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

Abibulla Atawulla Xi Zhou Yating Yang Bo Ma Fengyi Yang

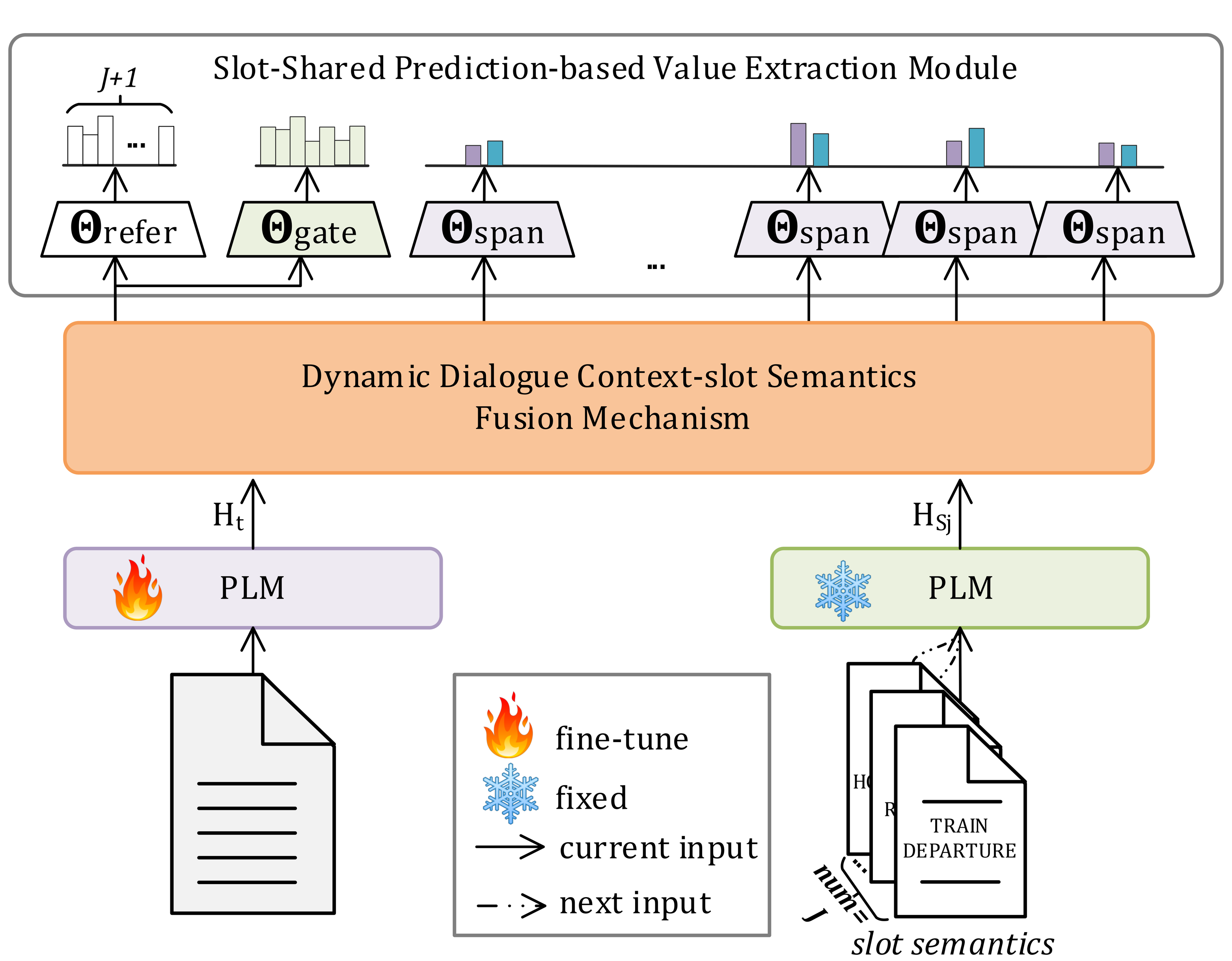
ICASSP 2023
4 - 10 JUNE, RHODES ISLAND, GREECE

INTRODUCTION

In multi-domain dialogue state tracking, some slots share the same candidate values, e.g., *taxi-destination* is usually accompanied with *taxi-departure*, the value of the *taxi-destination* may correspond to *restaurant-name*. However, the existing span prediction-based dialogue state tracking methods generally adopt slot-independent value extraction architecture, which ignore the above value sharing. Besides, the slot-independent design leads to poor scalability.

We propose a Slot-shared Span Prediction based Network with a general value extraction module for all slots to tackle these problems. To ensure that the value extraction module is able to distinguish different slots, we introduce a Dynamic Fusion Mechanism to extract different slot-aware features. The mechanism 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.

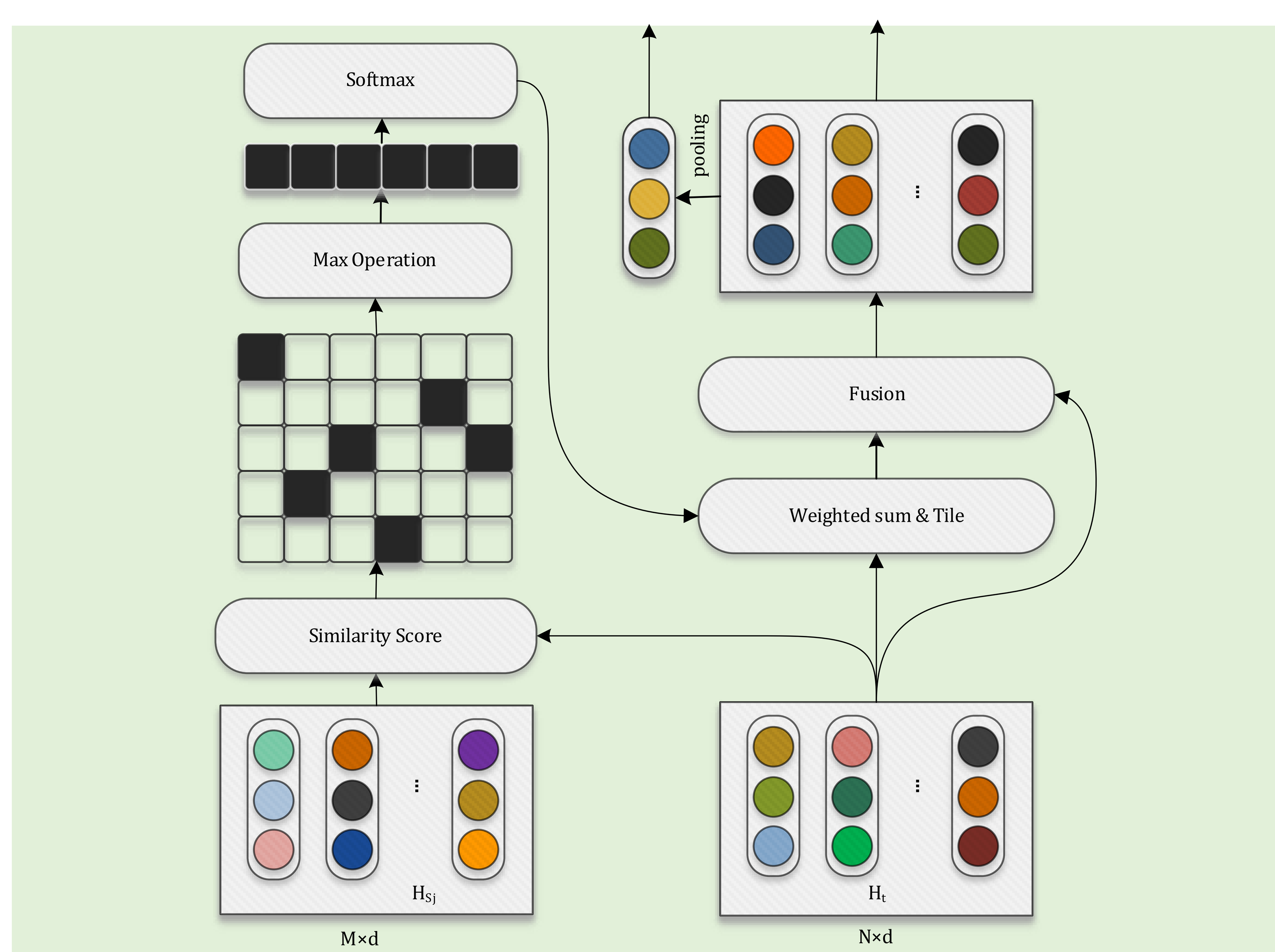
MODEL ARCHITECTURE



DYNAMIC FUSION MECHANISM

Input: dialogue context representations H_t at first turn; slot semantic representations sets $H_S = \{H_{s1}, \dots, H_{sj}\}$;
Output: token level slot-aware dialogue context representations sets U_t and sentence level dialogue representations sets $u_t^{[CLS]}$.

for $j = 1,2 \dots J$ **do**
Compute the similarity matrix D_j between H_t and H_{sj} ;
Calculate attention weight $B_j = [b_1, b_1, \cdots, b_N]$;
Tile the weighted sum $\tilde{h} = \sum_{n=0}^{N} b_n . h_t^n$ for N times and obtain $\widetilde{H}_t$;
Fuse the H_t with $\widetilde{H}_t$ and obtain token level slot-aware dialogue context representations $U_t^j = W^T.[H_t; H_t \odot \widetilde{H}_t]$, sentence level dialogue representations $u_{t,j}^{[CLS]} = BERT_{pooling}(U_t^j)$;
Fuse the auxiliary features to $u_{t,j}^{[CLS]}$ and obtain enhanced feature $\tilde{u}_{t,j}^{[CLS]} = u_{t,j}^{[CLS]} \oplus a_t^{ds} \oplus a_t^{inform}$;
end
Collect the token and sentence level dialogue represents under the all different slot semantics: the $U_t = \{U_t^1, \dots, U_t^J\}$ and $\tilde{u}_t^{[CLS]} = \{\tilde{u}_{t,1}^{[CLS]}, \dots, \tilde{u}_{t,J}^{[CLS]}\}$;
return U_t and $\tilde{u}_t^{[CLS]}$**;**



Overview of Dynamic Fusion Mechanism

EXPERIMENTAL RESULTS

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**

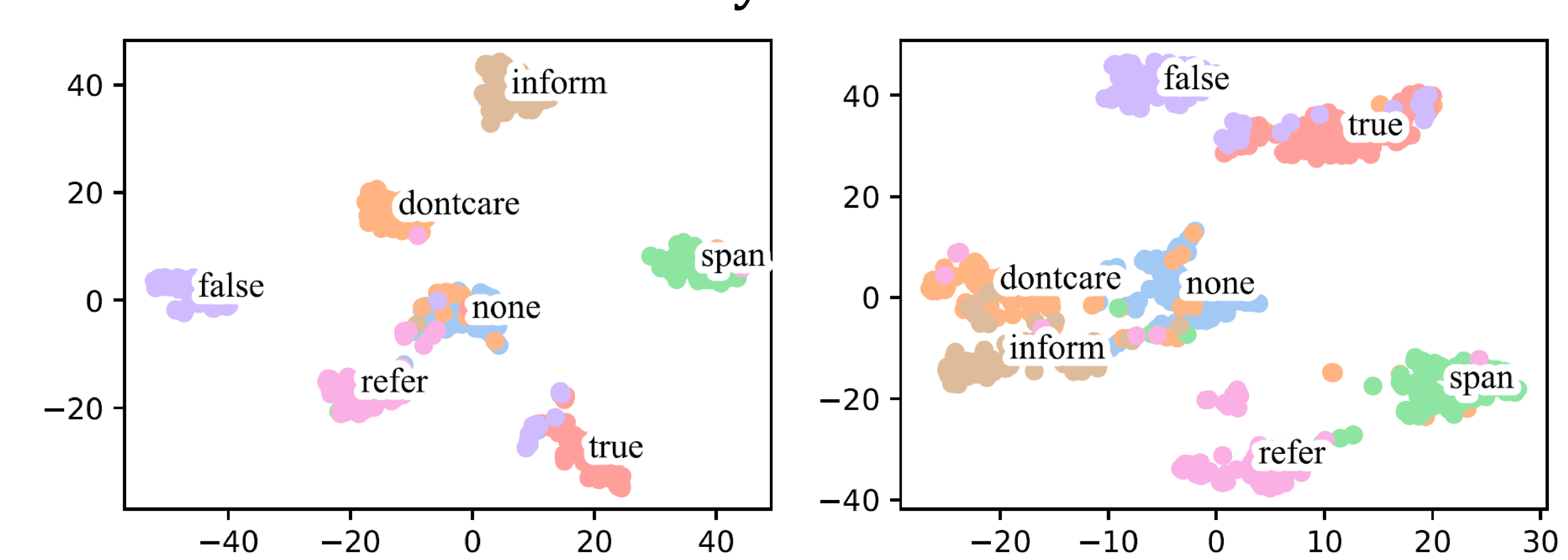
Joint goal accuracy on MultiWOZ 2.1 and 2.2.

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

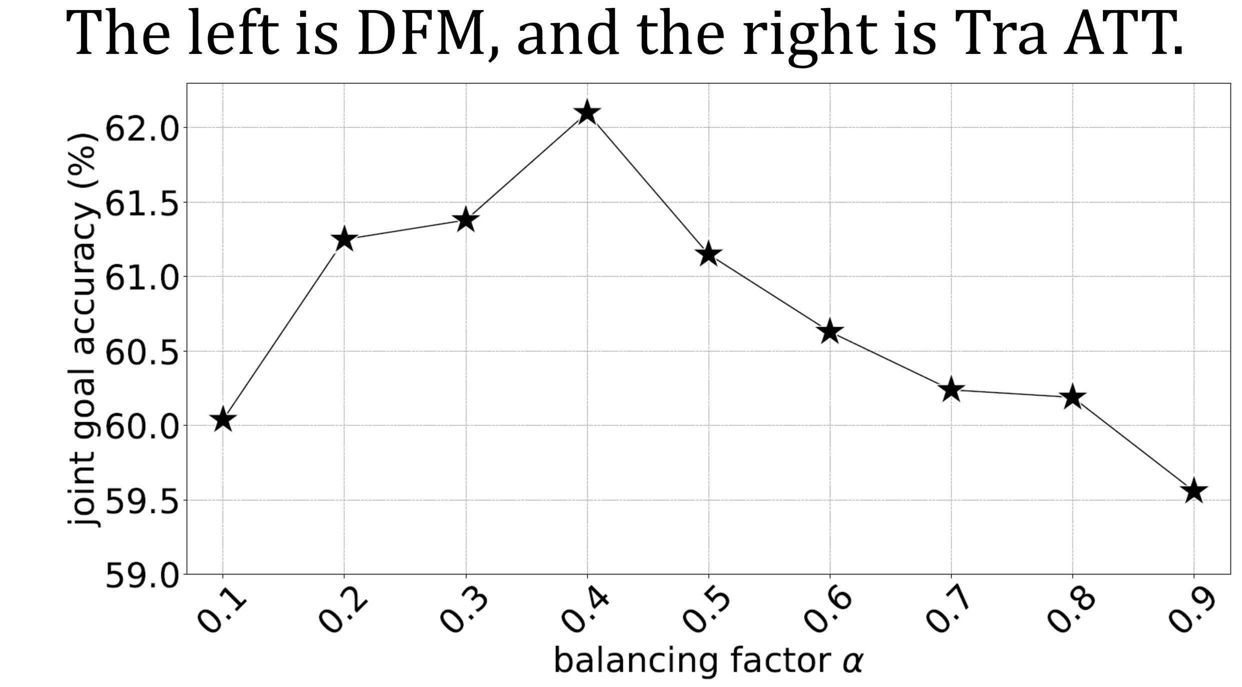
Domain expansion experiment on MultiWOZ 2.2.

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 .



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



The impact of balancing factor α