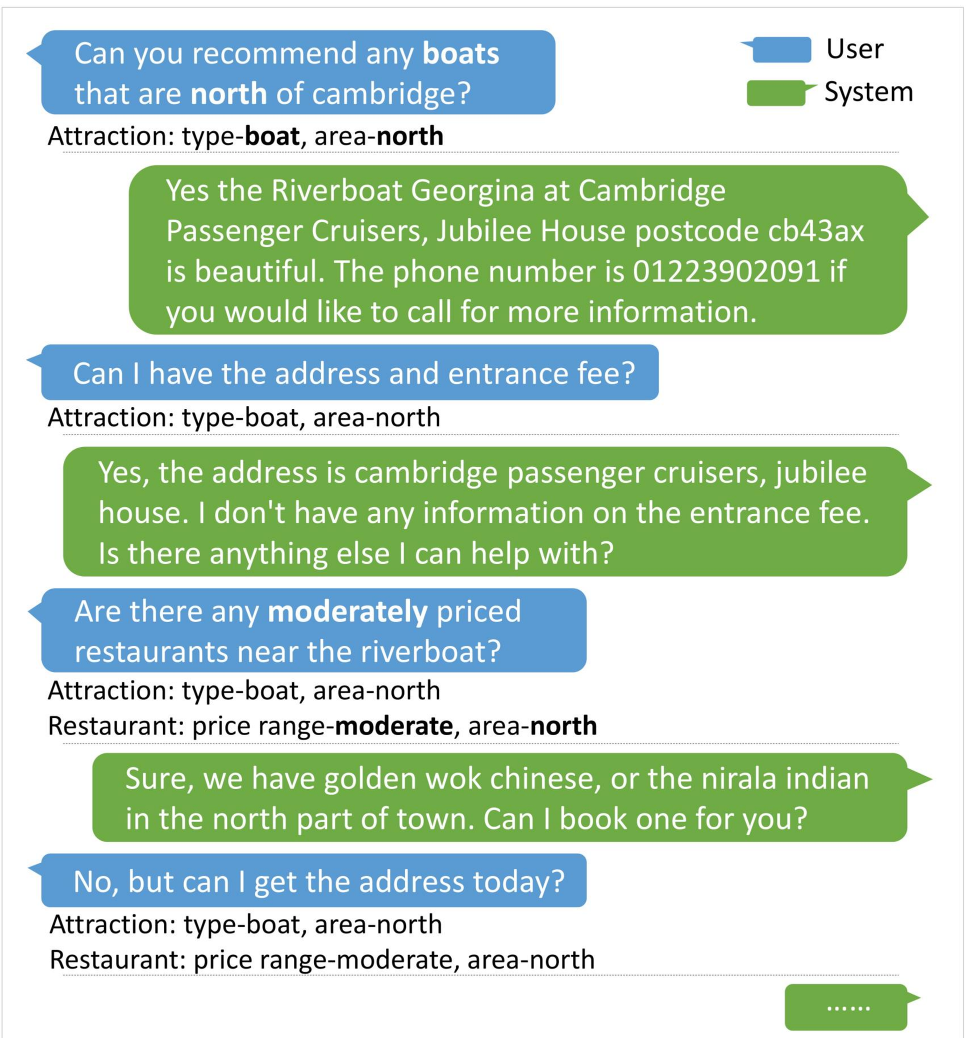
PROGRESSIVE DIALOGUE STATE TRACKING FOR MULTI-DOMAIN DIALOGUE SYSTEMS

Jiahao Wang*, Minqian Liu† and Xiaojun Quan*
*School of Computer Science and Engineering, Sun Yat-sen University
†School of Computer Science and Engineering, South China University of Technology

Paper 2250

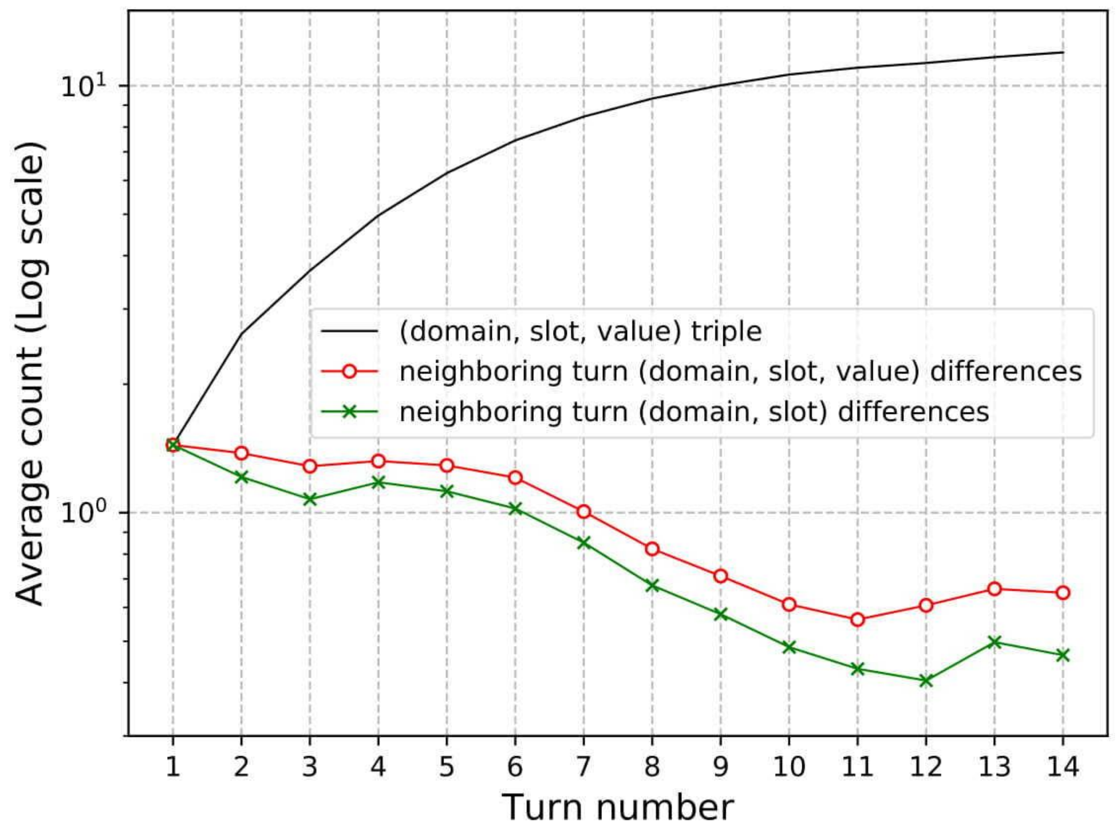
Introduction

Task-oriented dialogue systems aim to facilitate people with such services such as taxi booking or hotel reservation through multi-turn natural language conversations. The dialogue state tracker keeps close track of the dialogue states to manage information about the tasks. The dialogue states are usually organized in triples such as **domain-slot-value**.



Our work is inspired by two critical observations in multi-domain dialogue state tracking data:

- ***Accumulating state triples***. The number of triples in dialogue states increases with the growth of dialogue turns.
- ***Adjacent state dependencies***. Although the states are accumulating, the difference between two adjacent turns is constantly small.



The contributions of our work:

- We propose to divide DST into two successive stages, i.e., progressive domain-slot tracking and shrunk value prediction, based on our two observations.
- We adopt three levels of embeddings and attentions to model the domain-slot structure and capture the information on different levels.
- The progressive tracker predicts domain-slot pairs in parallel and reduce the number of domain-slot candidates significantly for value prediction, making our model more scalable and efficient.

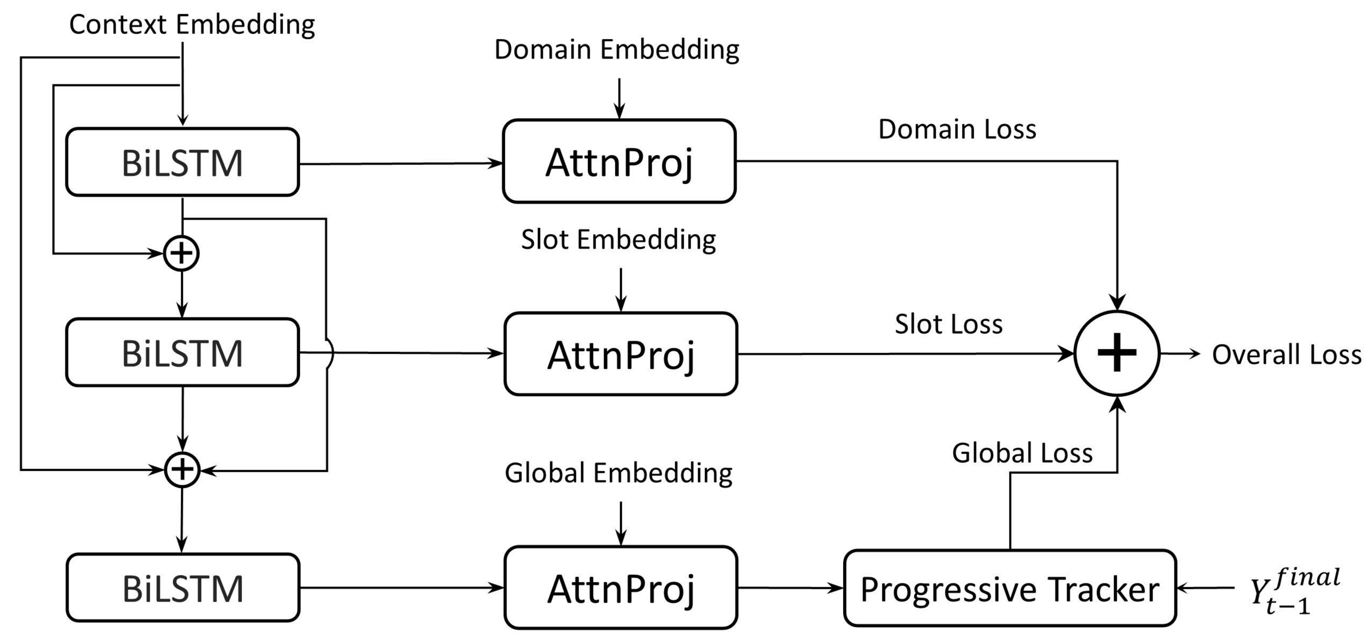
Method

The dialogue state tracking is divided into two successive procedures: domain-slot tracking and shrunk value prediction.

1. The domain-slot tracking model contains a multi-level attention module and a progressive domain-slot tracker.
2. The shrunk value prediction model takes the predicted domain-slots and the dialogue context as inputs, and outputs the corresponding values for each domain-slot with a pointer generator network.

- **Multi-Level Attention**

The multi-level attention module use three levels of attention (domain, slot, global) to model the current-turn domain-slot predictions in a finer-grained manner.



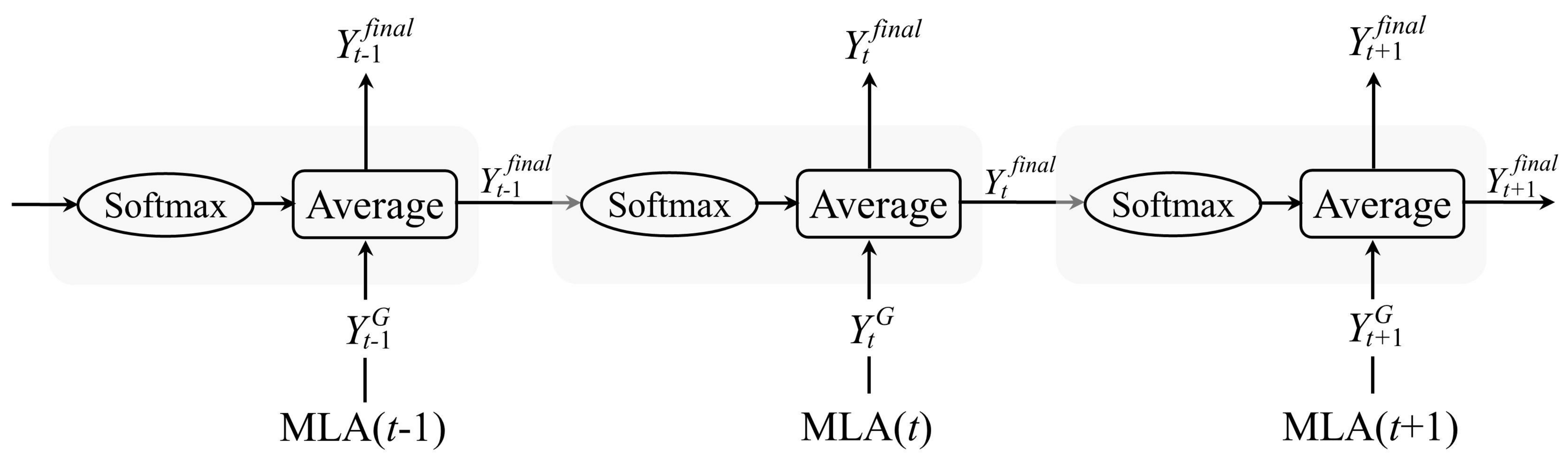
- **Progressive Domain-Slot Tracking**

The progressive domain-slot tracker takes domain-slot predictions of previous turn and current turn as input, and outputs the averaged results after smoothing the previous turn predictions.

$B^{(t)} = \{b_1^{(t)}, b_2^{(t)}, \dots, b_m^{(t)}\}$ is the dialogue state at t-th turn, where $b_i^{(t)}$ is the i-th domain-slot pair with label from {*pointer, dontcare, none*}.
The change from from $B^{(t-1)}$ to $B^{(t)}$ is $\nabla^{(t)} = \{\delta_1^{(t)}, \delta_2^{(t)}, \dots, \delta_m^{(t)}\}$.
$B_{ptd}^{(t-1)}$ represents the *pointed* and *dontcare* states at t-th turn.

$$\delta_i^{(t)} = \begin{cases} b_i^{(t)}, & b_i^{(t)} \neq b_i^{(t-1)} \\ \varnothing, & b_i^{(t)} = b_i^{(t-1)} \end{cases}$$

- ***Accumulating state triples***: $|B_{ptd}^{(t-1)}| < |B_{ptd}^{(t)}|$ for most utterances.
- ***Adjacent state dependencies***: Changed states $|\nabla^{(t)}|$ is small.



$$Y_t^{final} = \frac{\mathrm{Softmax}(Y_{t-1}^{final}) + Y_t^G}{2}$$

Result

- **Joint Goal Accuracy**

Model	MultiWOZ 2.0 Acc. (%)	MultiWOZ 2.1 Acc. (%)
GLAD [2]	35.57	-
DST Reader [17]	39.41	36.4
COMER [9]	45.72	-
TRADE [8]	48.62	45.6
NADST [18]	50.52	49.0
SAS [19]	51.03	-
DST-SC [20]	**52.24**	49.6
PRO-DST (Ours)	51.48	**49.9**

- **Ablation study** with Joint Domain-Slot Accuracy & Joint Goal Accuracy

Row	Model	Domain-Slot Acc.	Goal Acc.
1	Our Model	58.06	49.89
2	− output of previous turn	56.28	48.4
3	− smoothing of previous-turn output	51.14	44.75
4	− domain&slot attention module	47.92	41.99
5	− learnable domain&slot emb. + fixed domain&slot emb.	58.06	49.80

- The predictions in the previous turn effectively improve the performance, but the smoothing operation is critical.
- The information in the domain and slot levels improves the performance significantly.

- **Time complexity** of different value sequence generation-based models, *n* is the number of domain-slots.

Model	Best	Worst
COMER [9]	$\Omega(1)$	$O(n)$
TRADE [8]	$\Omega(n)$	$O(n)$
NADST [18]	$\Omega(1)$	$O(n)$
SAS [19]	$\Omega(n)$	$O(n)$
DST-SC [20]	$\Omega(n)$	$O(n)$
PRO-DST (Ours)	$\Omega(1)$	$O(n)$

Discussion

- The two observations can direct further research on developing more accurate domain-slot tracker, e.g., utilizing large-scaled pretrained language model.
- The three-level attentions enables finer-grained modeling of domain-slot predictions and can be extended to more complicate dialogue state ontology setting.
- The progressive domain-slot tracking mechanism can be improved to focus more on the state changes between neighboring turns.