SUMBT: Slot-Utterance Matching for Universal and Scalable Belief Tracking

Hwaran Lee*, Jinsik Lee*, Tae-Yoon Kim

SK T-Brain, AI Center, SK Telecom

{hwaran.lee, jinsik16.lee, oceanos}@sktbrain.com

I. Introduction

- In goal-oriented dialog systems, belief trackers estimate the probability distribution of slot-values at every dialog turn
- Previous neural approaches have modeled domain- and slot-dependent belief trackers, and have difficulty in adding new slot-values, resulting in lack of flexibility of domain ontology configurations
- We propose a new approach to **universal** and **scalable belief tracker**, called ***slot-utterance matching belief tracker (SUMBT)***
- SUMBT learns the **relations between domain-slot-types and slot-values** appearing in utterances through attention mechanisms based on contextual semantic vectors
- Furthermore, SUMBT predicts slot-value labels in **a non-parametric way**

II. SUMBT: Slot-Utterance Matching Belief Tracker

1) Contextual Semantic Encoders

- Encode pairs of system x_t^{sys} and user utterances x_t^{usr}, using the pretrained BERT:

$$U_t = \mathrm{BERT}\left([\mathbf{x}_t^{sys}, \mathbf{x}_t^{usr}]\right)$$

- Also, literally encode domain-slot-type s and slot-values v_t at turn t:

$$\mathbf{q}^s = \mathrm{BERT}_{\mathrm{sv}}(\mathbf{x}^s), \quad \mathbf{y}_t^v = \mathrm{BERT}_{\mathrm{sv}}(\mathbf{x}_t^v).$$

  - The slot-value BERT encoder is not fine-tuned

2) Slot-Utterance Matching

- Retrieve the relevant information corresponding to the domain-slot-type from the utterances using attention mechanism

$$\mathbf{h}_t^s = \mathrm{MultiHead}(\underbrace{Q}_{Q^s}, \underbrace{K, V}_{U_t})$$

3) Belief Tracker

- Model dialog flows using an RNN

$$\mathbf{d}_t^s = \mathrm{RNN}(\mathbf{d}_{t-1}^s, \mathbf{h}_t^s)$$

- Additionally, normalize the RNN output

$$\hat{\mathbf{y}}_t^s = \mathrm{LayerNorm}(\mathbf{d}_t^s)$$

4) Training Criteria

- Distance metric based classifier

$$p\left(v_t | \mathbf{x}_{\leq t}^{sys}, \mathbf{x}_{\leq t}^{usr}, s\right) = \frac{\exp\left(-d(\hat{\mathbf{y}}_t^s, \mathbf{y}_t^v)\right)}{\sum_{v' \in \mathcal{C}_s} \exp\left(-d(\hat{\mathbf{y}}_t^s, \mathbf{y}_t^{v'})\right)}$$

- Training loss

$$\mathcal{L}(\theta) = -\sum_{s \in \mathcal{D}} \sum_{t=1}^{T} \log p(v_t | \mathbf{x}_{\leq t}^{sys}, \mathbf{x}_{\leq t}^{usr}, s)$$

  - By training all domain-slot-types together, the model can learn general relations between slot-types and slot-values, which helps to improve performance.

IV. Conclusion

- SUMBT achieved *the state-of-the-art joint accuracy* performance in WOZ 2.0 and MultiWOZ corpora
- *Sharing knowledge by learning from multiple domain data* helps to improve performance
- As future work, we plan to explore whether SUMBT can continually learn new knowledge when domain ontology is updated
- ***Our implementation is open-published in*** `https://github.com/SKTBrain/SUMBT`

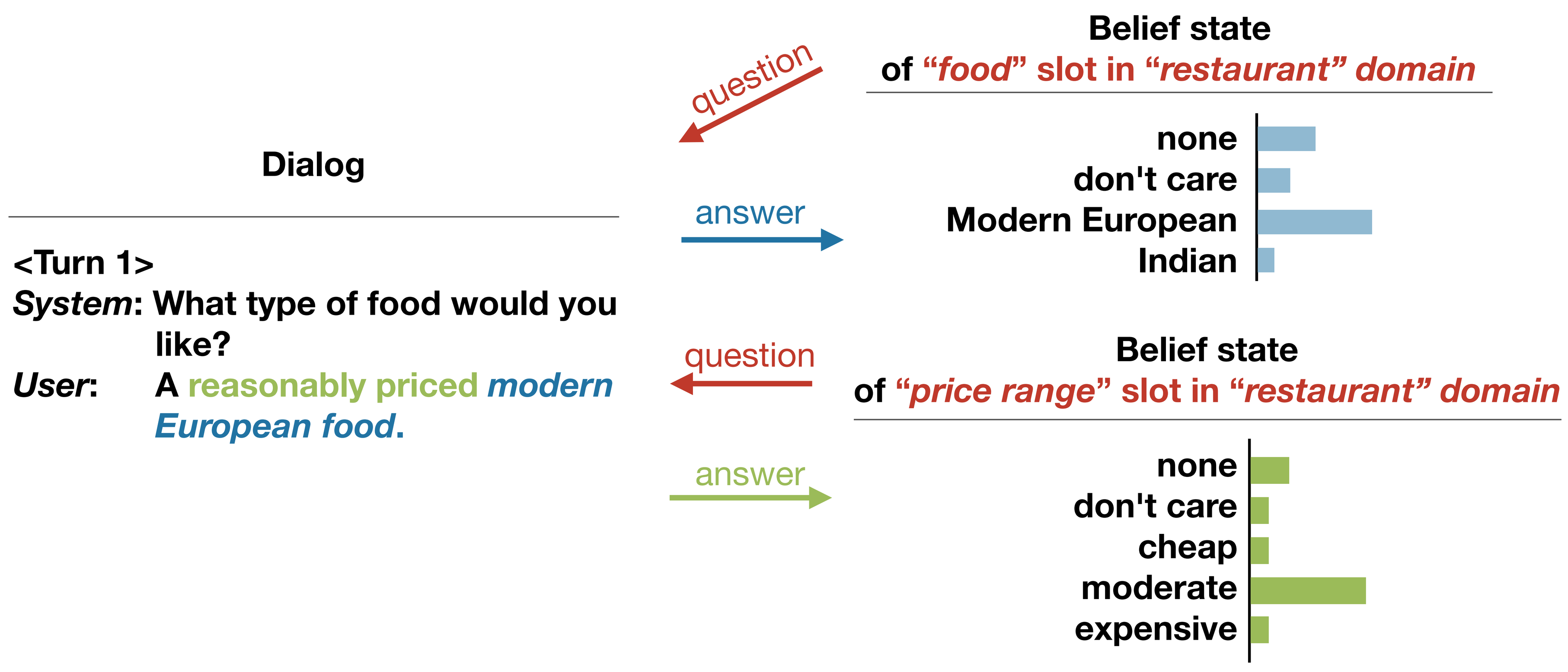

Figure 1. An example of multi-slot dialog state tracking and our motivation

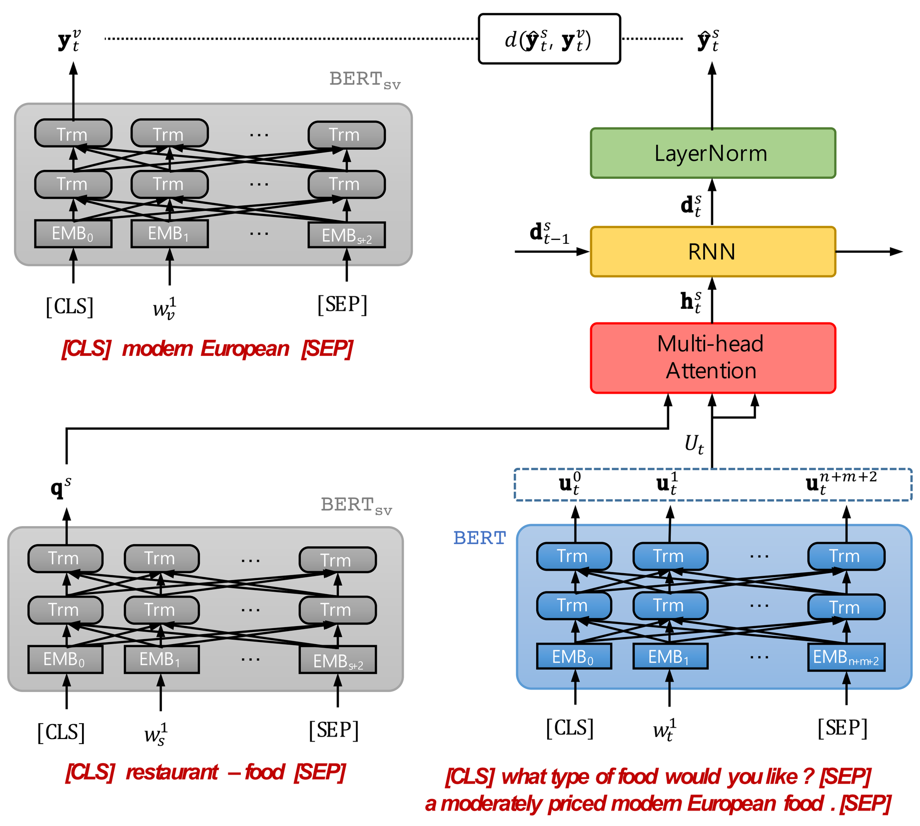

Figure 2. The architecture of slot-utterance matching belief tracker (SUMBT)

III. Experimental Results

1) WOZ 2.0

- Restaurant reservation domain, 3 slots (Area, Food, Price range)

Model	Joint Accuracy
NBT-DNN (Mrksic et al., 2017)	0.844
BT-CNN (Ramadan et al., 2018)	0.855
GLAD (Zhong et al., 2018)	0.881
GCE (Nouri and Hosseini-Asl, 2018)	0.885
StateNetPSI (Ren et al., 2018)	0.889
Baseline 1 (BERT+RNN)	0.892 (±0.011)
Baseline 2 (BERT+RNN+Ontology)	0.893 (±0.013)
Baseline 3 (slot-dependent SUMBT)	0.891 (±0.010)
Slot-independent SUMBT (proposed)	**0.910 (±0.010)**

Table 1. Joint Accuracy of on the evaluation dataset of WOZ 2.0 corpus

2) MultiWOZ

- Multi-domain conversation, 35 slots of 7 domains

Model	MultiWOZ		MultiWOZ (Only Restaurant)	
	Joint	Slot	Joint	Slot
MDBT (Ramadan et al., 2018)*	0.1557	0.8953	0.1789	0.5499
GLAD (Zhong et al., 2018)*	0.3557	0.9544	0.5323	0.9654
GCE (Nouri et al., 2018)*	0.3627	**0.9842**	0.6093	0.9585
TRADE (Wu et al., 2019)	**0.4862**	0.9692	0.6535	0.9328
SUMBT	0.4240 (±0.0187)	0.9599 (±0.0016)	**0.7858 (±0.0115)**	**0.9575 (±0.0025)**

Table 2. Joint Accuracy of on the evaluation dataset of MultiWOZ corpus

** The experiment results are reported in Wu et al., 2019*

3) Attention Visualization

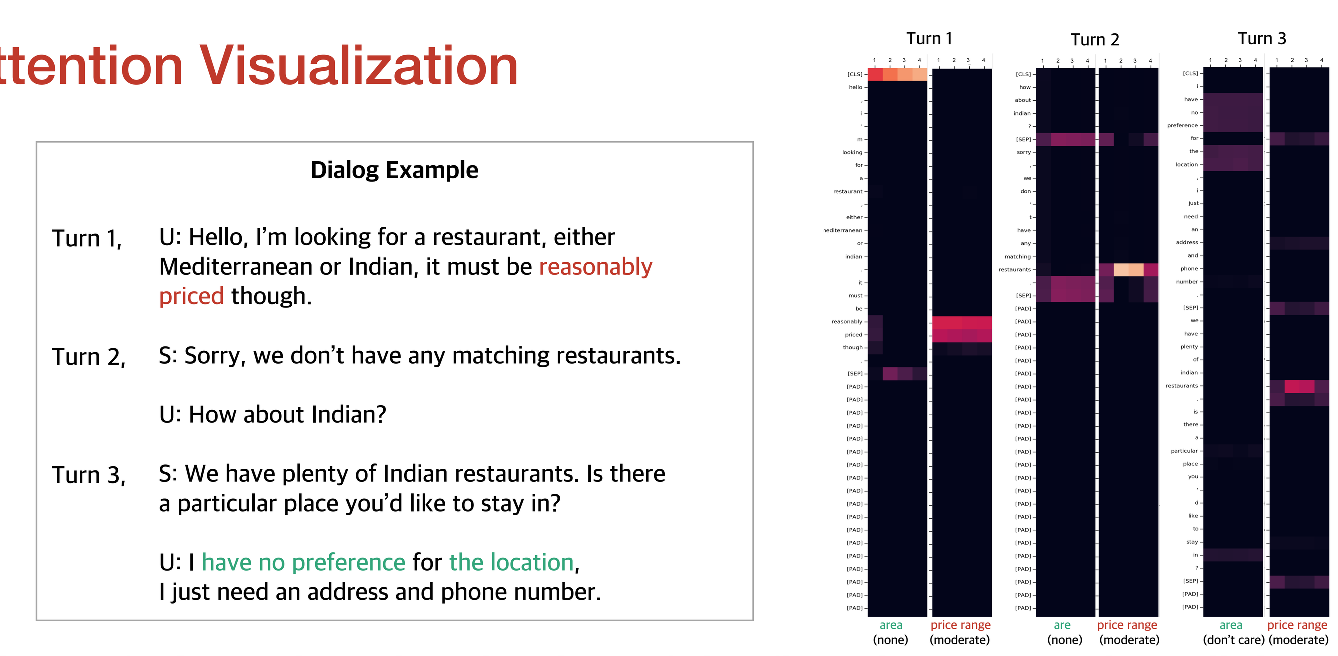

Figure 3. Attention visualzation of the first three turns in a dialog (WOZ2.0)

V. References

Nikola Mrkšić, Diarmuid Ó Séaghdha, Tsung Hsien Wen, Blaise Thomson, and Steve Young. 2017. Neural belief tracker: Data-driven dialogue state tracking. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, pages 1777–1788. Association for Computational Linguistics.

Osman Ramadan, Paweł Budzianowski, and Milica Gašić. 2018. Large-scale multi-domain belief tracking with knowledge sharing. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Short Papers), pages 432–437. Association for Computational Linguistics.

Victor Zhong, Caiming Xiong, and Richard Socher. 2018. Global-locally self-attentive dialogue state tracker. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Long Papers), pages 1458–1467. Association for Computational Linguistics.

Elnaz Nouri and Ehsan Hosseini-Asl. 2018. Toward scalable neural dialogue state tracking model. arXiv preprint, arXiv: 1812.00899.

Liliang Ren, Kaige Xie, Lu Chen, and Kai Yu. 2018. Towards universal dialogue state tracking. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 2780-2786. Association for Computational Linguistics.

Wu, C. S., Madotto, A., Hosseini-Asl, E., Xiong, C., Socher, R., & Fung, P. (2019). Transferable Multi-Domain State Generator for Task-Oriented Dialogue Systems. arXiv preprint arXiv:1905.08743.

The 57th Annual Meeting of the Association for Computational Linguistics (ACL), July 28th 2019, Florence, Italy