Toward Scalable Neural Dialogue State Tracking Model

Elnaz Nouri[1], Ehsan Hosseini-Asl[2]
[1] Microsoft Research, [2] Salesforce Research

Overview

- Dialog State Tracking (DST) is an important component of task-oriented dialogue systems which keeps track of the goal of the interaction and what has happened in the dialog.

- The latency in the current neural-based dialogue state tracking models prohibits them from being effectively deployed in production systems, albeit their highly accurate performance.

- Recently proposed Global-Local Self-Attention encoder (GLAD) [1] achieves state of arts results on WoZ and DSTC2 datasets.

- GLAD model used dedicated RNN models for each slot type inside three encoders, for user utterance, system action and slots.

- This paper proposes a new scalable and accurate neural dialogue state tracking model, by proposing a Globally Conditioned Encoder (GCE)

- The latency is improved during training and inference by **35%** on average, while preserving accuracy in predicting belief state, **88.51%** on joint goal and **97.38%** on turn request on WoZ dataset.

GLAD Model

Global-locally self-attentive encoder [1]

Proposed Globally-Conditioned Encoder (GCE) Model

- GCE employs the similar approach of learning slot-specific temporal and context representation of utterance and previous system actions

- Reducing the number of recurrent networks from (1 + #slots) to 1 in utterance, system actions and slot encoders
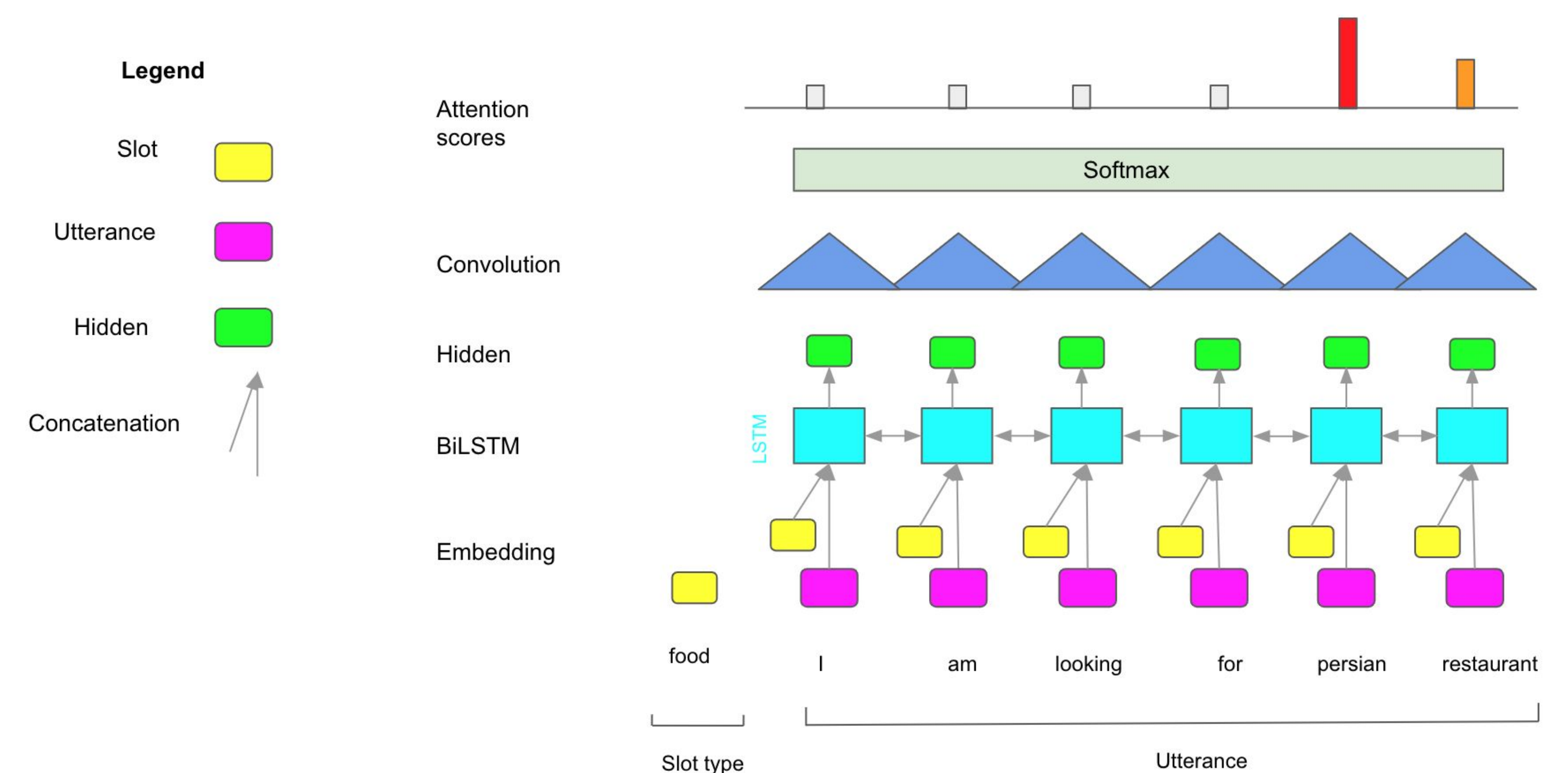
Evaluation

- The evaluation metric is based on joint goal and turn-level request and joint goal tracking accuracy. The joint goal is the accumulation of turn goals as described [1]
- Table1: Test accuracy on WoZ dataset (restaurant reservation)
- Table2: Time complexity for each batch of turn, and train and test epoch. Each batch contains 50 turns. All numbers are in second.
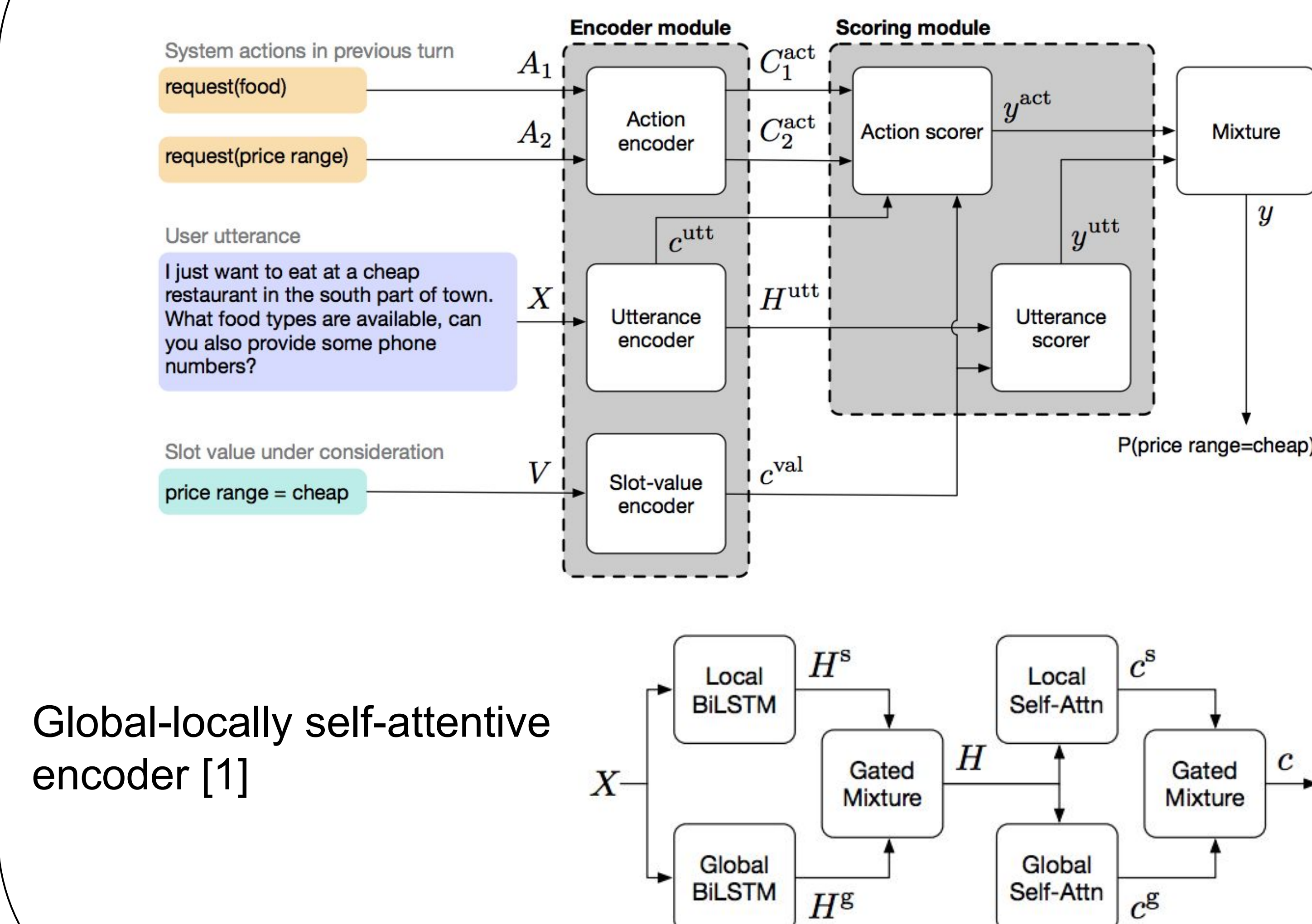
Table 1 Accuracy on WoZ		
Model	Joint Goal	Turn Request
Delex, Model [2]	70.8%	87.1%
Delex. + Semantic Dictionary [2]	83.7%	87.6%
Neural Belief Tracker-DNN [2]	84.4%	91.2%
Neural Belief Tracker - CNN[3]	84.2%	91.6%
GLAD [1]	88.1	97.1
GCE (Ours)	**88.5**	**97.38%**

WoZ				
	Train		Test	
Model	Turn	Toral	Turn	Total
GLAD [1]	1.78	89	2.32	76
GCE (Ours)	**1.16**	**60**	**1.92**	**63**

References

[1] Zhong, Victor, Caiming Xiong and Richard Socher. "Global-Locally Self-Attentive Dialogue State Tracker." in ACL (2018)
[2] Nikola Mrksic, Diarmuid Ó Séaghdha, Tsung-Hsien Wen, Blaise Thomson, and Steve J. Young. 2017. Neural belief tracker: Data-driven dialogue state tracking. In ACL
[3] Tsung-Hsien Wen, Lina Maria RojasBarahona, Milica Gasic, Nikola Mrksic, Pei hao Su, Stefan Ultes, , Steve J. Young, and David Vandyke. 2017. A network-based end-to-end trainable taskoriented dialogue system. In EACL