https://github.com/jasonwu0731/trade-dst

TRADE: Transferable Multi-Domain State Generator for Task-Oriented Dialogue Systems

Chien-Sheng (Jason) Wu, Andrea Madotto, Ehsan Hosseini-Asl, Caiming Xiong, Richard Socher and Pascale Fung

2019.07.29

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

salesforce

Short Conclusion

“

TRADE is a simple ***copy-augmented*** *generative model that can track dialogue states* ***without requiring ontology****. It is the current SOTA model in multi-domain DST. It also enables zero-shot and few-shot DST in an unseen domain.*

Dialogue Systems: Chit-Chat v.s. Task-Oriented

Chit-Chat Dialogue Systems

- No Specific goal
- Focus on generating natural responses
- The more turns the better

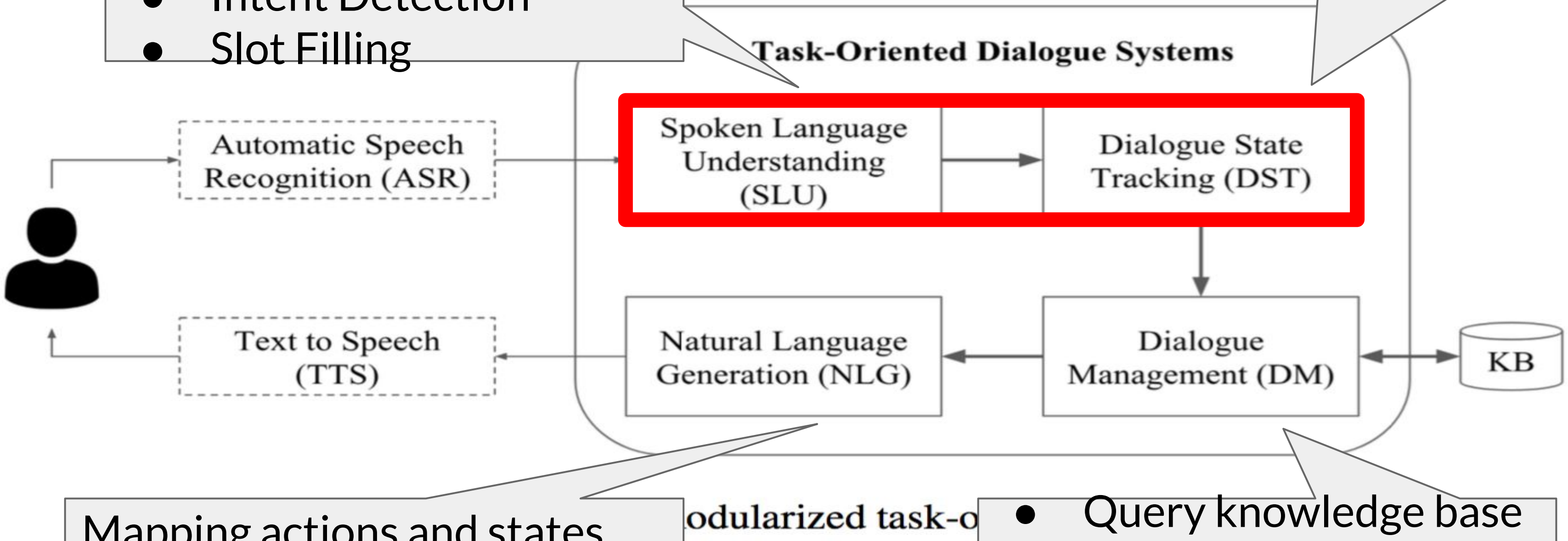
Task-Oriented Dialogue Systems

- Help users achieve their goal
- Focus on understanding users, tracking states, and generating next actions.
- The less turns the better

Modularized Dialogue Systems

- Domain Identification
- Intent Detection
- Slot Filling

Represent system's belief of user's goal as slot-value pairs

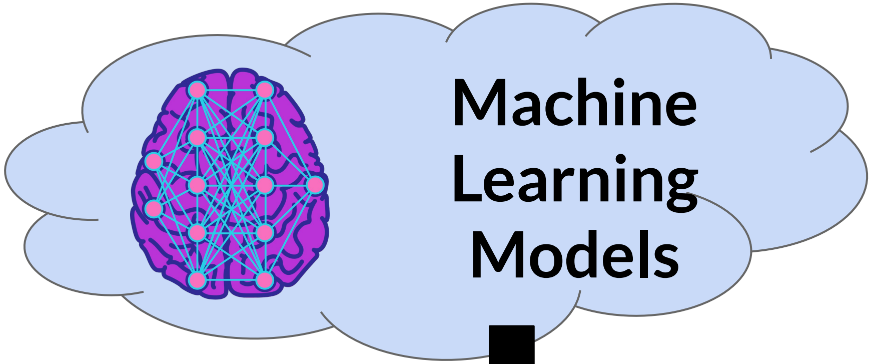


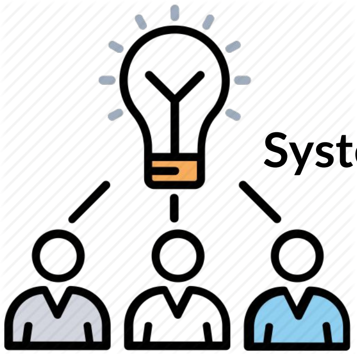
Mapping actions and states into natural language.

- Query knowledge base
- Dialogue policy (generate next action)

Problem: Multi-Domain Dialogue State Tracking (DST)

Multi-domain, Multi-turn Conversations





Dialogue States

System belief of a user's goal or Intention.

Ex: Restaurant:
{Location: Florence,
Price: Expensive, …}

A Dialogue Example

I'm looking for a **cheap pizza** restaurant in the city **center**.

Sure. There is a **Dominos** nearby. How many?

Four people **Tuesday** at **1 pm** please. Please make sure there is **NO PINEAPPLE** on the pizza!

Booked! ABCDE is your reservation code.

Also looking for some architectural place close to restaurant.

Florence Cathedral is famous. Would you like to head there?

Yes help me book a taxi between restaurant and church.

What time do you need the taxi?

Around **3pm** please.

You are all set. Have a good trip.

Restaurant: {*Price*: Cheap, *Type*: Pizza, *Area*: Center}

Restaurant: {*People*: Four, *Day*: Tuesday, *Time*: 1pm, *Others*: No pineapple}

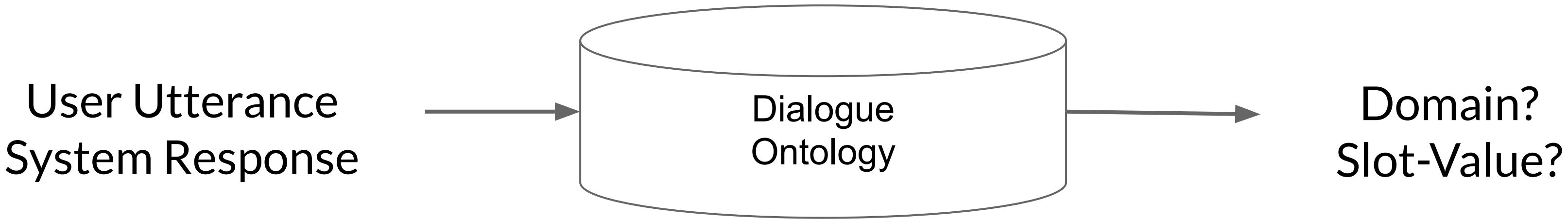
Attraction: {*Area*: *Center*, *Type*: *Architectural*}

Taxi: {*Destination*: *Florence Cathedral* , *Departure*: *Dominos*}

Taxi: {*LeaveAt*: *3pm*}

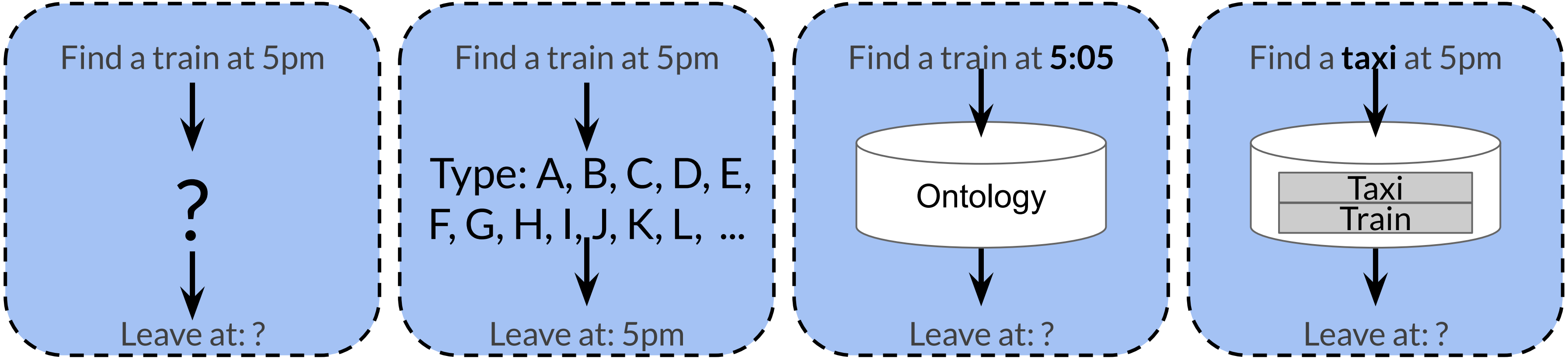
Ontology-based DST

- Given system response and current user utterance, each slot in each domain is predicted to be one of the **predefined values** in **ontology**.
- Models: **ScaleDST** (Rastogi et al., 2017); **MDBT** (Ramadan et al., 2018); **GLAD** (Zhong et al., 2018); **GCE** (Nouri et al., 2018)

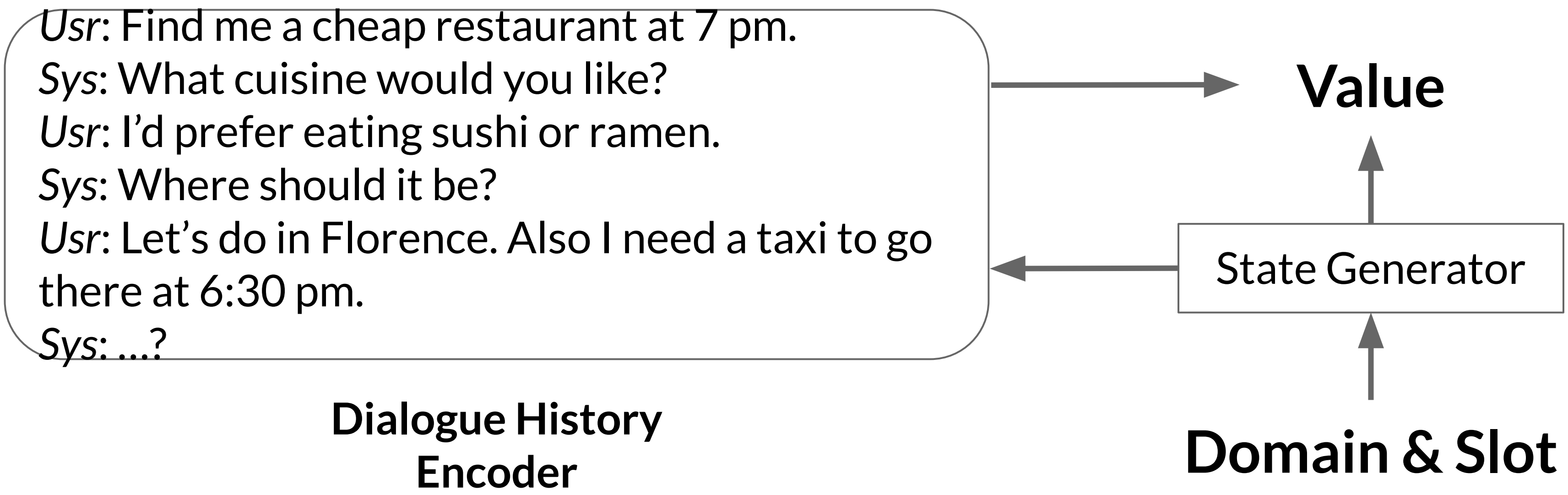


Challenges of Ontology-based DST

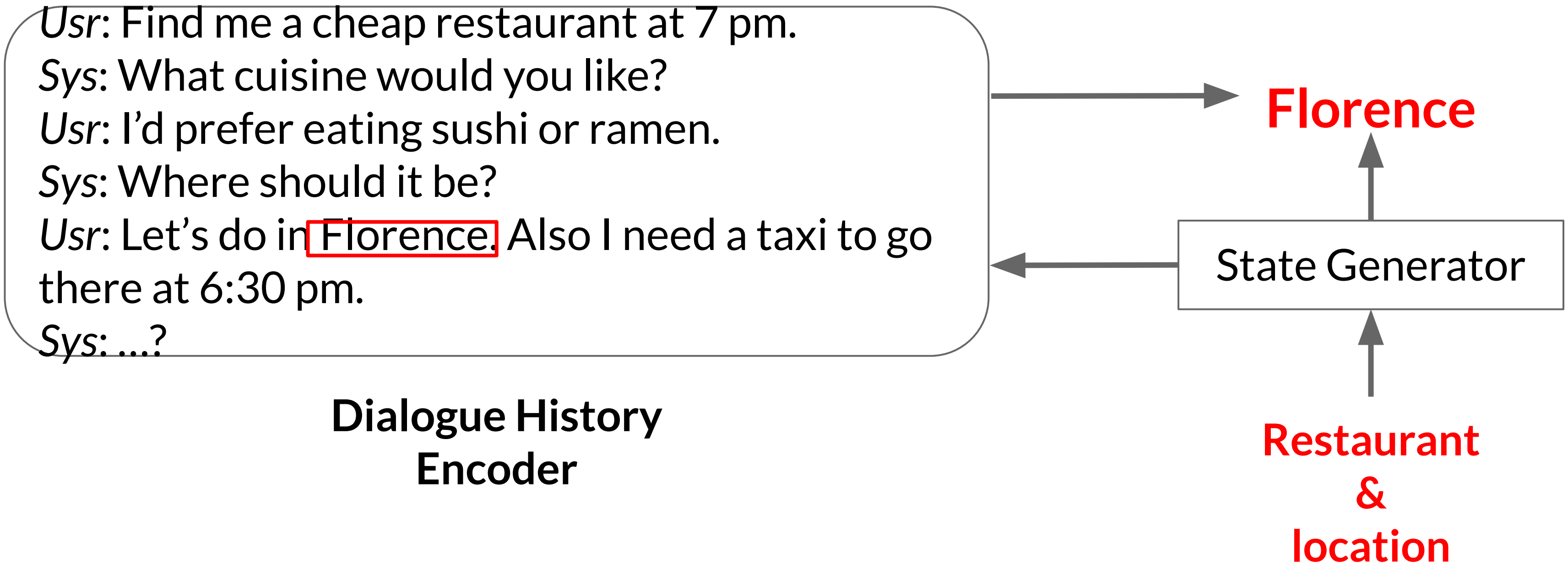
- Ontology is hard to obtain in real scenarios
- Need to track lots of slot values
- Cannot track unseen slot values
- Missing domain sharing capacities



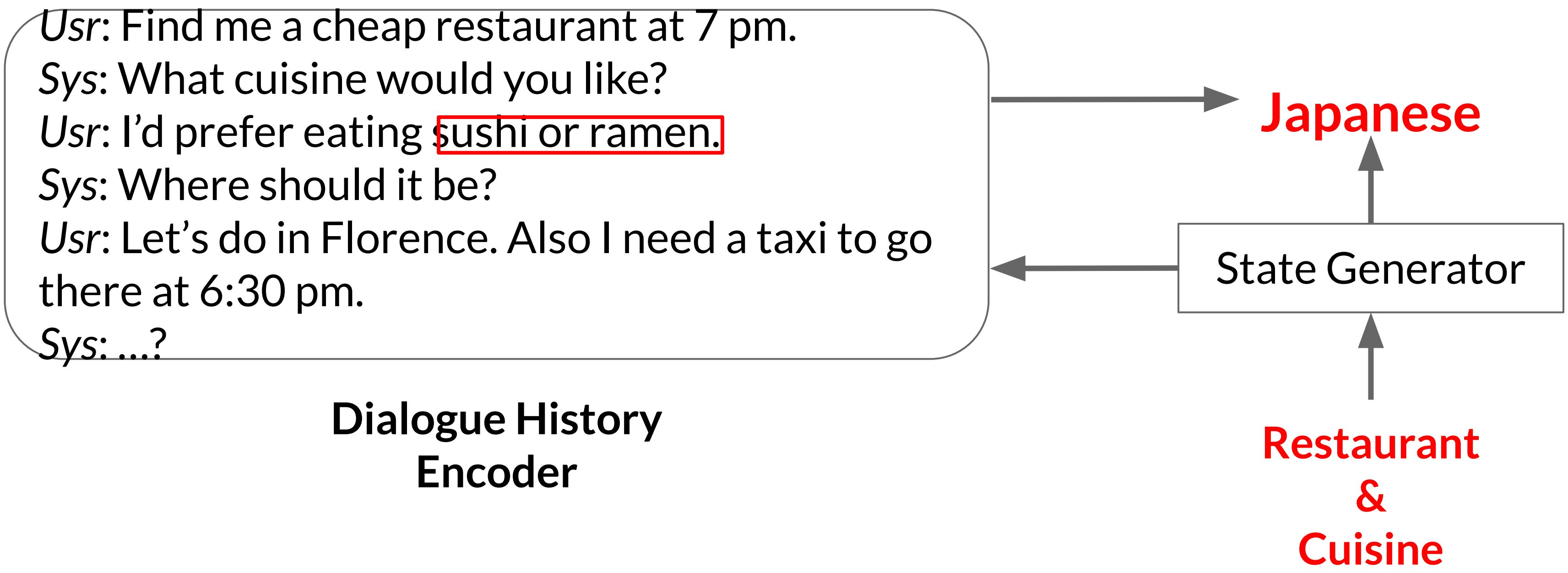
Ontology-free DST: Intuition



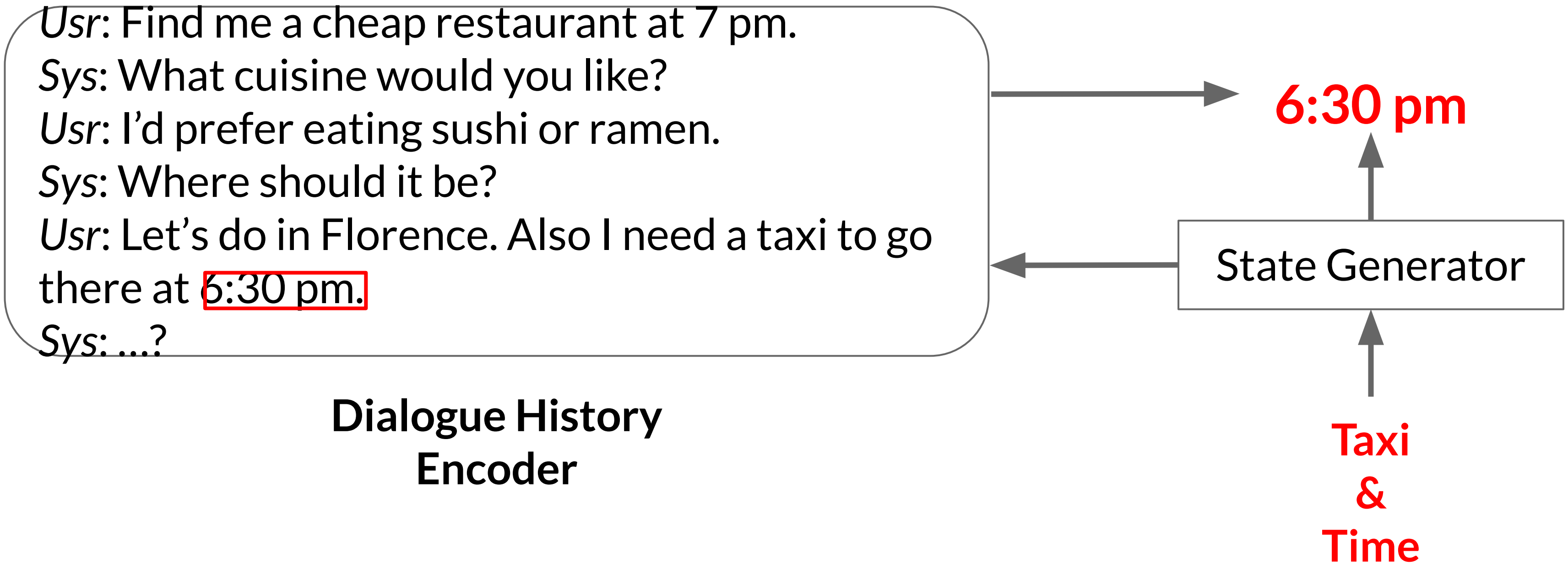
Ontology-free DST: Intuition



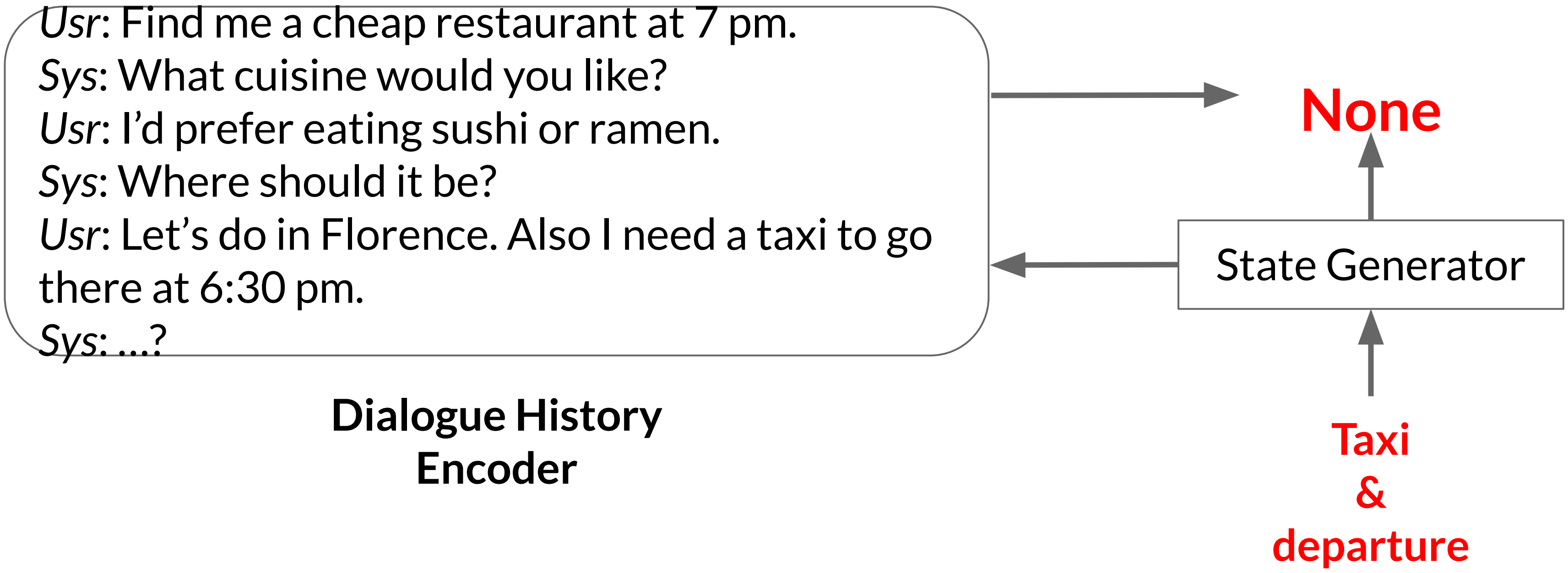
Ontology-free DST: Intuition



Ontology-free DST: Intuition

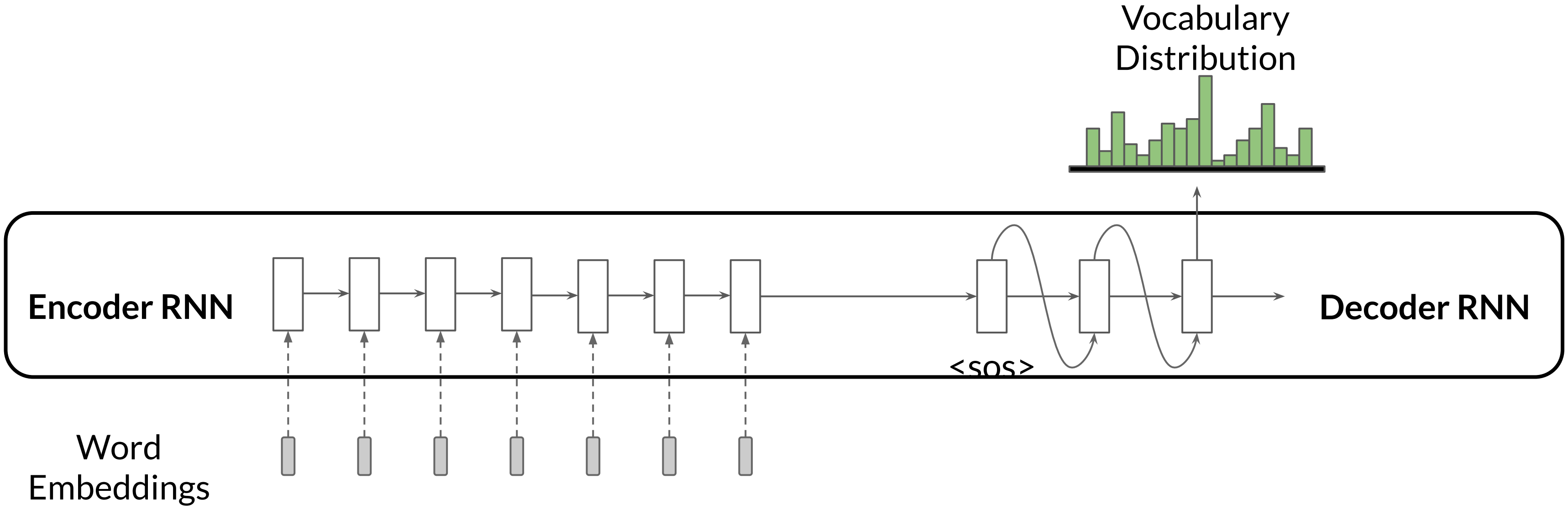


Ontology-free DST: Intuition

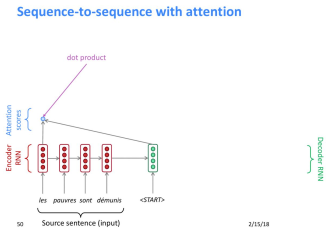


Sequence-to-Sequence (Seq2Seq)

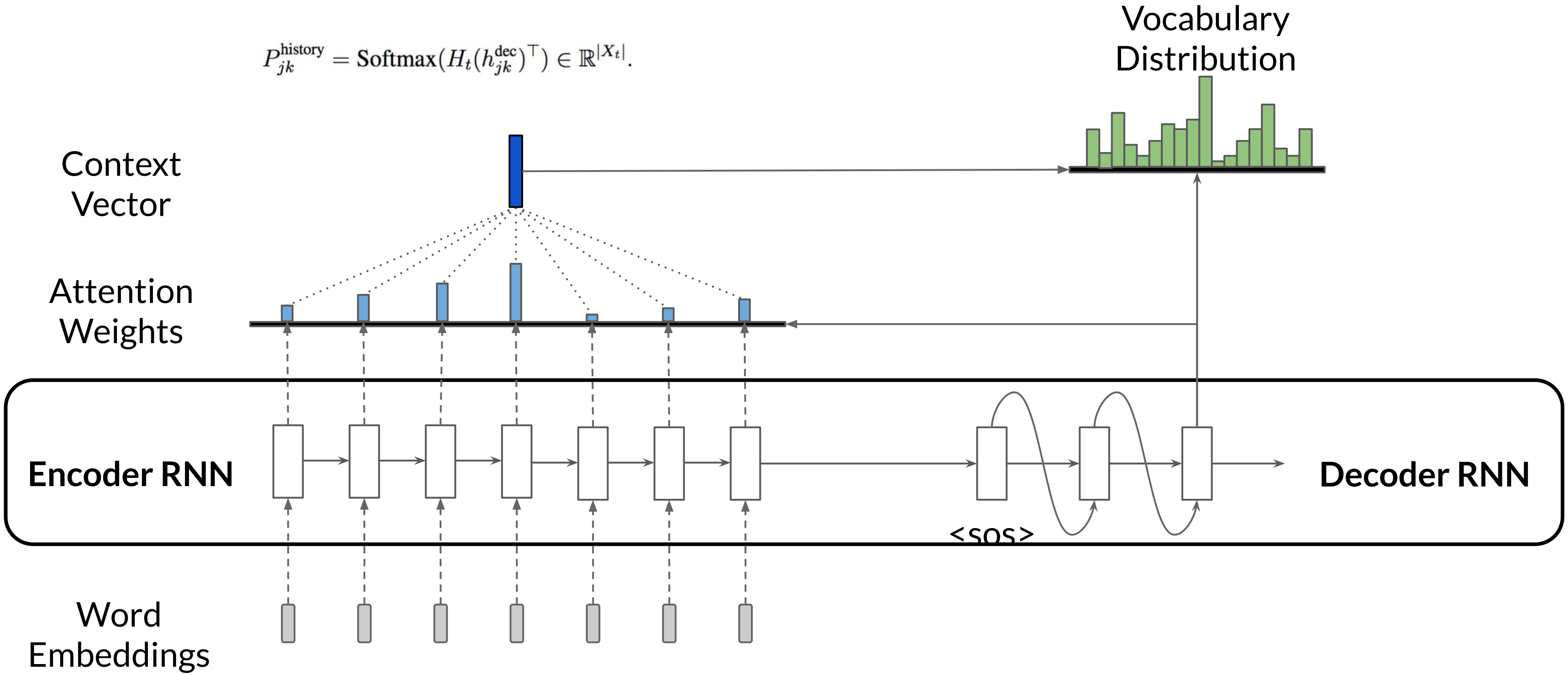
$$P_{jk}^{\text{vocab}} = \text{Softmax}(E(h_{jk}^{\text{dec}})^\top) \in \mathbb{R}^{|V|},$$



Seq2Seq with Attention

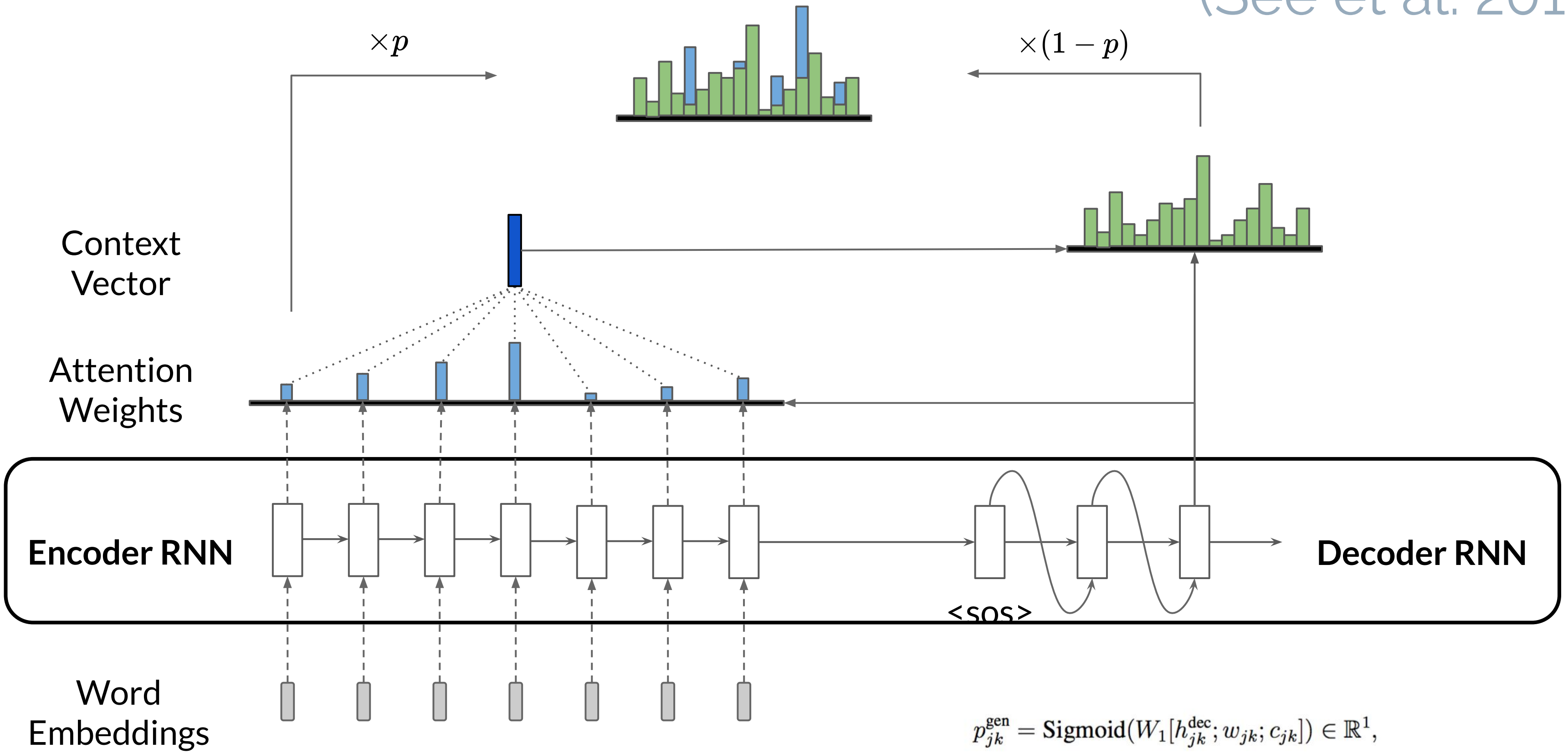


$$P_{jk}^{\text{history}} = \text{Softmax}(H_t (h_{jk}^{\text{dec}})^\top) \in \mathbb{R}^{|X_t|}.$$



Seq2Seq with (Soft) Copy Mechanism

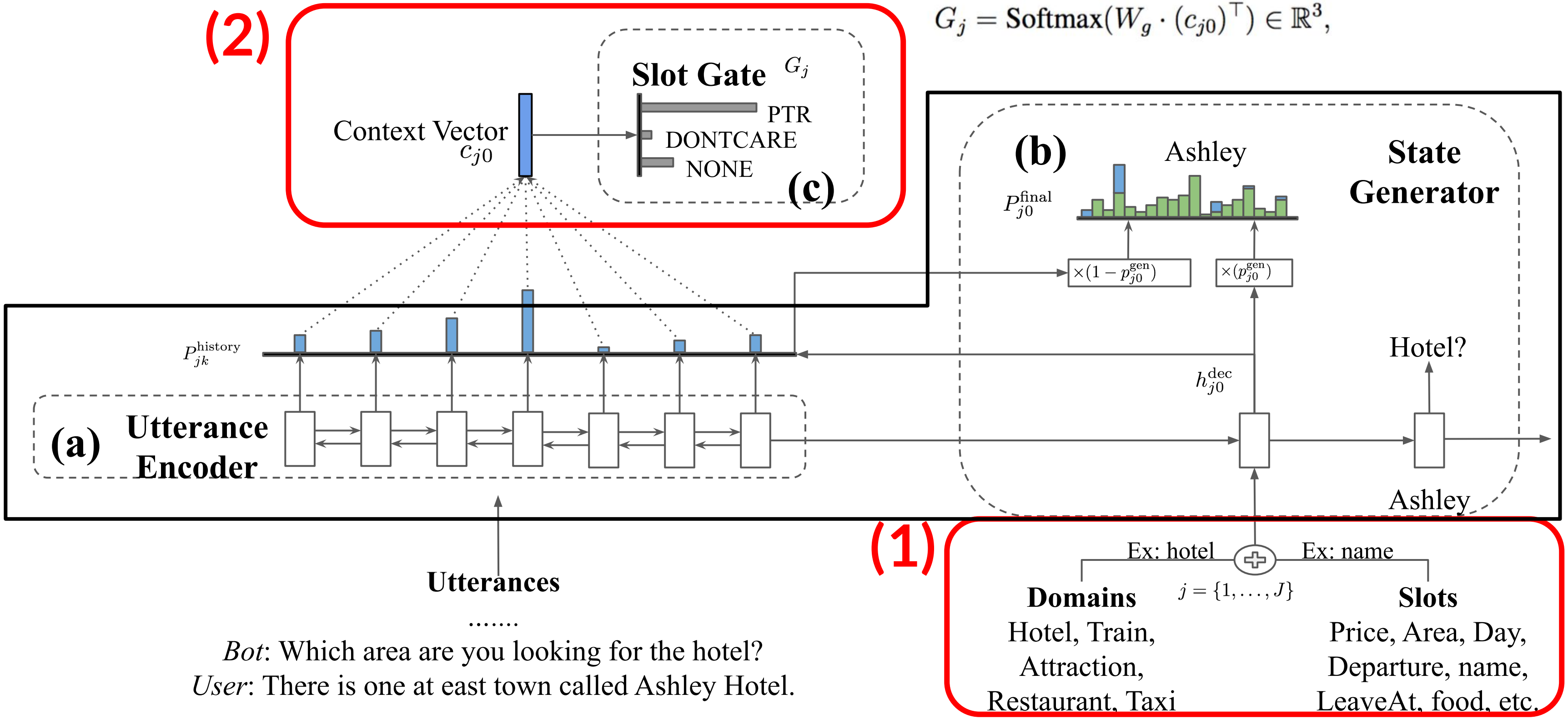
(See et al. 2017)



$$p_{jk}^{\text{gen}} = \text{Sigmoid}(W_1[h_{jk}^{\text{dec}}; w_{jk}; c_{jk}]) \in \mathbb{R}^1,$$

TRADE: Transferable Dialogue State Generator

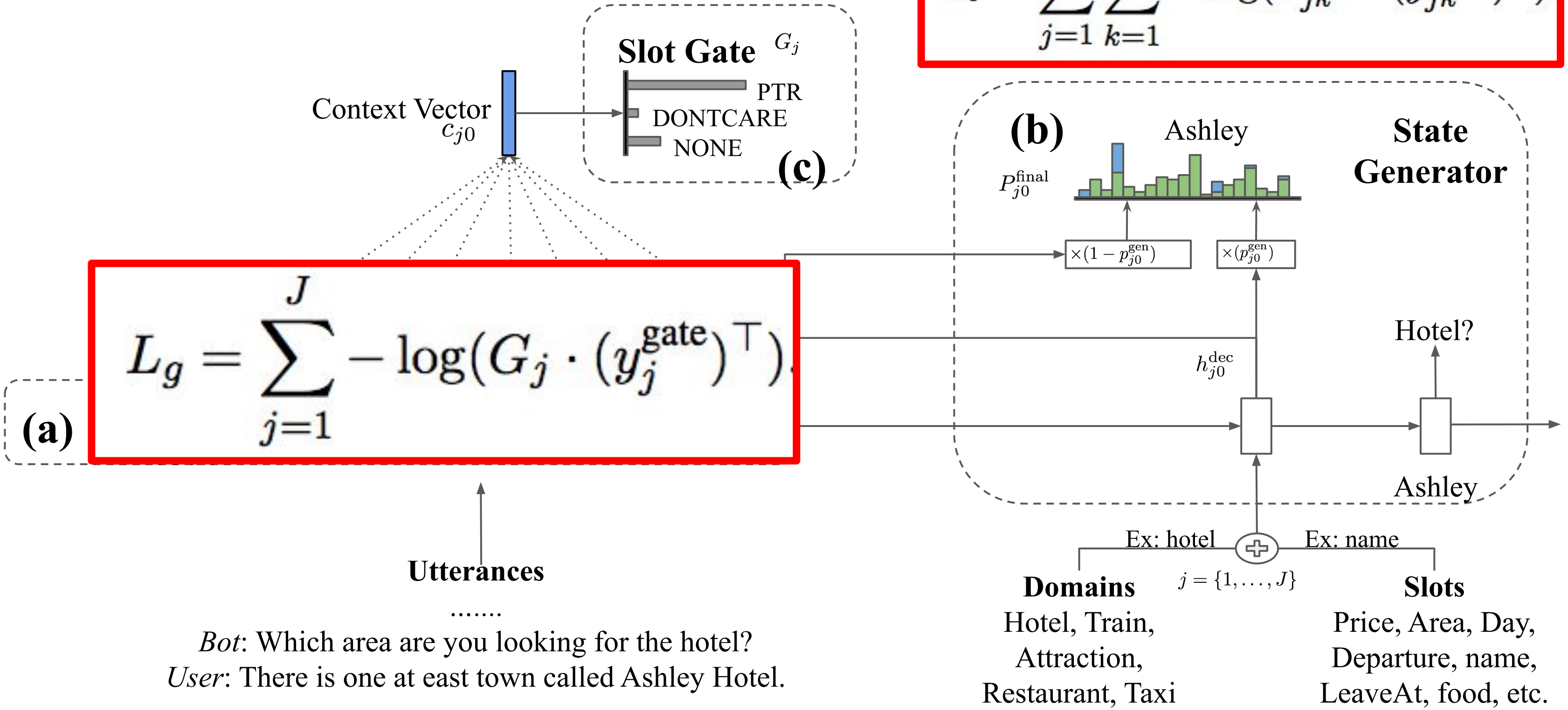
$$G_j = \text{Softmax}(W_g \cdot (c_{j0})^\top) \in \mathbb{R}^3,$$



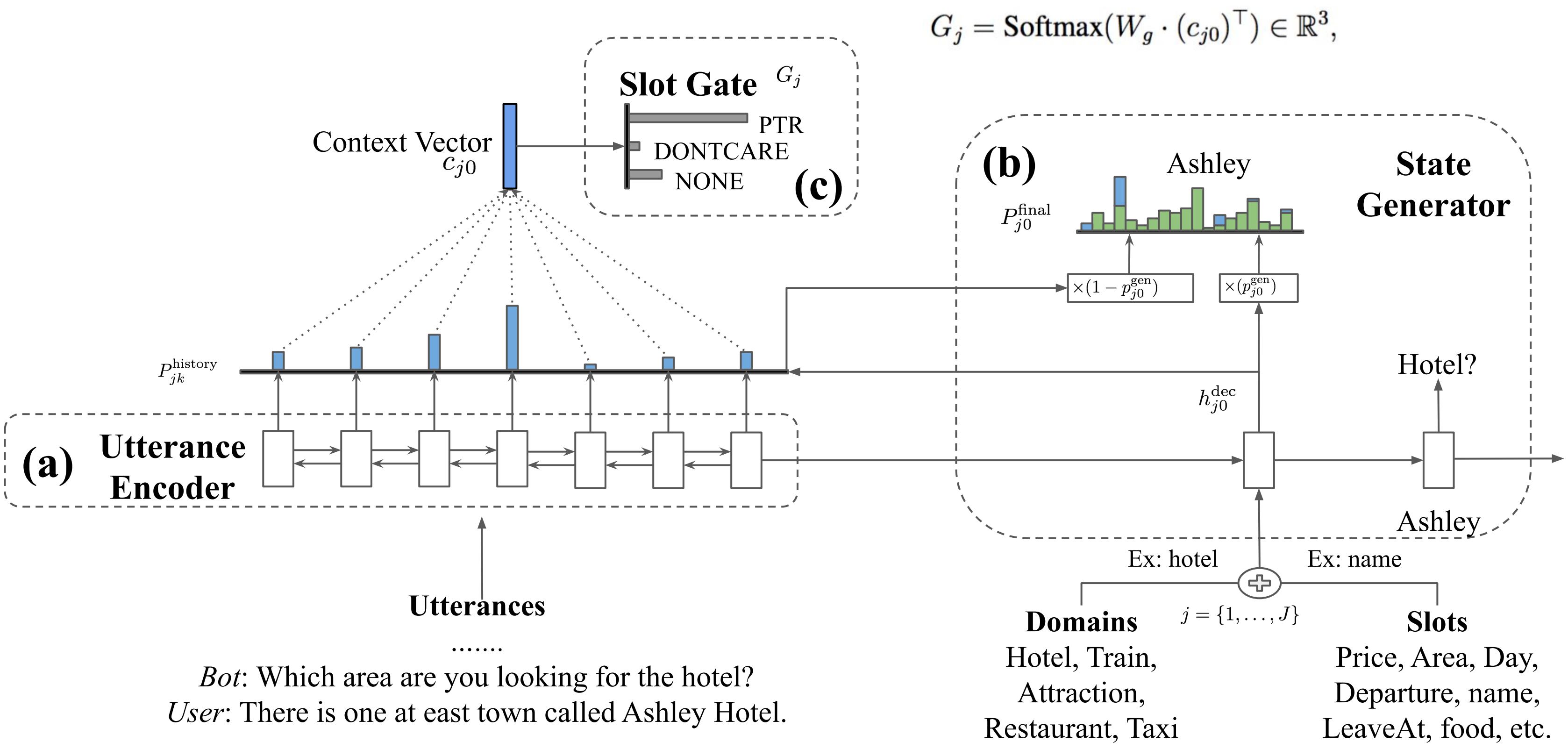
TRADE: Transferable Dialogue State Generator

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

$$L_g = \sum_{j=1}^{J} - \log(G_j \cdot (y_j^{\text{gate}})^\top).$$



TRADE: Transferable Dialogue State Generator



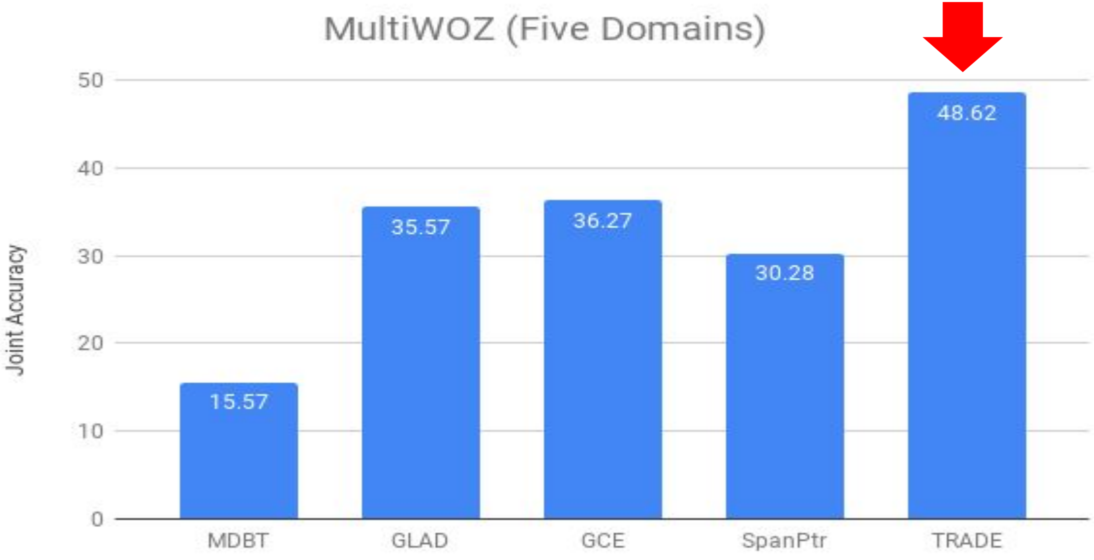
MultiWOZ Dataset (Budzianowski et al., 2018)

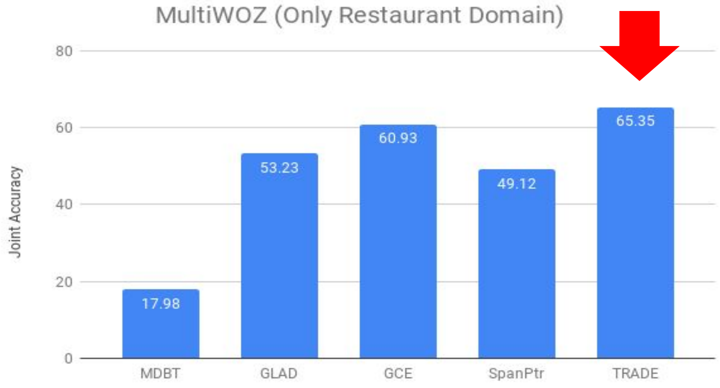
- The largest available human-human conversational corpus with DST labels (8438 dialogues with avg 13.68 turns).
- 5 domains (Hotel, Train, Attraction, Restaurant, Taxi) and 16 slots (food, leave at, area, etc).
- Total 30 domain-slot pairs and ~4500 slot values.

	Hotel	Train	Attraction	Restaurant	Taxi
Slots	price, type, parking, stay, day, people, area, stars, internet, name	destination, departure, day, arrive by, leave at, people	area, name, type	food, price, area, name, time, day, people	destination, departure, arrive by, leave by
Train	3381	3103	2717	3813	1654
Valid	416	484	401	438	207
Test	394	494	395	437	195

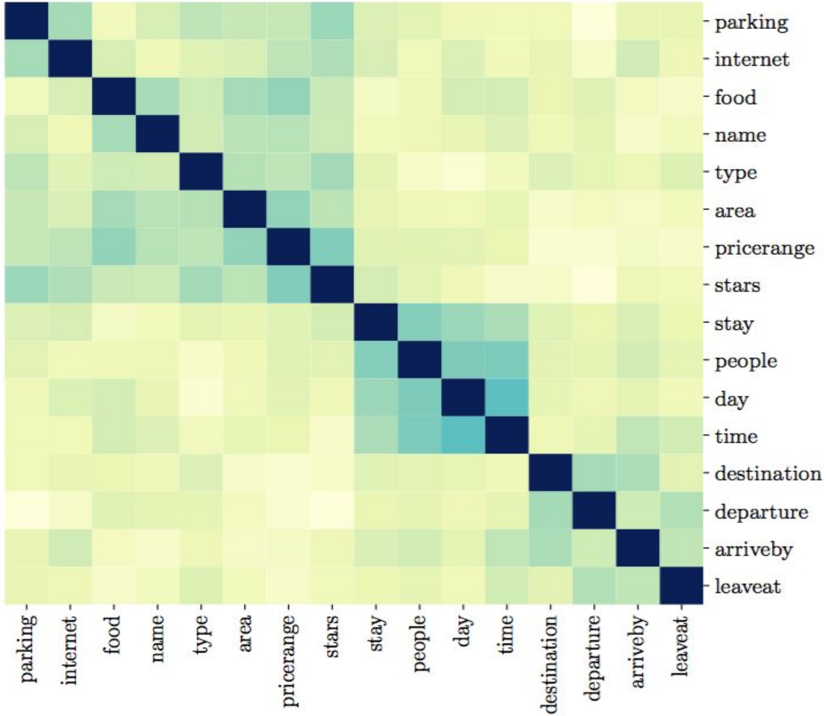
Multi-Domain Joint Training

MDBT (Ramadan et al., 2018)
GLAD (Zhong et al., 2018)
GCE (Nouri et al., 2018)
SpanPtr (Xu et al., 2018)





Multi-Domain Joint Training: Visualization

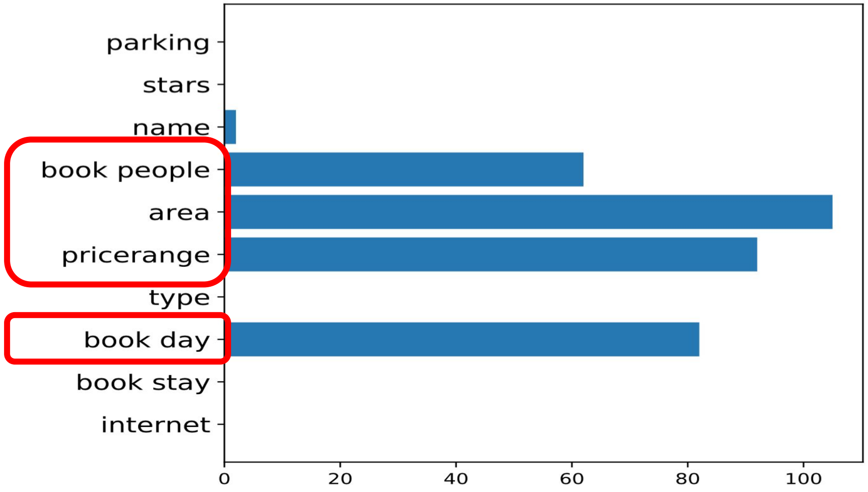


Zero-Shot Domain DST

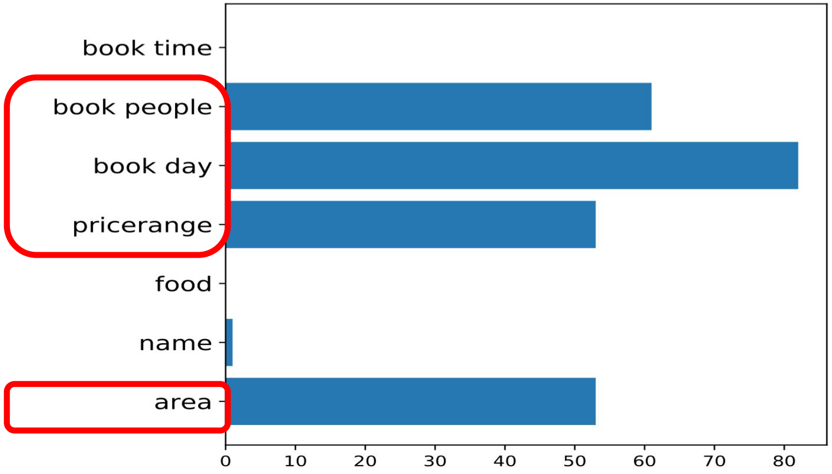
	Trained Single		Zero-Shot	
	Joint	*Slot*	*Joint*	*Slot*
Hotel	55.52	92.66	13.70	65.32
Train	77.71	95.30	22.37	49.31
Attraction	71.64	88.97	19.87	55.53
Restaurant	65.35	93.28	11.52	53.43
Taxi	76.13	89.53	**60.58**	73.92

Table 3: Zero-shot experiments on an unseen domain. In *taxi* domain, our model achieves 60.58% joint goal accuracy without training on any samples from *taxi* domain. *Trained Single* column is the results achieved by training on 100% single-domain data as a reference.

Unseen Domain Testing (Zero-Shot): Correctness Analysis



Hotel



Restaurant

Few-Shot Domain Expansion DST: (1% unseen domain data)

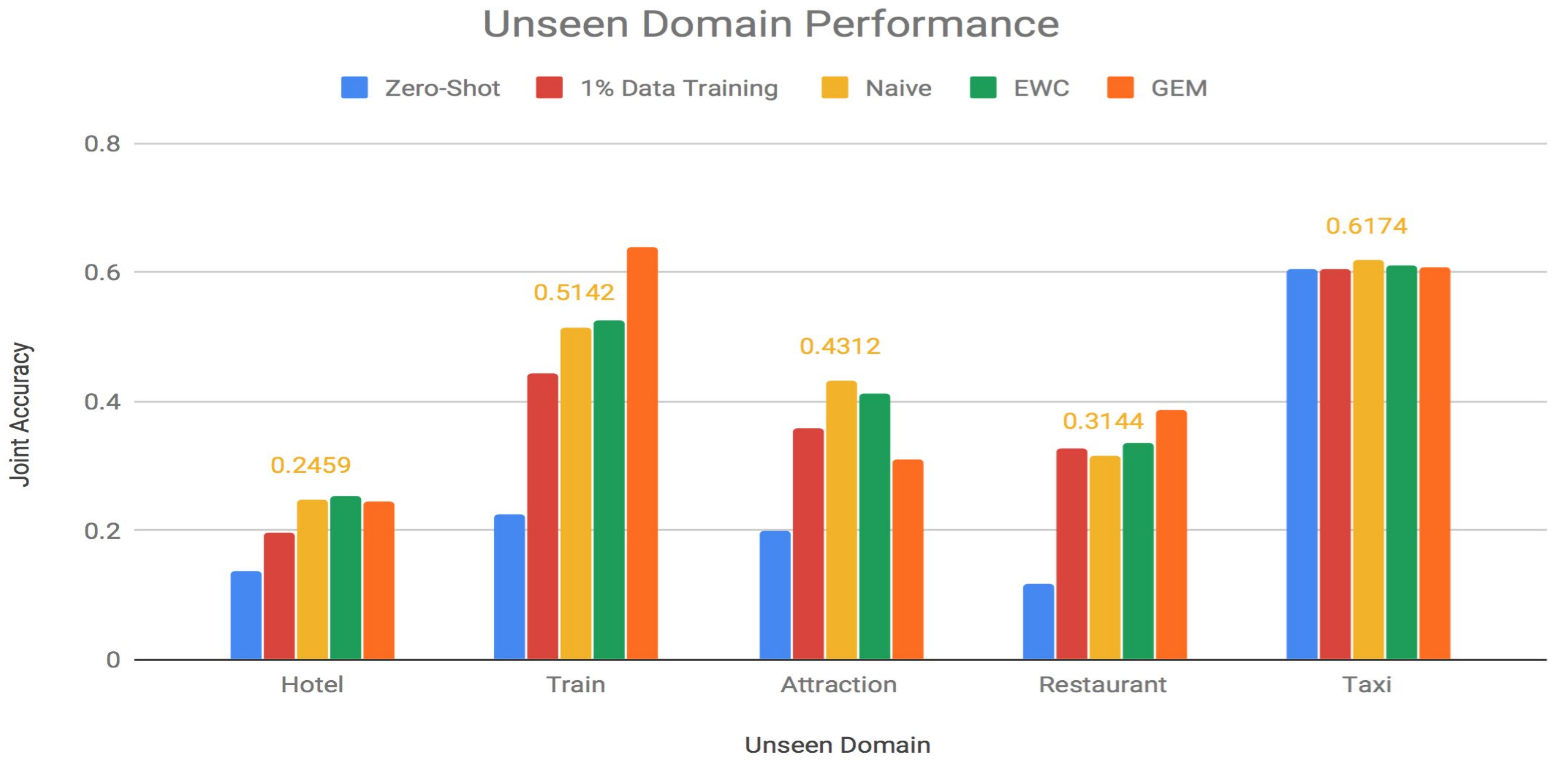
- Why?
  - Be able to quickly adapt to new domains.
  - Not require retraining with all the data from previously learned domains (not available and time-consuming).
- How?
  - **Naive** fine-tuning; **EWC** (Kirkpatrick et al., 2017); **GEM** (Lopez-Paz et al., 2017).
- What?
  - Unseen domain performance
  - Trained domains performance

$$L_{ewc}(\Theta) = L(\Theta) + \sum_i \frac{\lambda}{2} F_i(\Theta_i - \Theta_{S,i})^2$$

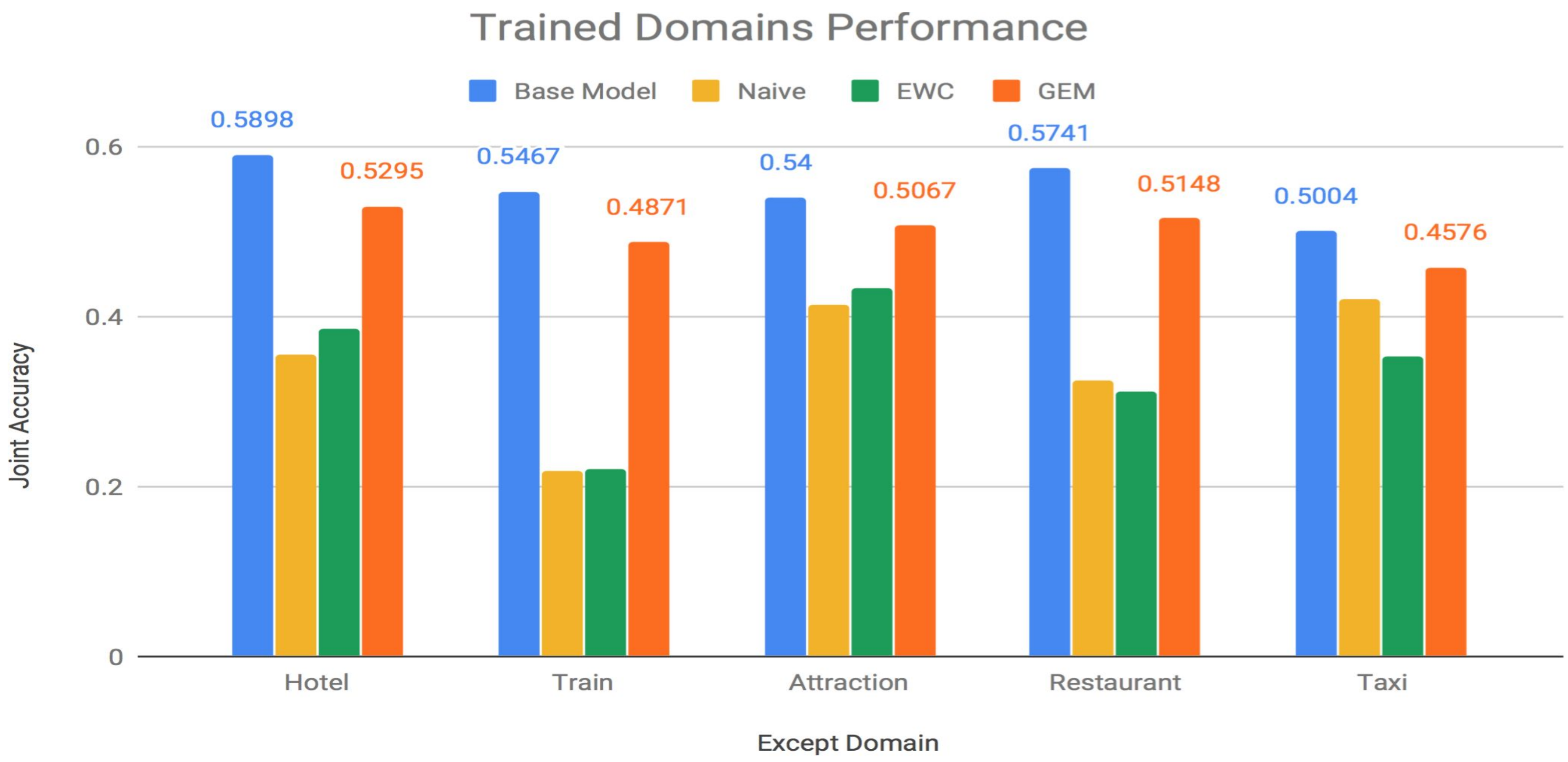
$$\text{Minimize}_\Theta \ L(\Theta)$$
$$\text{Subject to } L(\Theta, K) \leq L(\Theta_S, K),$$

Unseen Domain Performance (Few-Shot)

Unseen Domain Performance (Few-Shot)



Trained Domains Performance (Few-Shot)



MultiWOZ 2.1 (Eric et al., 2019)

- A correction version of original MultiWOZ dataset, resulting in changes to 32% of state annotations across 40% of the dialogue turns.

Model	MultiWOZ 2.0	MultiWOZ 2.1
FJST	40.2%	38.0%
HJST	38.4%	35.55%
TRADE	**48.6%**	**45.6%**
DST Reader	39.41%	36.4%
HyST	42.33%	38.1%

Type	Conversation	MultiWOZ 2.0	MultiWOZ 2.1
Delayed Markups	User: I'd also like to try a Turkish restaurant. Is that possible? Agent: I'm sorry but the only restaurants in that part of town serve either Asian food or African food.	restaurant.food: None	restaurant.food: Turkish
		nt.food: Turkish	restaurant.food:Turkish
		me: The Cambridge Belfry n.name: belf	hotel.name: The Cambridge Belfry attraction.name: None
		veAt: Thursday : Not Mentioned	train.leaveAt: None train.day: Thursday
		n.area: cent	attraction.area: Centre
		nt.pricerange: None	restaurant.pricerange: Dontcare
	again: Cambridge to Bishop Stafford on Thursday.	train.destination: Bishop Stortford	train.destination: Bishops Stortford

Table 5: Examples of annotation errors between MultiWOZ 2.0 and 2.1

Thank you!
Any Questions?

Chien-Sheng (Jason) Wu

Andrea Madotto

Ehsan Hosseini-Asl

Caiming Xiong

Richard Socher

Pascale Fung

Code: **https://github.com/jasonwu0731/trade-dst**