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

Anonymous Authors

Dialogue State Tracking

- 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 prompt to provide information about
  - Slot name, slot description, slot type, possible values, priming examples, questions …

`[usr] while in cambridge i need a hotel that has free parking and free wifi. … [usr] no, that s it. goodbye!`

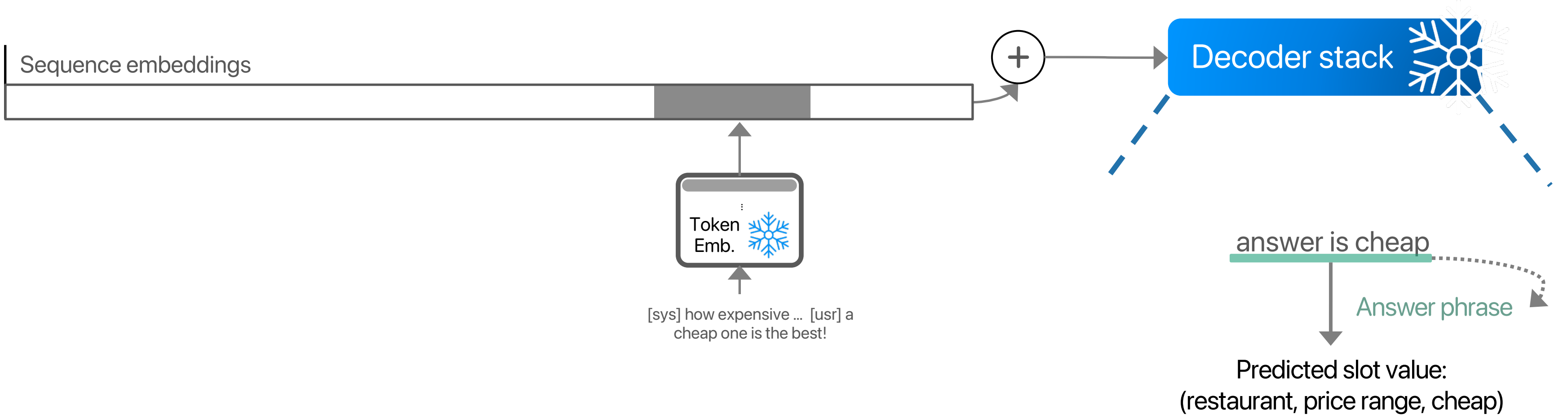
Restaurant, book time	*11:30*
Restaurant, name	*null*
Hotel, type	*expensive*
Hotel, internet	*null*
…	

Parameter-Efficient Dialogue State Tracking

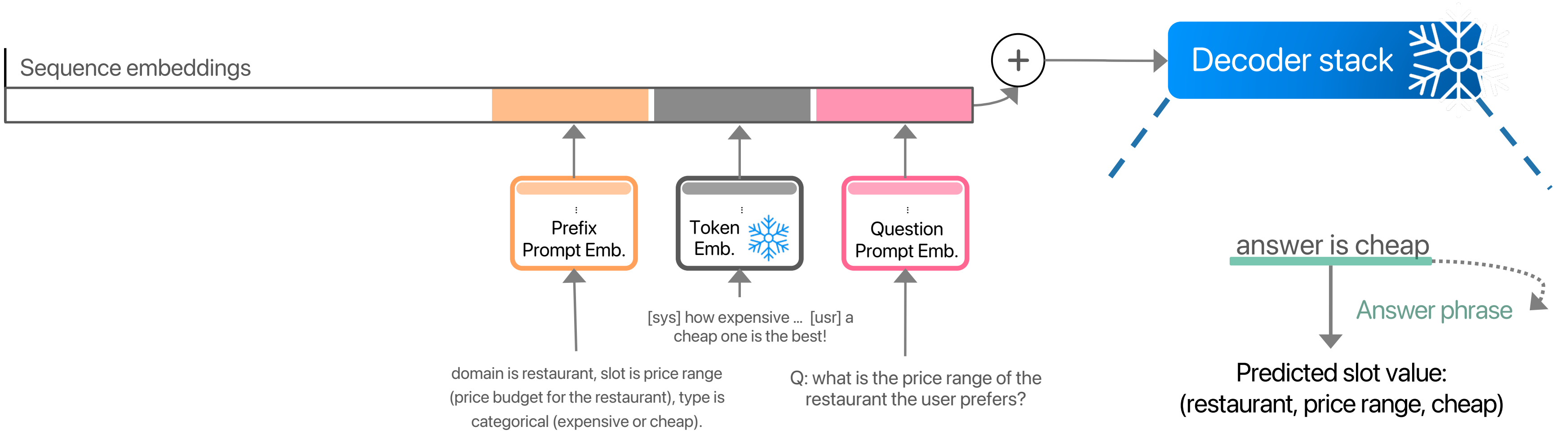
- Existing works all fine-tune the entire LM along with the prompt token embeddings
- Real-world deployment needs to train and host separate LMs for different domains and tasks
- Limited data is available for new domain or task

Parameter-efficient and **data-efficient**
DST model for low-resource settings

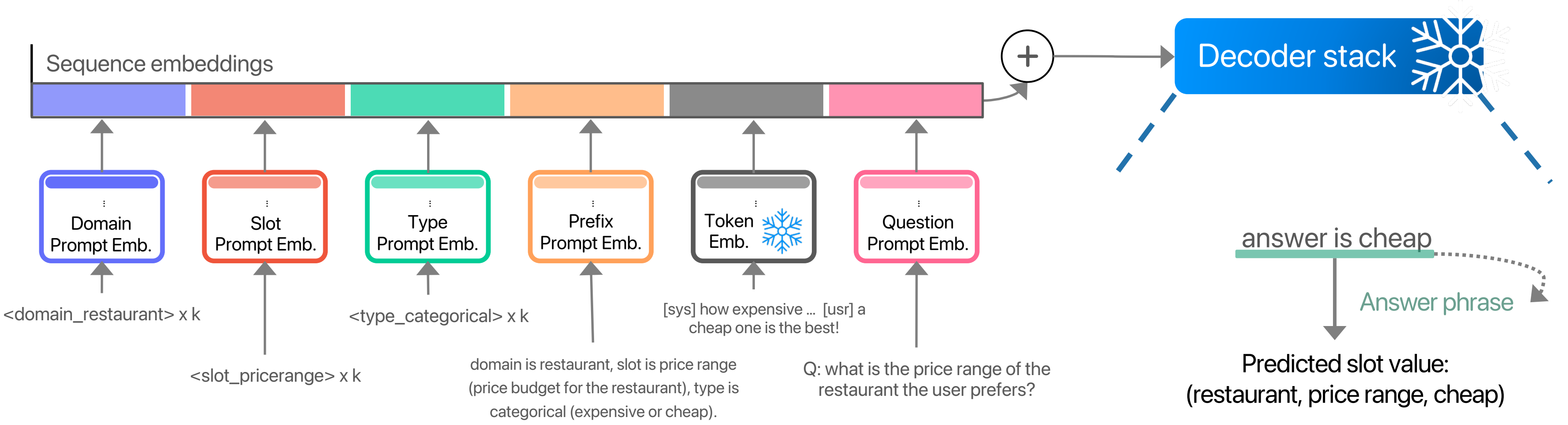
Generative seq2seq dialogue state tracking



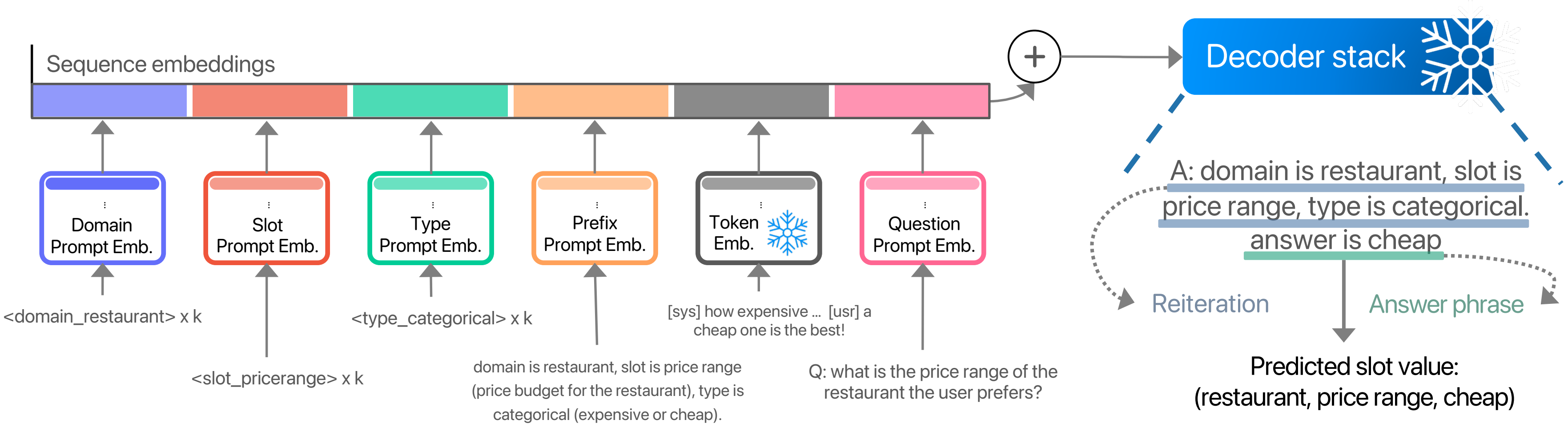
Word-mapping prompt tokens



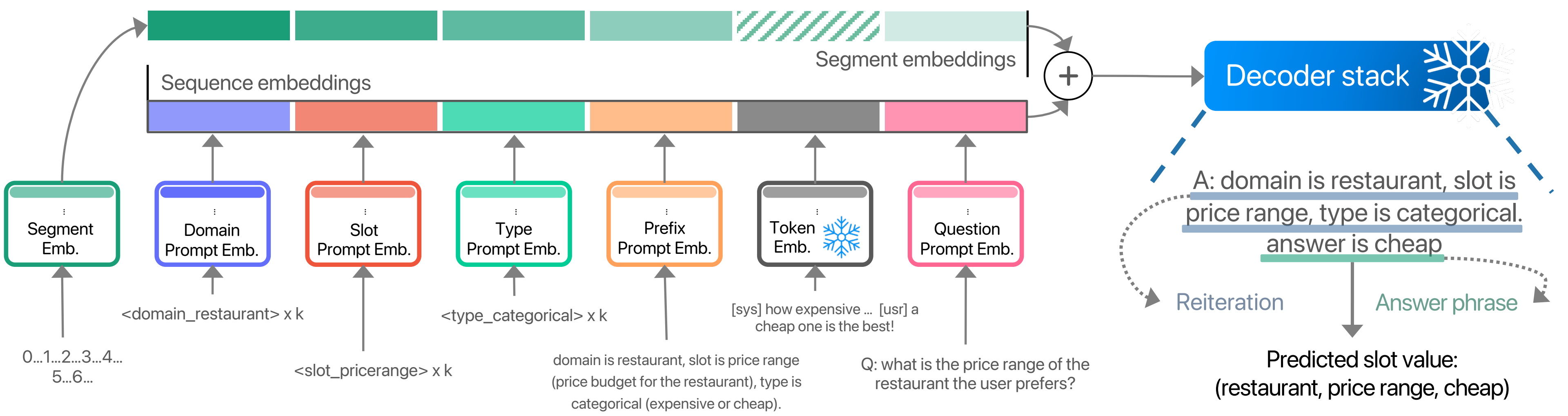
Task prompt tokens



Task metadata in objective



Distinguishing segments



Experimental setup

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																			
DSTQA																			
T5DST	60M																		
Lee et al.	60M																		
Li et al.	335M																		
Ours	271K																		
		Taxi (4 slots, 1% = 15 conv.)						Train (6 slots, 1% = 29 conv.)						Average					
TRADE																			
DSTQA																			
T5DST	60M																		
Lee et al.	60M																		
Li et al.	335M																		
Ours	271K																		

Dataset: MultiWOZ 2.0

Low-resource setting

Experimental setup

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																			
DSTQA																			
T5DST	60M																		
Lee et al.	60M																		
Li et al.	335M																		
Ours	271K																		
		Taxi (4 slots, 1% = 15 conv.)						Train (6 slots, 1% = 29 conv.)						Average					
TRADE																			
DSTQA																			
T5DST	60M																		
Lee et al.	60M																		
Li et al.	335M																		
Ours	271K																		

Dataset: MultiWOZ 2.0

Low-resource setting

Experimental setup

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																			
DSTQA																			
T5DST	60M																		
Lee et al.	60M																		
Li et al.	335M																		
Ours	271K																		
		Taxi (4 slots, 1% = 15 conv.)						Train (6 slots, 1% = 29 conv.)						Average					
TRADE																			
DSTQA																			
T5DST	60M																		
Lee et al.	60M																		
Li et al.	335M																		
Ours	271K																		

Dataset: MultiWOZ 2.0

Low-resource setting

Experimental setup

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																			
DSTQA																			
T5DST	60M																		
Lee et al.	60M																		
Li et al.	335M																		
Ours	271K																		
		Taxi (4 slots, 1% = 15 conv.)						Train (6 slots, 1% = 29 conv.)						Average					
TRADE																			
DSTQA																			
T5DST	60M																		
Lee et al.	60M																		
Li et al.	335M																		
Ours	271K																		

Evaluation metric: Joint Goal Accuracy

SOTA JGA using 0.5% or less parameters

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

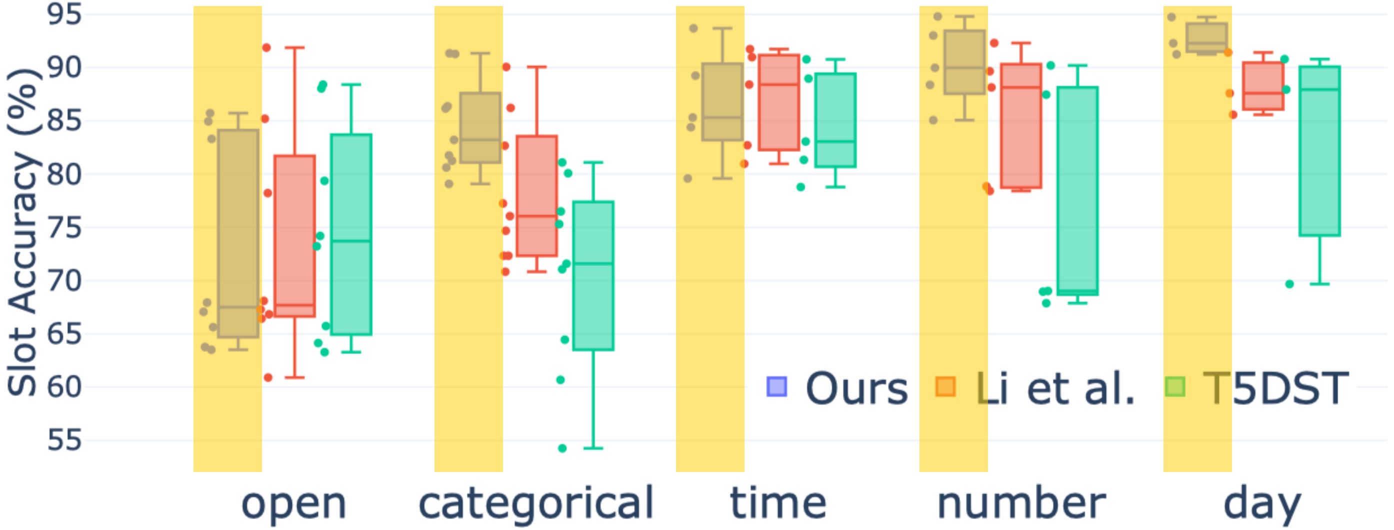
Our model achieves higher JGA than all baselines when using 1% or less training data while uses only 0.08% and 0.45% of parameters

SOTA JGA using 0.5% or less parameters

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

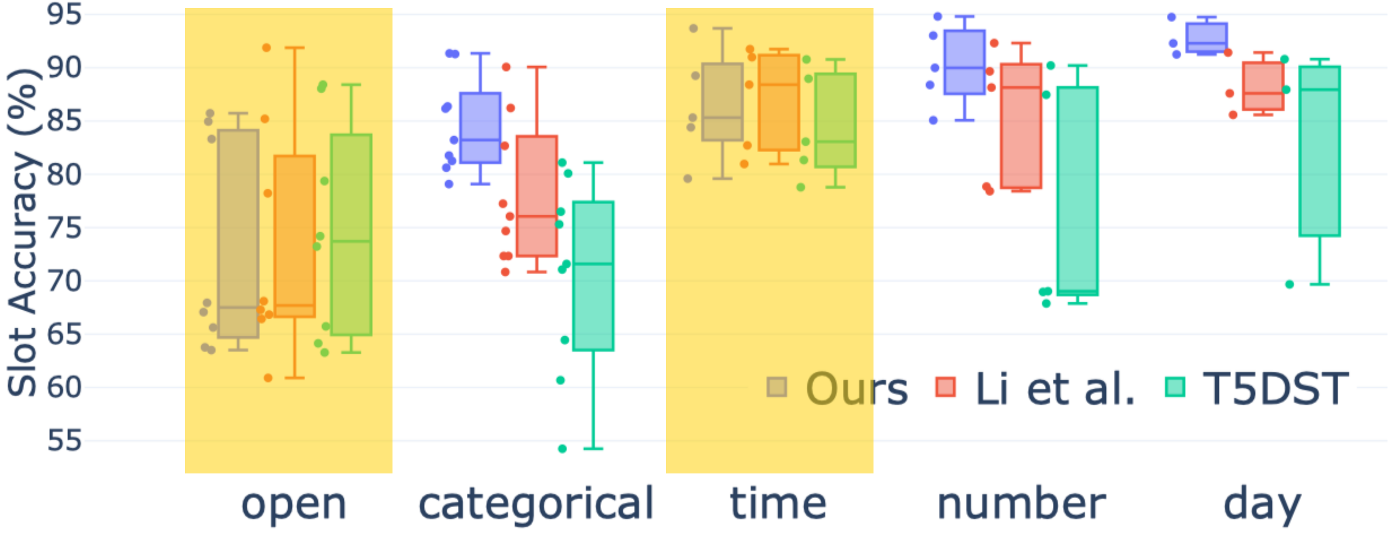
Our model achieves higher JGA than all baselines when using 1% or less training data while uses only 0.08% and 0.45% of parameters

More superior for categorical slots



Relatively worse performance on "open" slots

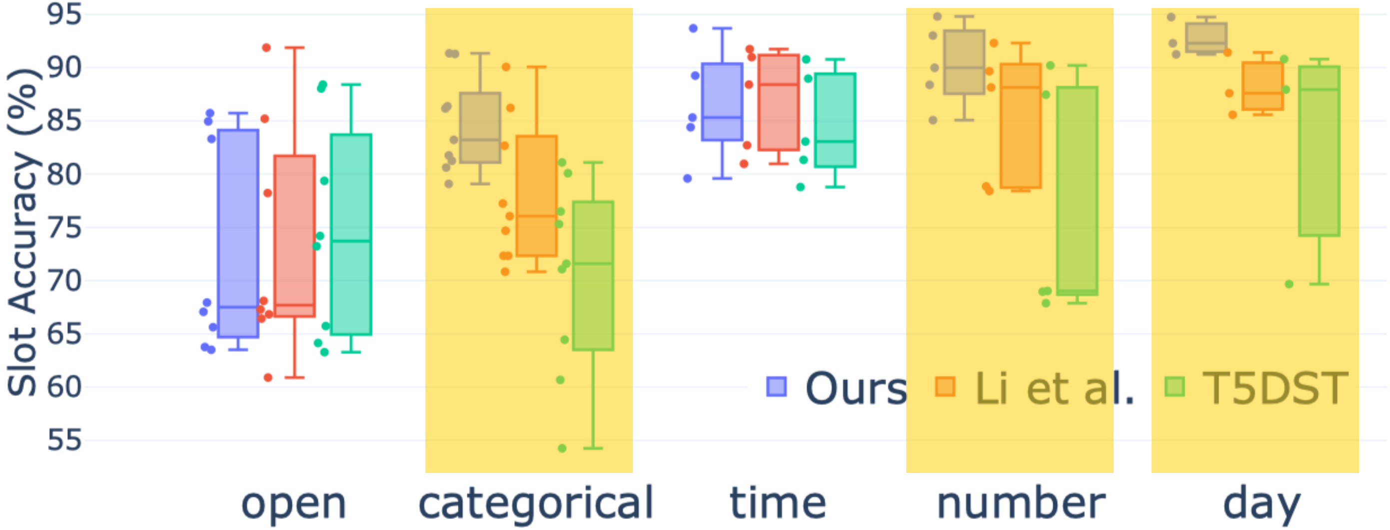
More superior for categorical slots



Compared with baselines

- **Comparably** on slots with larger output candidate space
- More **superior** for slots with smaller output candidate space

More superior for categorical slots



Compared with baselines

- **Comparably** on slots with larger output candidate space
- More **superior** for slots with smaller output candidate space

Ablation study

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

Removing segment embedding leads to on average 7.8 points JGA drop

Ablation study

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

Removing reiteration technique leads to an almost 2 points JGA drop

Conclusion

- We propose a parameter-efficient DST model using prompt tuning
  - It represents tasks with soft prompt tokens
  - Segment awareness
  - Reiteration
- Our model achieves state-of-the-art low-resource DST performance with less than 0.5% parameters compared with fine-tuning LM.

Thanks!