

Augmenting Natural Language Generation with External Memory

Minglong Sun, Bo-Hsiang Tseng, Dr Milica Gasic

Department of Engineering, University of Cambridge

Introduction

- Natural language generation: a crucial step in the Spoken Dialogue Systems.
- Previous approaches:
 - Rule-based and Corpus-based approaches
- State-of-art RNN based generator:
 - Jointly optimize sentence planning and surface realization
 - Fewer heuristics but more natural, efficient and diverse system output
- Motivation for memory module:
 - Better capture long-term dependencies
 - Zero-shot or k-shot learning
 - Improve variability of the output

Semantically Conditioned LSTM-based generator

The semantically conditioned LSTM generator[1] is based on a recurrent NN architecture in which a 1-hot encoding \mathbf{w}_t of a token w_t is input at each time step t conditioned on a recurrent hidden layer \mathbf{h}_t and outputs the probability distribution of the next token w_{t+1} . This way, the network can produce a sequence of tokens which can be lexicalized to form the required utterance.

The LSTM architecture can be defined as below:

$$\mathbf{i}_t = \sigma(\mathbf{W}_{wi}\mathbf{w}_t + \mathbf{W}_{hi}\mathbf{h}_{t-1}) \quad (1)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_{wf}\mathbf{w}_t + \mathbf{W}_{hf}\mathbf{h}_{t-1}) \quad (2)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_{wo}\mathbf{w}_t + \mathbf{W}_{ho}\mathbf{h}_{t-1}) \quad (3)$$

$$\hat{\mathbf{c}}_t = \tanh(\mathbf{W}_{wc}\mathbf{w}_t + \mathbf{W}_{hc}\mathbf{h}_{t-1}) \quad (4)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \hat{\mathbf{c}}_t \quad (5)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \quad (6)$$

The generator introduces a sentence planning cell based on a sigmoid control gate and a dialogue act(DA):

$$\mathbf{r}_t = \sigma(\mathbf{W}_{wr}\mathbf{w}_t + \alpha\mathbf{W}_{hr}\mathbf{h}_{t-1}) \quad (7)$$

$$\mathbf{d}_t = \mathbf{r}_t \odot \mathbf{d}_{t-1} \quad (8)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \hat{\mathbf{c}}_t + \tanh(\mathbf{W}_{dc}\mathbf{d}_t) \quad (9)$$

Semantically Conditioned LSTM

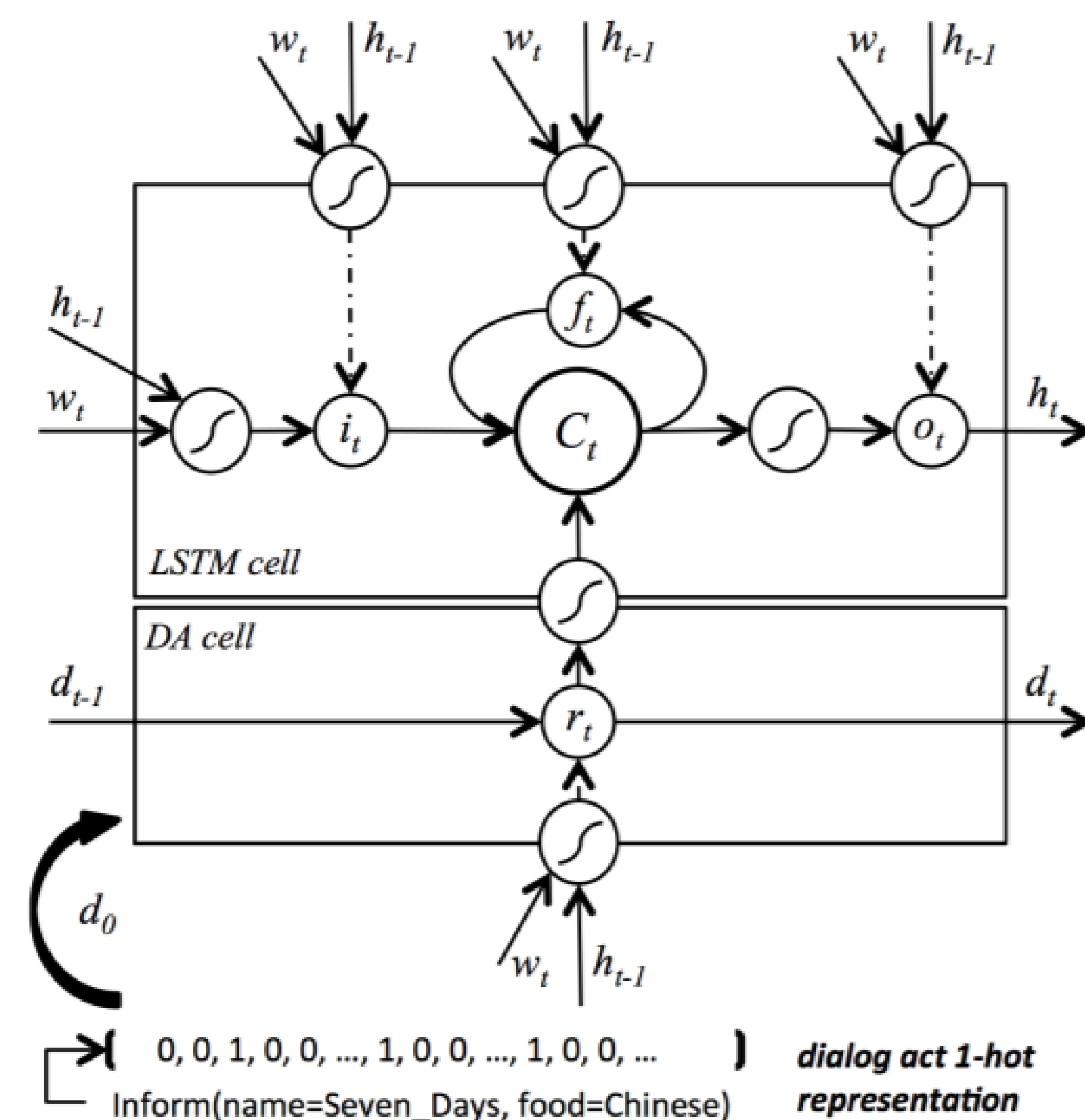


Figure: Semantically conditioned LSTM generator[1]

Using external memory to augment generation models

Using external memory module has been studied extensively recently. Memory in the neural network has the capability to better monitor long-term dependencies and to store useful information from previous training examples. The interaction of the memory module with the environment involves read and write operations which are based on an associated attention mechanism. A few things can be considered when designing the memory module:

Ways to address the memory:

- Content-based addressing: access memories based on the similarity to a given cue
- Location-based addressing: access memories based on their position
- Combination of both

Initialization of the memory:

- Learned memory: treat memory as model parameters, initialized randomly
- Fixed memory: memory is seen as the database of additional information

Soft/Hard attention mechanisms:

- Make the memory module fully-differentiable or not

Which layer to use memory and how to use the output from the memory module

Conditional VAE

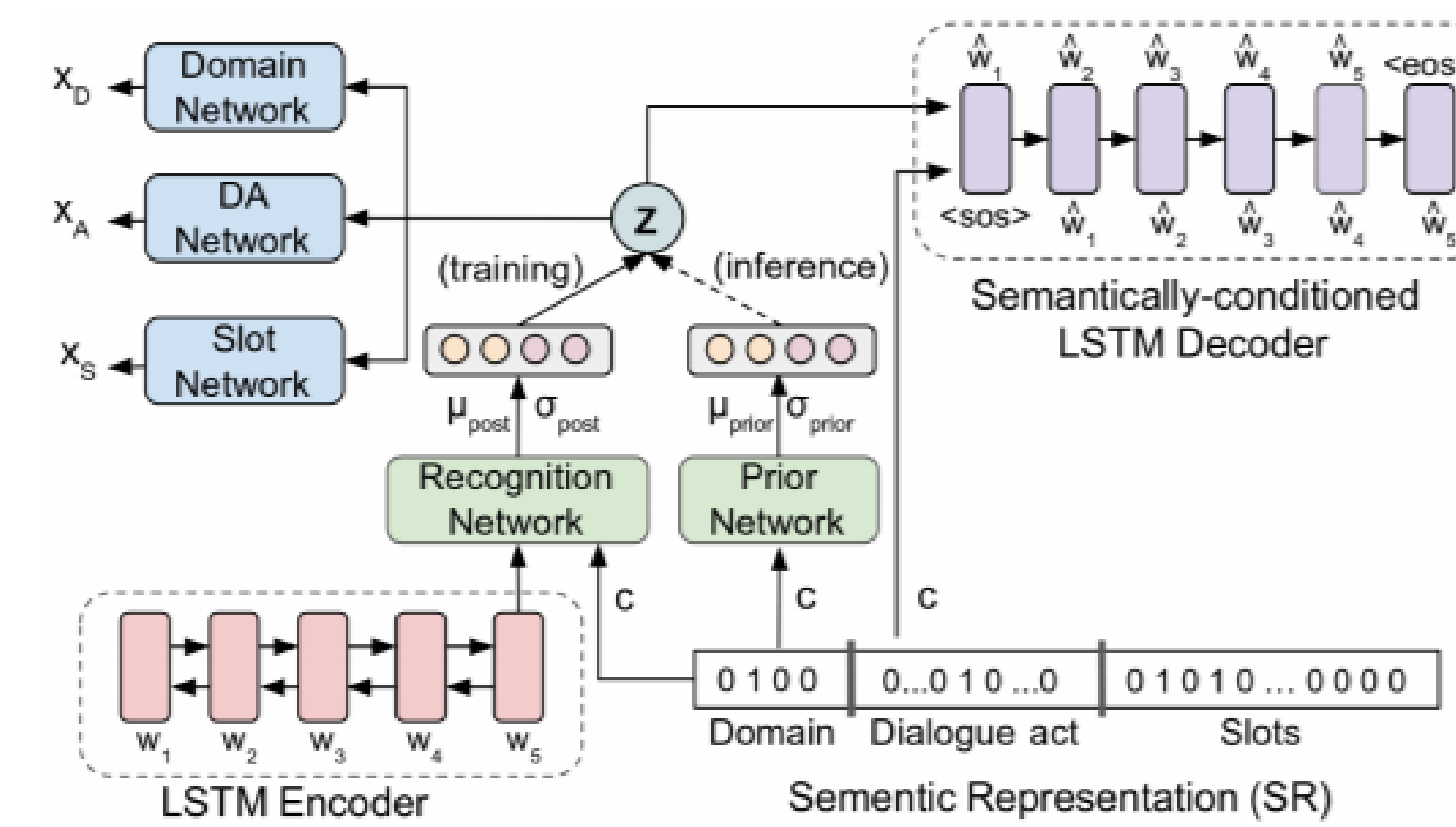


Figure: Variational auto-encoder with SC-LSTM[2]

By combining the Variational auto-encoder with the Semantically Conditioned LSTM[2], we can improve the generation of the sentences. Semantically conditioned VAE encodes the latent information from both sentences and semantic representation and feeds latent variable z into the SCLSTM decoder as a better initialization of decoder.

Some comparisons and results

The metrics used to evaluate system performance in our task are Slot Error Rates, BLEU-4 scores and perplexity. Averaged results over 5 random seeds for Restaurant Domain are summarized as below:

| Model Type | Slot Error(%) | BLEU-4 |
|------------------------|---------------|--------|
| Vanilla LSTM | 12.21 | 0.525 |
| SC-LSTM | 5.49 | 0.537 |
| CVAE+SCLSTM | 4.86 | 0.539 |
| Memory module for CVAE | 4.38 | 0.546 |

Table: Averaged Results for different generation models

The bottom model architecture uses an external memory module to augment the Conditional VAE. The memory is initialized randomly and uses soft content-based attention mechanisms.

Future Experiments

- Different memory mechanisms
 - Key-value style memory module
 - memory module as trainable parameters (a combination of memory keys and values)
- Using dynamic external memory module augmenting LSTM decoder

References

- [1] Tsung-Hsien Wen, Milica Gasic, Nikola Mrksic, Pei-Hao Su, David Vandyke, and Steve Young. Semantically conditioned lstm-based natural language generation for spoken dialogue systems. *arXiv preprint arXiv:1508.01745*, 2015.
- [2] Bo-Hsiang Tseng, Florian Kreyszig, Pawel Budzianowski, Inigo Casanueva, Yen-Chen Wu, Stefan Ultes, and Milica Gasic. Variational cross-domain natural language generation for spoken dialogue systems.