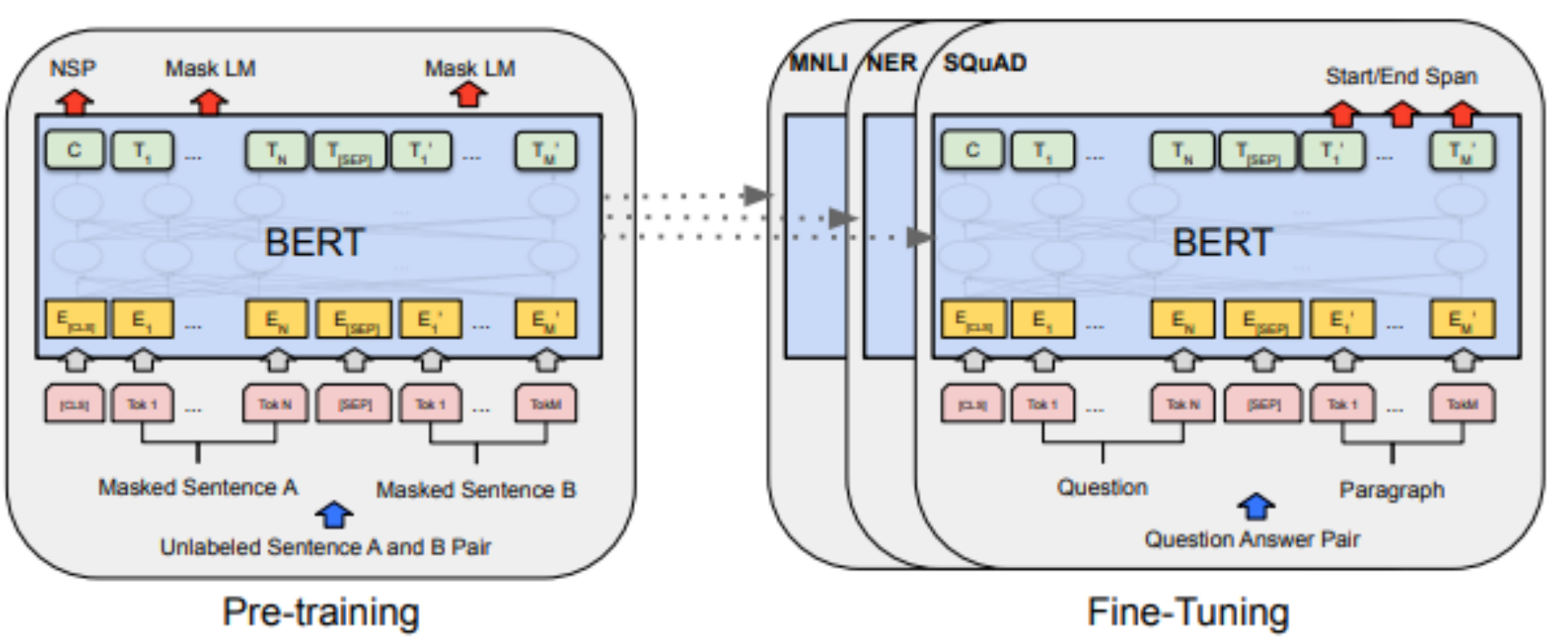
$$p_{jk}^{\text{gen}} = \text{Sigmoid}(W_1 \cdot [h_{jk}^{\text{dec}}; w_{jk}; c_{jk}]) \in \mathbb{R}^1$$
$$c_{jk} = P_{jk}^{\text{history}} \cdot H_t \in \mathbb{R}^{d_{hdd}}$$

$$L_v = \sum_{j=1}^{J} \sum_{k=1}^{|Y_j|} -\log(P_{jk}^{\text{final}} \cdot (y_{jk}^{\text{value}})^\top).$$

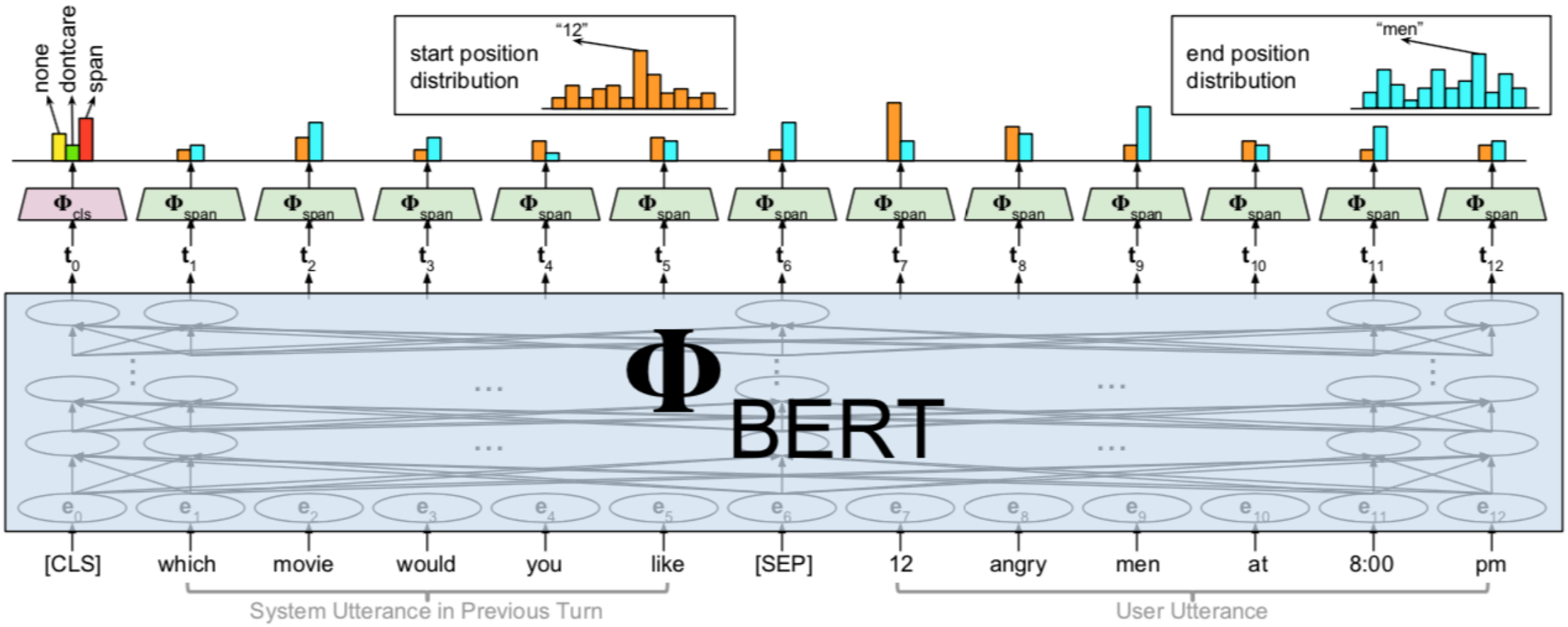
DeepPavlov.ai

MIPT MOSCOW INSTITUTE OF PHYSICS AND TECHNOLOGY

BERT


NSP
Mask LM
Mask LM
BERT
Masked Sentence A
Masked Sentence B
Unlabeled Sentence A and B Pair
Pre-training
MNLI
NER
SQuAD
Start/End Span
BERT
Question
Paragraph
Question Answer Pair
Fine-Tuning

BERT Dialogue State Tracker



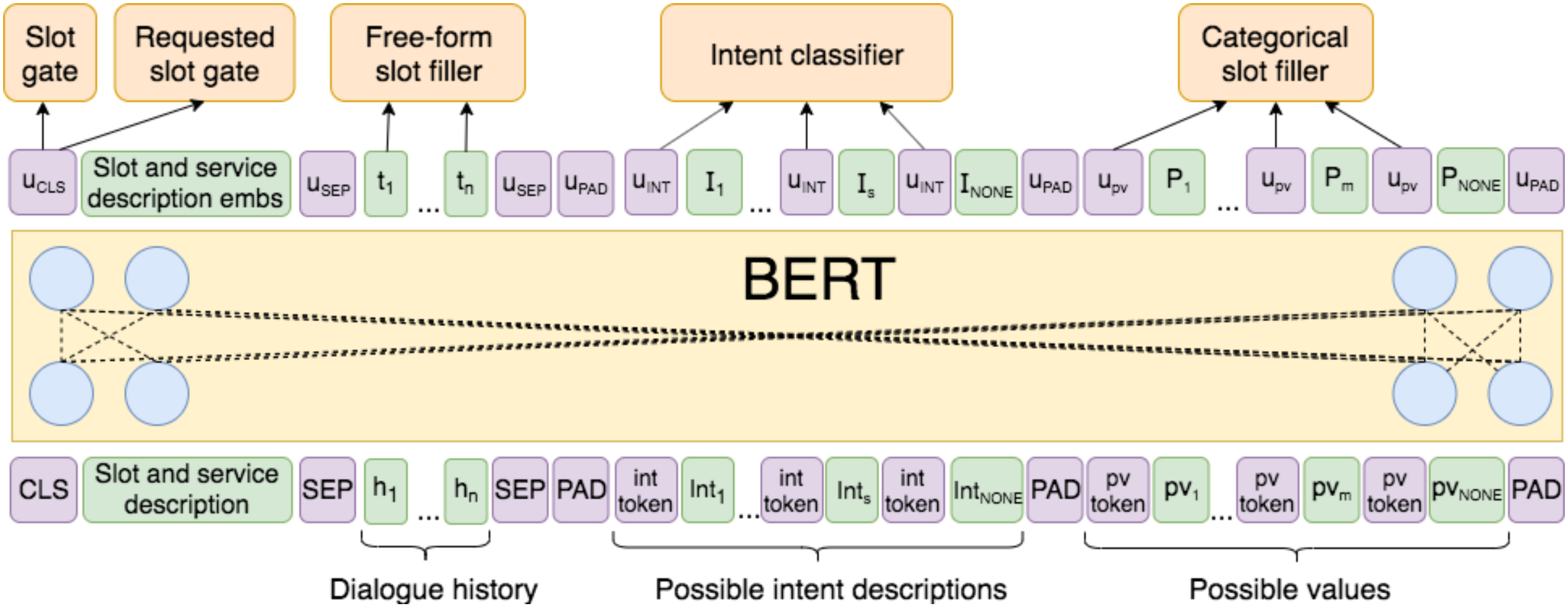
Question Answering on SQuAD

A: Vice President

Q: Who is Mike Pence?

The U.S. is ready to engage in talks about North Korea's nuclear program even as it maintains pressure on Kim Jong Un's regime, the Washington Post reported, citing an interview with **Vice President** [A] Mike Pence. Pence and South Korea's President Moon Jae-in agreed on a post-Olympics strategy during conversations at the Winter Olympics in the South Korean resort of Pyeongchang that Pence dubbed "maximum pressure and engagement at the same time." Pence spoke in an interview on his way home from the Winter Olympics. "The point is, no pressure comes off until they are actually doing something that the alliance believes represents a meaningful step toward denuclearization," the Post quoted Pence as saying. "So the maximum pressure campaign is going to continue and intensify. But if you want to talk, we'll talk."

Our model architecture



Results

	Active Intent Accuracy	Requested Slot F1	Average Goal Accuracy	Joint Goal Accuracy
GOLOMB, dev scores	0.660	0.969	**0.817**	**0.539**
Baseline, dev scores	**0.908**	**0.973**	0.740	0.411
GOLOMB, test scores	0.747	0.971	0.750	0.465

Thank you for your attention!