论文分享

胡佳颖

Large-Scale Multi-Domain Belief Tracking with Knowledge Sharing

Osman Ramadan, Paweł Budzianowski, Milica Gašić
Department of Engineering,
University of Cambridge, U.K.
{oior2,pfb30,mg436}@cam.ac.uk

ACL, 2018

Motivation

- The tasks that dialogue systems are trying to solve are becoming increasingly complex, requiring scalability to multi-domain, semantically rich dialogues.

Motivation

- Most current approaches have difficulty scaling up with domains because of the dependency of the model parameters on the dialogue ontology.

- The **larger** the ontology, the more **flexible** and multi-purposed the system is, but the **harder** it is to train and maintain a good quality BT.

Slot value数目过多，计算量大

训练数据通常不足，各个slot value不能被充分学习

Model

- **Core idea**:
  - Leverage semantic similarities between the utterances and ontology terms to compute the belief state distribution.
  - The model parameters only learn to model the interactions between turn utterances and ontology terms in the semantic space, rather than the mapping from utterances to states.

The same semantic space

Information is shared between both slots and across domains.

the number of parameters does not increase with the ontology size.

$$\mathbf{d}_k = \mathbf{h}_k^d \odot \tanh(\mathbf{W}_d\, \mathbf{e}_d + \mathbf{b}_d),$$

$$\mathcal{P}_t(d) = \sigma(\mathbf{w}_d\, \{\mathbf{d}_{usr} \oplus \mathbf{d}_{sys}\} + b_d),$$

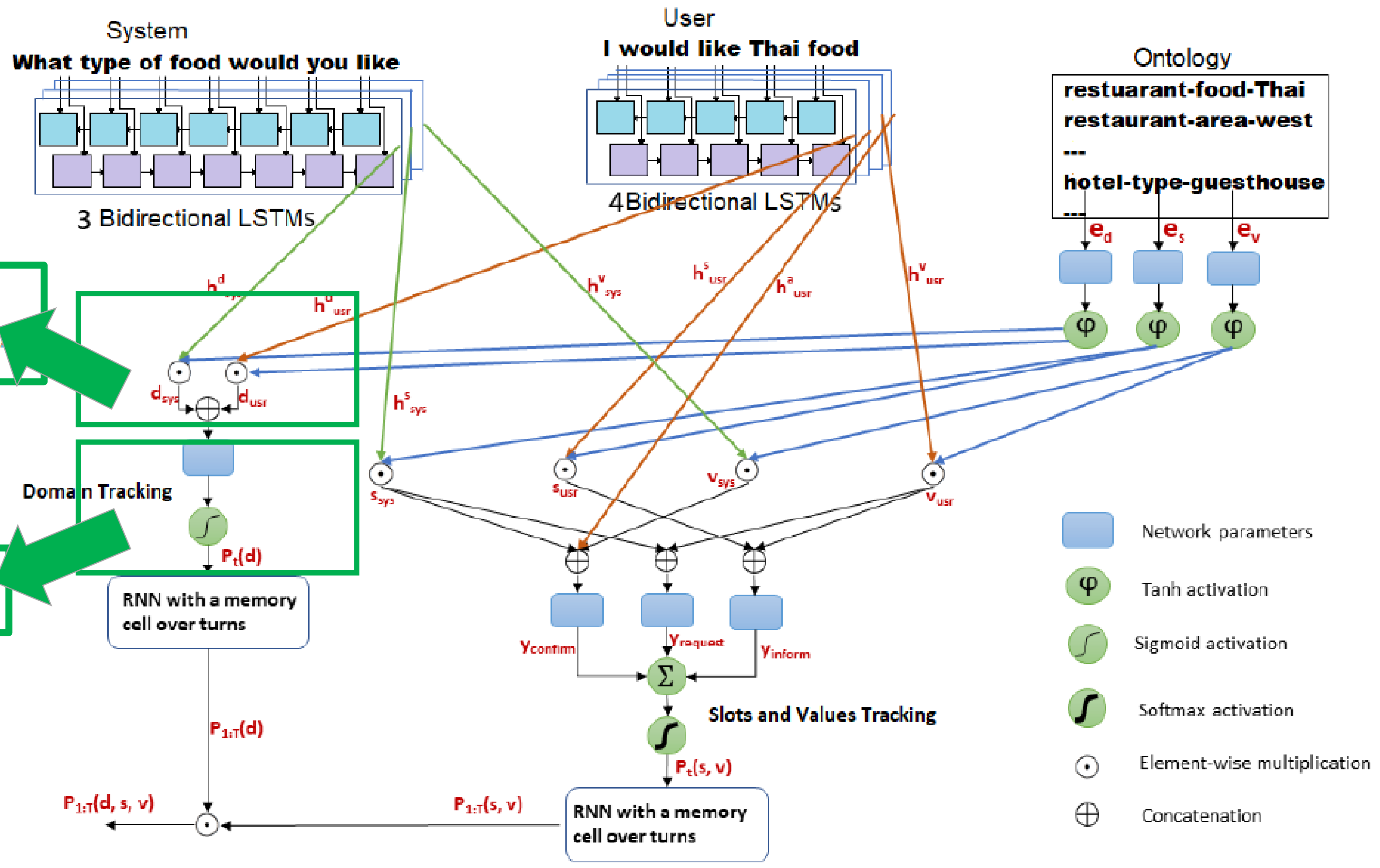

Figure 1: The proposed model architecture, using Bi-LSTMs as encoders. Other variants of the model use CNNs as feature extractors (Kim, 2014; Kalchbrenner et al., 2014).

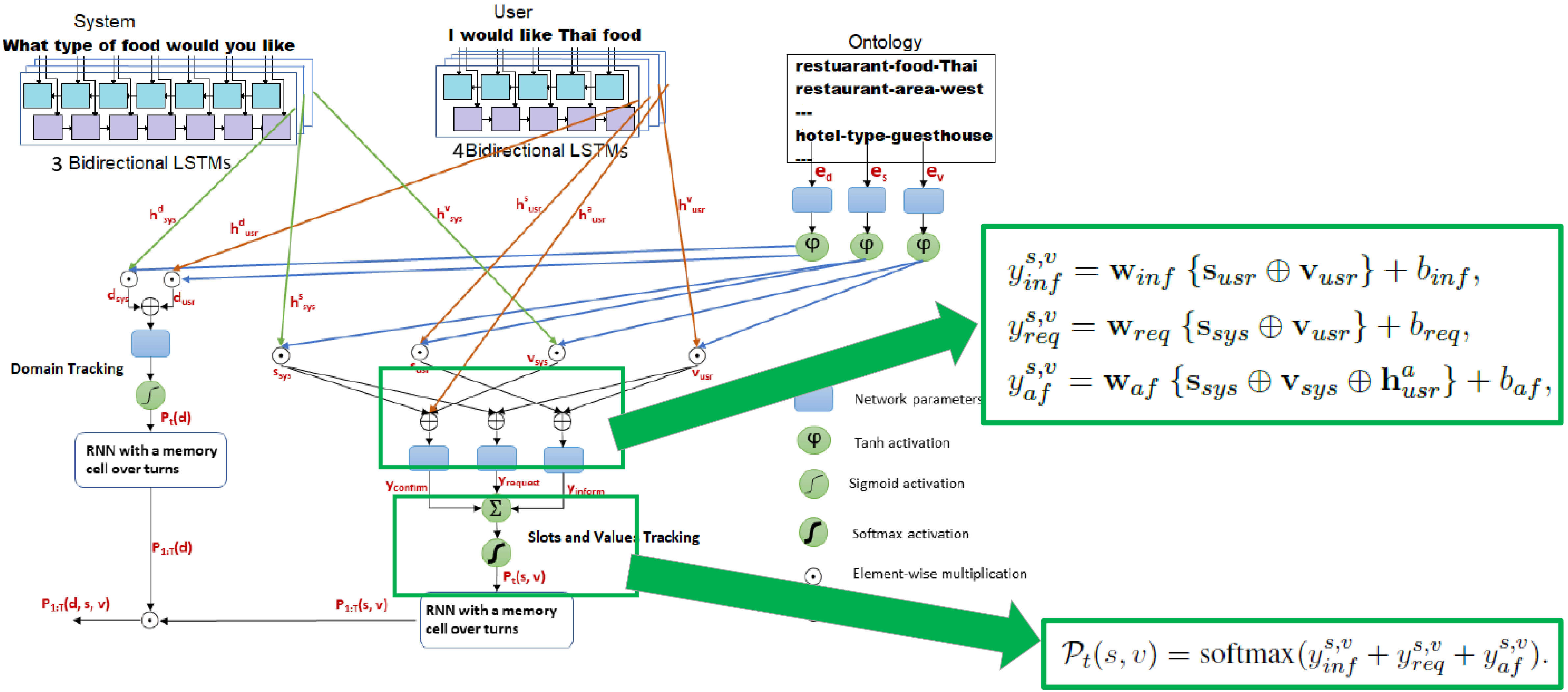

Figure 1: The proposed model architecture, using Bi-LSTMs as encoders. Other variants of the model use CNNs as feature extractors (Kim, 2014; Kalchbrenner et al., 2014).

Three kinds of information for determining the slot value

- Inform
  - The user is informing the system about his/her goal
  - “I am looking for a restaurant that serves Turkish food.”

- Request
  - The system is requesting information by asking the user about the value of a specific slot.
  - System: “When do you want the taxi to arrive?”
  - User: “19:30.”

- Confirm
  - The system wants to confirm information about the value of a specific slot.
  - System: “Would you like free parking?”
  - User: “Yes/No.”

Loss Function

The loss function for the **domain tracking** is:

$$\mathcal{L}_d = -\sum_{n=1}^{N} \sum_{\mathrm{d} \in \mathcal{D}} t^n(\mathrm{d}) \log \mathcal{P}_{1:T}^n(\mathrm{d}),$$

The loss function for the **slots and values** is:

$$\mathcal{L}_{s,v} = -\sum_{n=1}^{N} \sum_{\mathrm{s},\mathrm{v} \in \mathcal{S},\mathcal{V}} t^n(\mathrm{s},\mathrm{v}) \log \mathcal{P}_{1:T}^n(\mathrm{s},\mathrm{v}),$$

Thanks to the disjoint training, the learning of slot and value belief states are not restricted to a specific domain.

Evaluation

- Dataset:
  - **WOZ 2.0**
  - The main goal of the data collection was to acquire human-human conversations between a tourist visiting a city and a clerk from an information center.

  - **New WOZ(proposed by the authors)**

Evaluation

	WOZ 2.0			New WOZ (only restaurants)		
Slot	**NBT-CNN**	**Bi-LSTM**	**CNN**	**NBT-CNN**	**Bi-LSTM**	**CNN**
Food	88.9	96.1	**96.4**	78.3	84.7	**85.3**
Price range	93.7	**98.0**	97.9	92.6	**95.6**	93.6
Area	94.3	97.8	**98.1**	78.3	82.6	**86.4**
Joint goals	84.2	85.1	**85.5**	57.7	59.9	**63.7**

Table 1: WOZ 2.0 and new dataset test set accuracies of the NBT-CNN and the two variants of the proposed model, for slots *food, price range, area* and *joint goals*.

Evaluation

New WOZ (multi-domain)		
Model	**F1 score**	**Accuracy %**
Uniform Sampling	0.108	10.8
Bi-LSTM	0.876	**93.7**
CNN	**0.878**	93.2

Table 2: The overall F1 score and accuracy for the multi-domain dialogues test set.[4]

Thank You