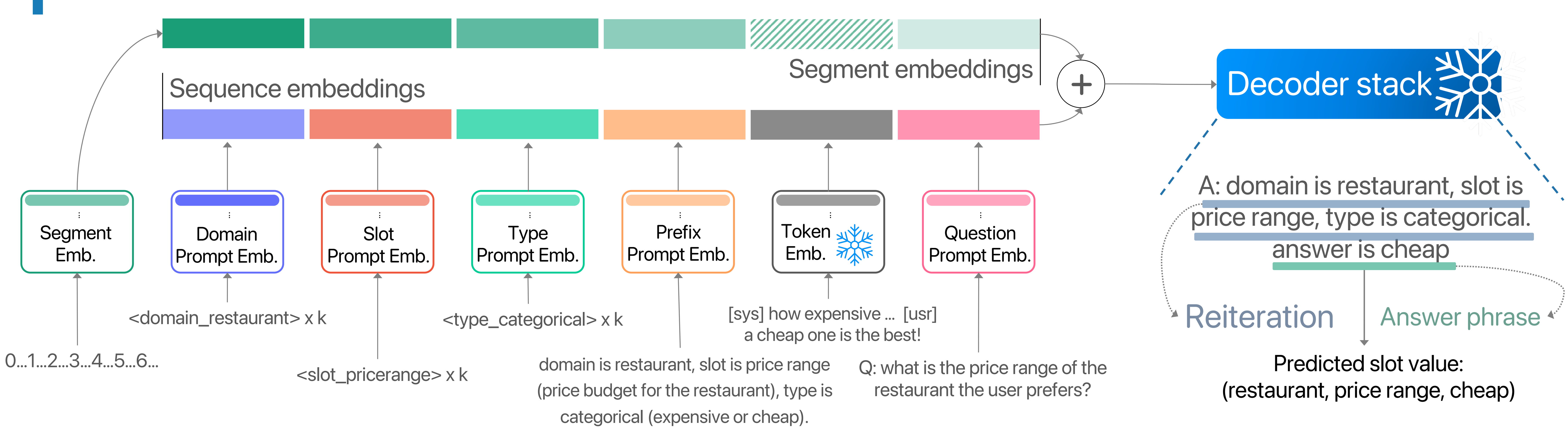
Parameter-Efficient Low-Resource Dialogue State Tracking by Prompt Tuning

Motivation

- Dialogue state tracking extracts structured conversation progress in a list of (slot, value) pairs from unstructured dialogue utterances
- Existing works formulate DST as a **conditional generation task with prompts** to provide information about slot name, slot description, slot type, possible values, priming examples, questions etc.
- Challenges
  - Existing works all **fine-tune the entire LM** along with the prompt token embeddings
  - Real-world deployment needs to train and host separate LMs for different domains and tasks
  - Limited data is available for new domains or tasks
- We propose a **parameter-efficient** and **data-efficient** DST model for **low-resource** settings, which only needs to update 0.5% of parameters compared with baselines while achieving state-of-the-art performance

Method



1 Task-specific parameters

Domain
Slot
Type

Task prompt tokens: Shared across instances of the same task, represent domain, slot, and type information

Prefix
Question

Word-mapping prompt tokens: Obtain task knowledge contained in natural language instruction and optimize human-created prompts with continuous embedding space; shared across instances with the same words

2 Task metadata in objective

Reiterate the querying task metadata before generating the answer; explicit task information as a part of the objective

3 Distinguishing segments

Use randomly initialized segment embedding to distinguish segments with diverse formats (prompt segments, answer segment, system turns and user turns in dialogue history)

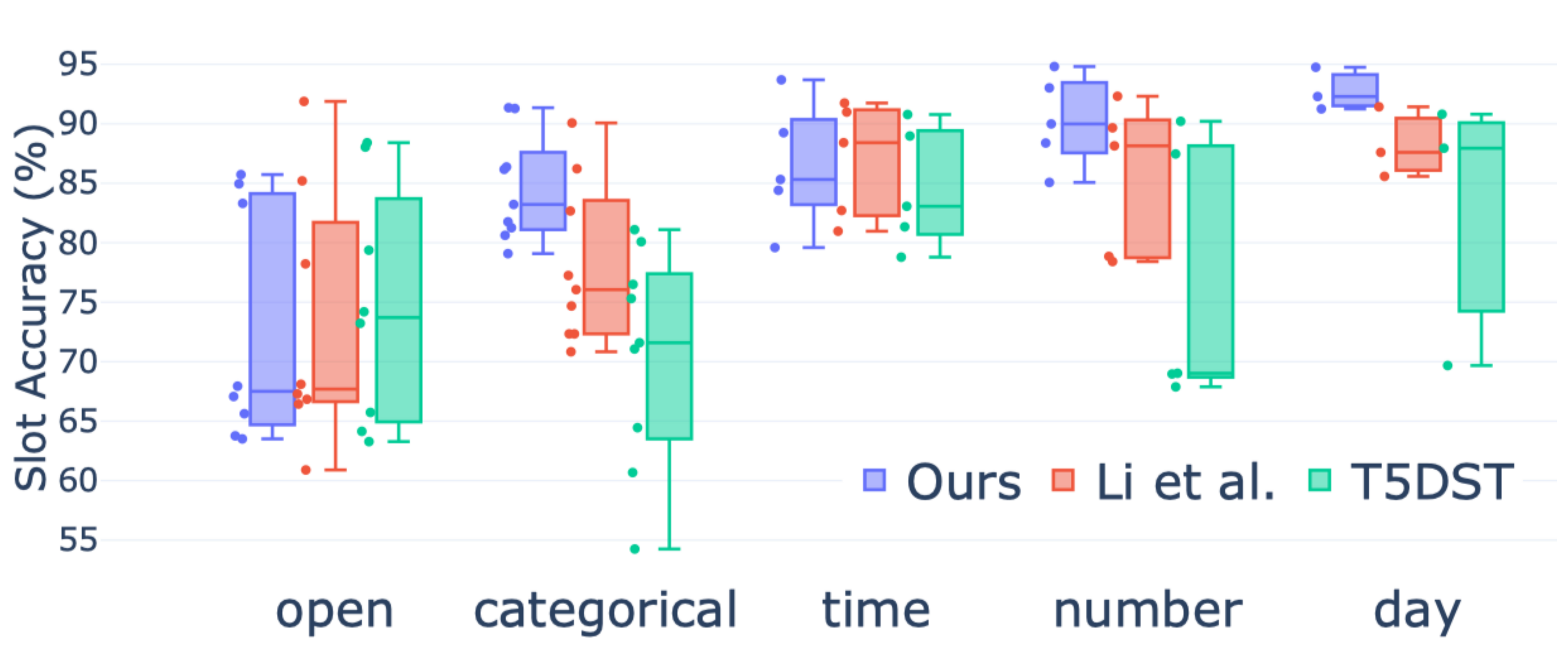
Experiments

Model	Params#	5	10	20	1%	5%	10%	5	10	20	1%	5%	10%	5	10	20	1%	5%	10%
		Attraction (3 slots, 1% = 27 conv.)						Hotel (10 slots, 1% = 33 conv.)						Restaurant (7 slots, 1% = 38 conv.)					
TRADE		—	—	—	—	52.19	58.46	—	—	—	—	31.93	41.29	—	—	—	—	47.31	53.65
DSTQA		—	—	—	—	51.58	61.77	—	—	—	—	33.08	**49.69**	—	—	—	—	35.33	54.27
T5DST	60M	4.77	21.93	30.57	40.68	52.12	60.13	8.19	13.46	17.94	18.63	38.76	46.13	13.80	19.51	22.79	29.47	**53.32**	58.44
Lee et al.	60M	6.33	19.12	34.53	37.56	54.34	58.75	9.31	15.76	22.07	24.41	**40.11**	42.98	15.87	19.66	22.15	30.96	48.94	**58.59**
Li et al.	335M	7.90	27.09	35.63	42.18	49.13	60.85	12.49	15.15	19.44	24.04	37.88	46.47	17.27	22.30	25.68	30.70	49.75	58.50
Ours	271K	**33.56**	**39.41**	**45.75**	**47.28**	**56.99**	**63.61**	**15.63**	**18.18**	**22.50**	**33.01**	38.24	45.60	**19.76**	**25.72**	**27.65**	**34.40**	50.81	55.79
		Taxi (4 slots, 1% = 15 conv.)						Train (6 slots, 1% = 29 conv.)						Average					
TRADE		—	—	—	—	59.03	60.51	—	—	—	—	48.82	59.65	—	—	—	—	47.86	54.71
DSTQA		—	—	—	—	58.25	59.35	—	—	—	—	50.36	61.28	—	—	—	—	45.72	57.27
T5DST	60M	48.22	53.74	58.27	58.19	59.23	69.03	12.31	21.93	36.45	43.93	69.27	69.48	17.46	26.11	33.20	38.18	54.54	60.64
Lee et al.	60M	45.32	49.93	58.58	58.52	60.77	**71.23**	13.57	25.02	38.52	50.26	69.32	69.72	18.08	25.90	35.17	40.34	54.70	60.25
Li et al.	335M	50.99	57.47	58.49	58.26	**61.68**	69.23	17.56	27.42	39.27	45.32	**71.69**	73.45	21.24	29.89	35.70	40.10	54.03	**61.70**
Ours	271K	**51.11**	**59.63**	**60.89**	**60.33**	61.63	63.00	**18.95**	**30.95**	**50.34**	**52.05**	69.51	**75.00**	**27.80**	**34.78**	**41.43**	**45.41**	**55.44**	60.60

Setting: low-resource Joint Goal Accuracy on MultiWOZ 2.0, compared with prompt-based generative DST models

Results

- Higher JGA than all baselines using 1% or less training data while using less than 0.5% of parameters
- Comparably performance with baselines when using 5% or 10% data



Compared with baselines, our model is comparable on "open" and "time" slots, and superior on "categorical", "number" and "day" slots

Model	Attr.	Hotel	Rest.	Taxi	Train	Avg
w/o segment emb.	34.35	23.18	27.33	59.69	43.30	37.57
w/o reiteration	45.08	27.57	33.48	59.89	51.08	43.42
Full model	47.28	33.01	34.40	60.33	52.05	45.41

Ablation study shows effectiveness of **segment awareness** and the **reiteration** technique