A SLOT-SHARED SPAN PREDICTION-BASED NEURAL NETWORK FOR MULTI-DOMAIN DIALOGUE STATE TRACKING

Abibulla Atawulla[1,2,3] Xi Zhou[1,2,3,†] Yating Yang[1,2,3,†] Bo Ma[1,2,3] Fengyi Yang[1,2,3]

[1] Xinjiang Technical Institute of Physics & Chemistry, Chinese Academy of Sciences, Urumqi, China
[2] University of Chinese Academy of Sciences, Beijing, China
[3] Xinjiang Laboratory of Minority Speech and Language Information Processing, Urumqi, China

Outline

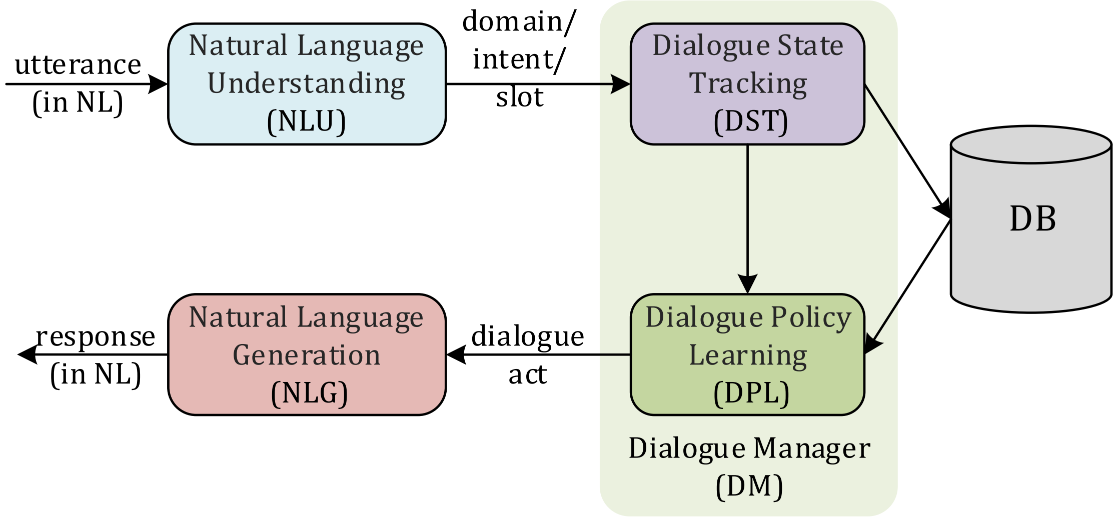
- Background
- Motivation
- Methods
- Experiments
- Conclusion

Background

- A typical task-oriented dialogue system consists of four key components, i.e., natural language understanding (NLU), dialogue state tracking (DST), dialogue policy learning (DPL) and natural language generation (NLG).

- Dialogue state tracking aims to estimating the dialogue state at each dialogue turn, where the state is represented in forms of a set of slot-value pairs.

- As an intermediate module, its performance directly affects subsequent modules, e.g., DPL.



Motivation

- In multi-domain dialogue state tracking, we focus on the value sharing among slots, i.e., some slots share the candidate slot values.

e.g., taxi-destination is usually accompanied with taxi-departure, the value of the taxi-destination may correspond to hotel-name, the value of the hotel-book day may be the same as train-day .

Dialogue

A1: The <u>Hilton</u> hotel is located on regent street. Would you like this one?

U1: I would like to book a room for <u>3 nights</u> for <u>2 people</u> starting on <u>Friday</u> please.

...

A3: Excellent. The booking worked. Your reference number is XEFDAS4S.

U3: I also booking looking to book a train that <u>arrives</u> there from Cambridge on the <u>same day</u> by <u>09:00</u>

DST model →

Dialogue State

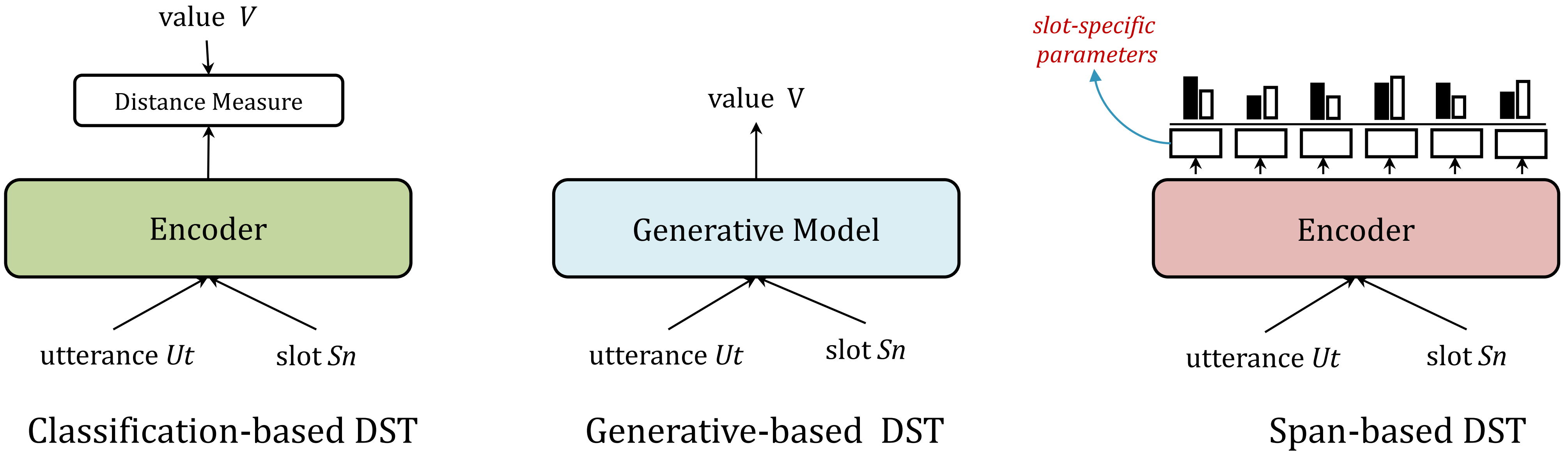
S1:hotel-name=Hilton hotel
hotel-book stay=3
hotel-book day=Friday
hotel-book people=2

...

S3: S1 + S2 +
train-day= Friday
train-leaveat= 09:00
train-destination= Hilton hotel
train-departure= Cambridge

Motivation

- The existing span prediction-based dialogue state tracking methods generally adopt slot-independent value extraction architecture, which ignore the mentioned value sharing phenomenon.
- Besides, the slot-independent design leads to poor scalability.



Methods

- Our model contains three main modules:

  - A Representing and Encoding Module is aimed to contextualize the dialogue context and slot semantics respectively.

  - A Dynamic Fusion Mechanism fuses the current dialogue context with different slot semantics to obtain slot-aware dialogue context representations.

  - A Slot-shared Value Extraction Module is responsible to extract values based on the slot-aware dialogue context representations.

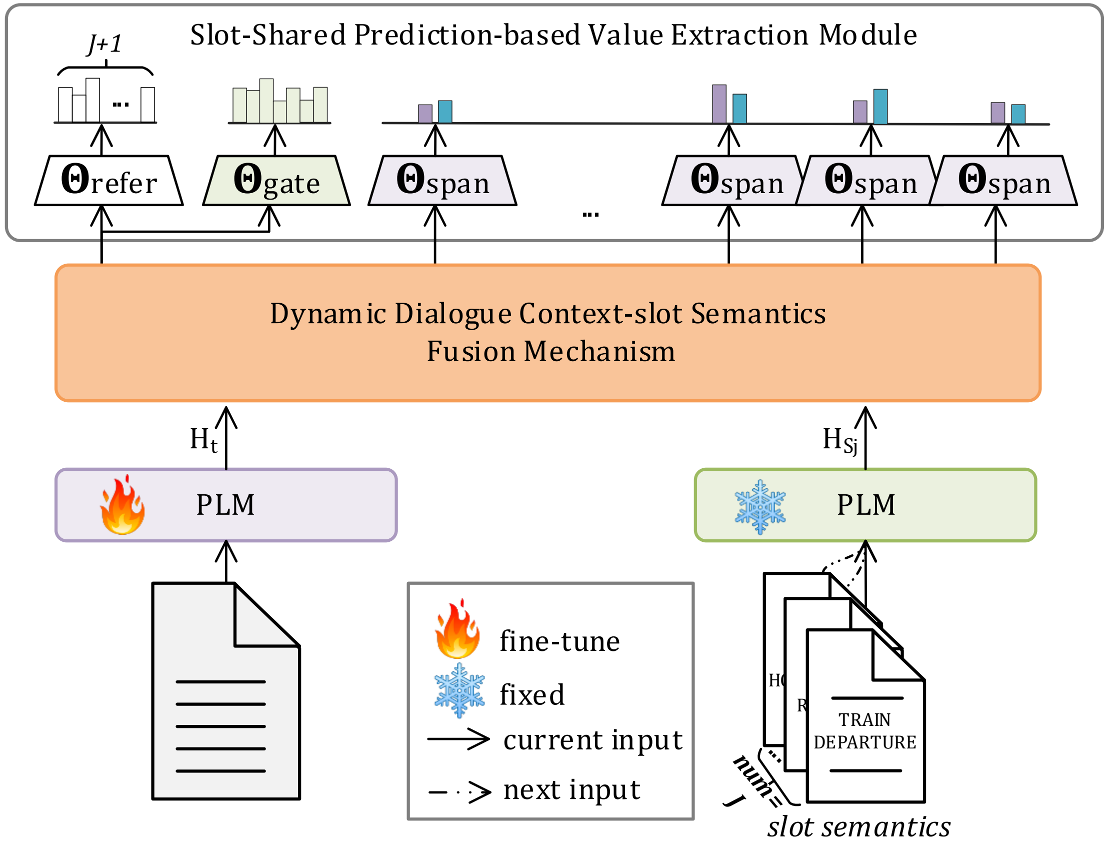


Fig 1: Illustration of our model

Methods

Dynamic Dialogue Context-slot Semantics Fusion Mechanism (DFM)

- A slot semantics can be describe as the concatenation of the slot name slot description and slot categorical type.
- The design of FDM is inspired by BiDAF[1] , a typical Reading comprehension model.
- DFM plays the routing role, highlighting different dialogue context tokens for different slots. Specifically, DFM firstly calculates similarity matrixes between the dialogue context and different slots, and then determines important dialogue context token with respect to each slot.

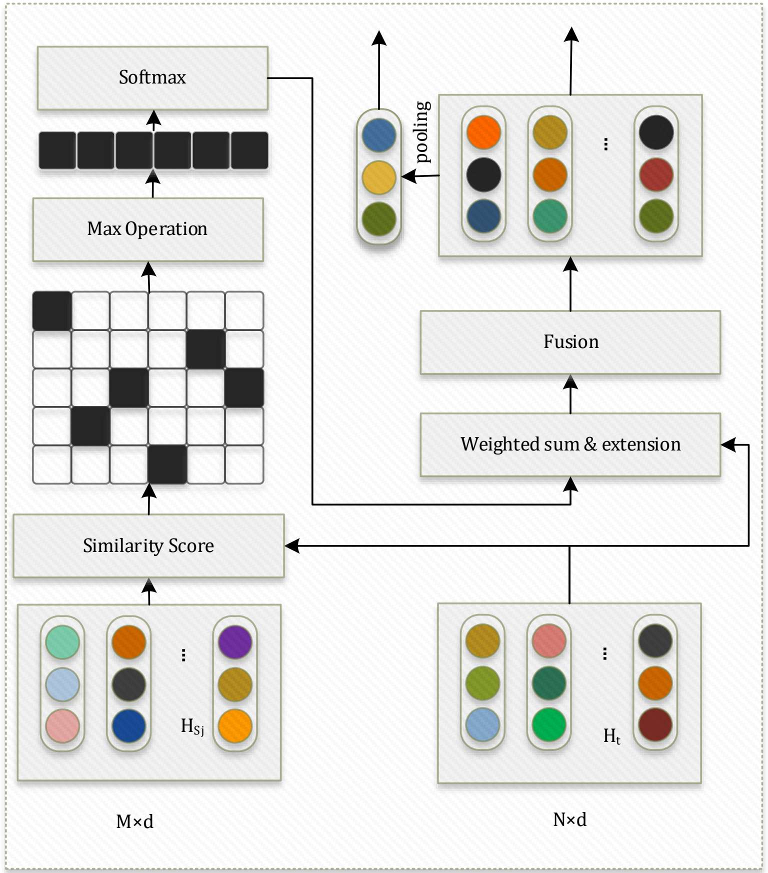


Fig 2: Dynamic Dialogue Context-slot Semantics Fusion Mechanism

[1] SEO M, KEMBHAVI A, FARHADI A, et al. Bidirectional attention flow for machine comprehension [J]. arXiv preprint arXiv:161101603, 2016.

Experiments

- Datasets Multi-domain Wizard-Of-Oz (MultiWOZ) series: Proposed by the University of Cambridge in 2018, it will consist of seven areas, including taxi, restaurant, train, attraction and hotel.
- Data set collection methods: Humans and Humans (H2H).

Datasets	**MultiWOZ**
Domain Num	7
Dialogue Num	8438
Average dialogue Turn Num	13.7
Dialogue state represent	slot-value
Slot Num	25
Value Num	4510

Experiments

Comparisons with other DST models

Model	Type	MultiWOZ 2.1	MultiWOZ 2.2
TRADE[5]	G	46.00	45.40*
DS-DST[22]	C+S	51.21	51.70*
SOM-DST[6]	G	53.68	53.81*
TripPy[16]	S	55.29	-
STAR[23]	C	56.36	-
AG-DST[7]	G	-	57.26
SDP-DST[9]	G	56.66	57.60
MSP-L[15]	S	57.20	57.70
SSNet	S	**59.48**	**62.10**

Joint goal accuracy on MultiWOZ 2.1 and 2.2.

Domain expansion and Ablation study

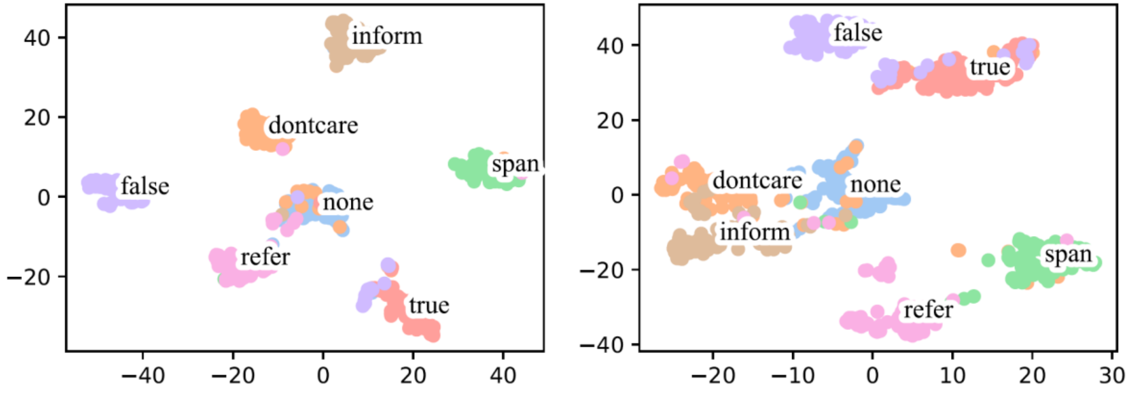
JGA (%)	SA (%)	Per-slot Accuracy (%)			
		destination	*departure*	*leaveby*	*arriveat*
61.65	77.71	66.82	72.45	87.29	84.27

Domain expansion experiment on MultiWOZ 2.2.

Fusion Method	JGA(%)	SA(%)
Tra_ATT	49.77 $\downarrow_{12.33}$	96.83 $\downarrow_{1.11}$
Mean	60.03 $\downarrow_{2.07}$	97.80 $\downarrow_{0.14}$
DFM	**62.10**	**97.94**

Ablation study on MultiWOZ 2.2 .

Analysis Experiments



t-SNE visualization of different fusion methods. The left is DFM, and the right is Tra ATT.

Class—Number	DFM	Tra_ATT
none — 209897	1.00	0.99
dontcare — 235	0.62	0.02
span — 7397	0.96	0.93
true — 416	0.76	0.70
false — 145	0.85	0.77
inform — 2702	0.93	0.75
refer — 368	0.73	0.58

F1 score of each class using different fusion methods

Conclusion

- we propose a Slot-shared Span Prediction-based Neural Network (SSNet) to improve the scalability of multi-domain dialogue state tracking model.

- In addition, we introduce a Dynamic Fusion Mechanism to fuse the dialogue context with all slot semantics , which can distinguish similar slots effectively.

- Experimental results show that SSNet achieves 59.48% and 62.10% joint goal accuracy on MultiWOZ 2.1 and MultiWOZ 2.2 datasets, respectively.

Thank you for your time and patience!

Multilingual Information Technology Lab, Urumqi, China
Abibulla Atawulla
Email: aibibulaatawula19@mails.ucas.sc.cn