SEntNet: Source-aware Recurrent Entity Network for Dialogue Response Selection

Jiahuan Pei[1], Arent Stienstra[1], Julia Kiseleva[2], Maarten de Rijke[1]

[1] University of Amsterdam
[2] Microsoft Research AI

August 12, 2019

Overview

1. Introduction
   - 1.1 Task-oriented Dialogue Systems (TDSs)
   - 1.2 Dialogue Response Selection in TDSs
   - 1.3 Motivation
2. Method
3. Experimental setup
4. Results
5. Conclusion & Future work

1.1 Task-oriented Dialogue Systems (TDSs)

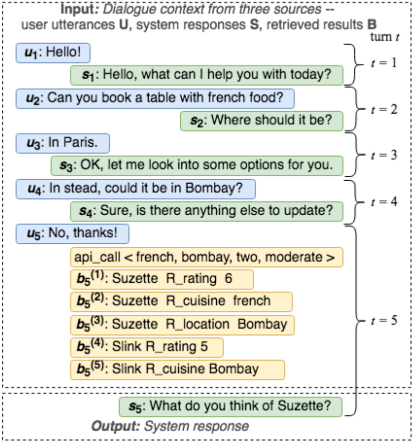
Task-oriented DSs → complete certain tasks or goals on specific domain (e.g., *finding restaurants*)

User: I'm looking for an affordable restaurant
System: How about Thai food?
User: Yes please, in central Cambridge
System: The House serves cheap Thai food
User: Where is it?
System: The House is at 106 Regent Street

Non-task-oriented DSs → provide reasonable responses & entertainment on open domain (e.g., chatbots)

User: Hey, do you know Monica?
System: Yes, I found her via AM leaks. She has the same hobbies as I and loves to meet in secluded area within.
User: Seriously?
System: Nope.
...

1.2 Dialogue Response Selection in TDSs



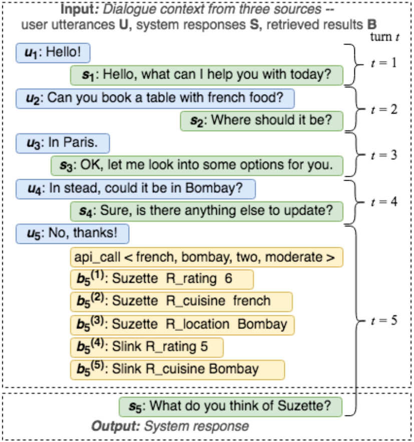
Traditionally,

- **Given**: a dialogue context $\mathbb{C}_t = (u_1, s_1, \ldots, u_t, s_t, [b_t^1, b_t^2, \ldots, b_t^\lambda])$
- **Goal**: select a response s_t from candidates by

$$\psi_\Theta(\mathbb{C}_t) \rightarrow s_t. \tag{1}$$

- **Problem**. Obtaining the important information from a complex, long dialogue context is challenging.

1.3 Motivation

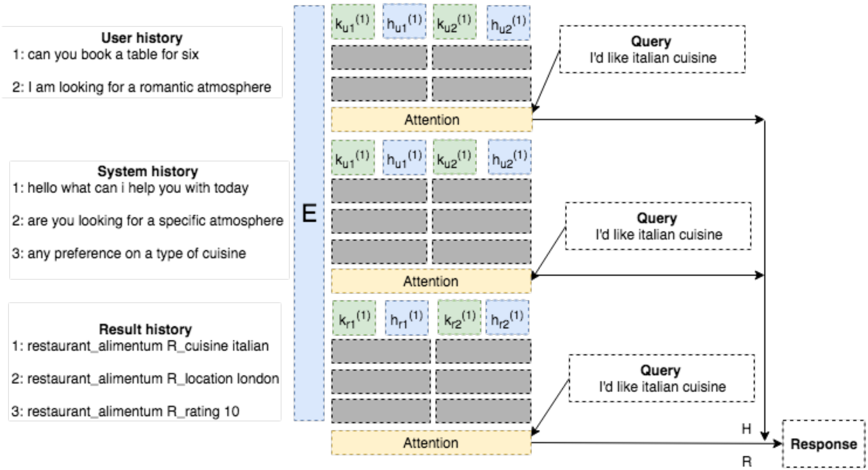


- **Given**: a dialogue context $(\mathbb{U}_t, \mathbb{S}_{t-1}, \mathbb{B}_t)$:
  - $\mathbb{U}_t = (u_1,\ u_2,\ \ldots,\ u_t)$ are user utterances;
  - $\mathbb{S}_{t-1} = (s_1,\ s_2,\ \ldots,\ s_{t-1})$ are system responses; and
  - $\mathbb{B}_t = (b_t^1, b_t^2, \ldots, b_t^\lambda)$ is λ-best retrieved results from an external knowledge base (KB).
- **Goal**:

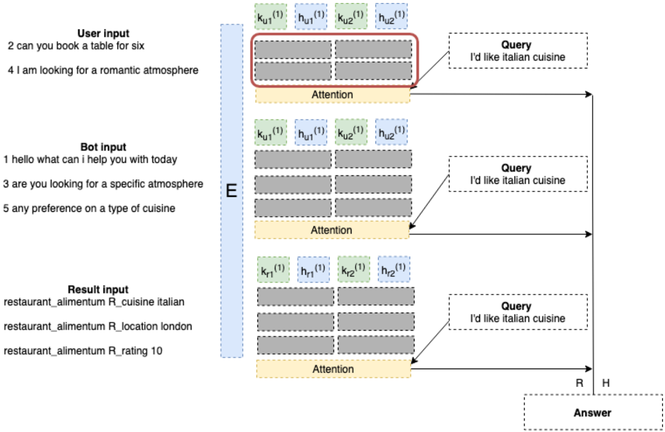
$$\psi_\Theta(\mathbb{U}_t, \mathbb{S}_{t-1}, \mathbb{B}_t) \rightarrow s_t. \qquad (2)$$

- **Solution**. Source-specific memories for different usage of words and syntactic structure.

2.1 Source-aware Recurrent Entity Network (SEntNet)



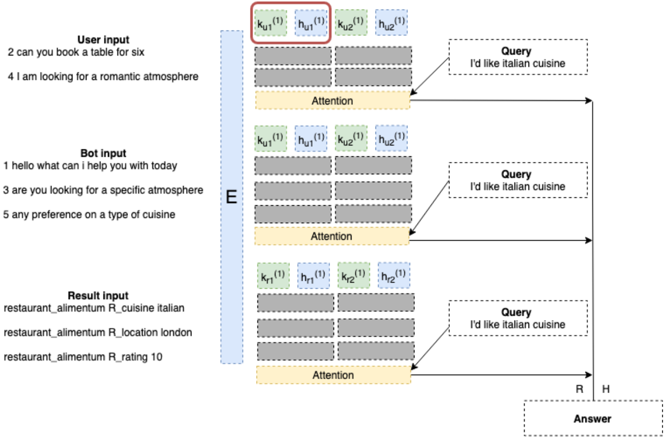
2.2 SEntNet – Input module



- The embedding of the i-th utterance $e_{i(\mathcal{S})}$ for source $\mathcal{S}$ is:

$$e_{i(\mathcal{S})} = \Sigma_x f_x \odot w_x^i + l_x^i \in \mathbb{R}^d. \tag{3}$$

2.2 SEntNet – Dynamic memory module (1)

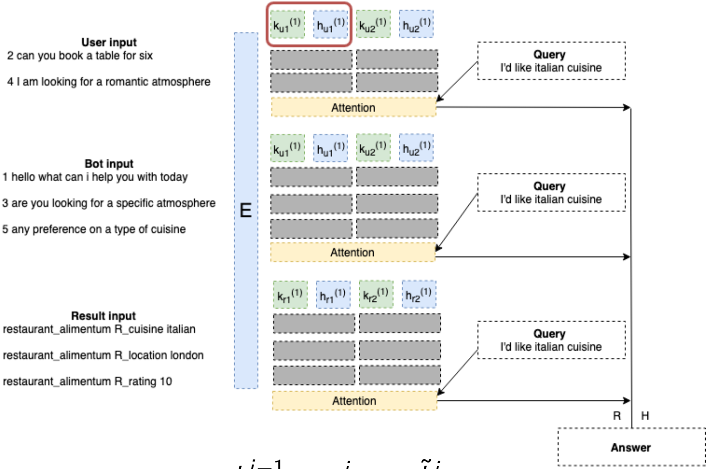


- For the i-th utterance from $\mathcal{S}$ in the dialogue, the memory block for the j-th entity is updated as:

$$g^i_{j(\mathcal{S})} = \sigma(e^T_{i(\mathcal{S})} h^{i-1}_{j(\mathcal{S})} + e^T_{i(\mathcal{S})} k^{i-1}_{j(\mathcal{S})}) \in \mathbb{R}^d \quad (4)$$

$$\tilde{h}^i_{j(\mathcal{S})} = \phi(G_{\mathcal{S}} h^{i-1}_{j(\mathcal{S})} + V_{\mathcal{S}} k^{i-1}_{j(\mathcal{S})} + W_{\mathcal{S}} e_{i(\mathcal{S})}) \in \mathbb{R}^d \quad (5)$$

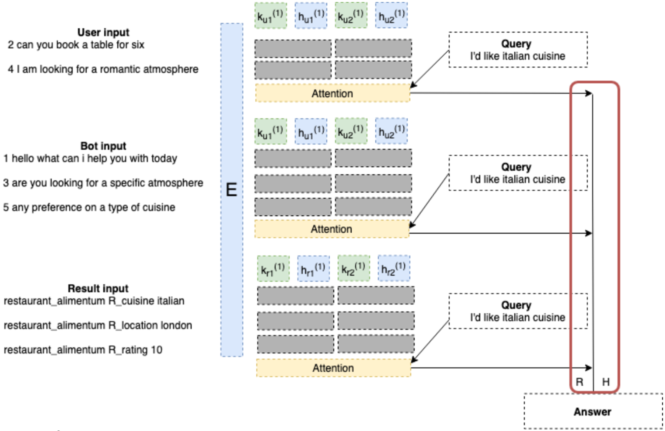
2.2 SEntNet – Dynamic memory module (2)



$$h^i_{j(\mathcal{S})} = \frac{h^{i-1}_{j(\mathcal{S})} + g^i_{j(\mathcal{S})} \odot \tilde{h}^i_{j(\mathcal{S})}}{\|h^{i-1}_{j(\mathcal{S})} + g^i_{j(\mathcal{S})} \odot \tilde{h}^i_{j(\mathcal{S})}\|} \in \mathbb{R}^d \tag{6}$$

$$h_{j(\mathcal{S})} = h^1_{j(\mathcal{S})} \oplus h^2_{j(\mathcal{S})} \oplus \cdots \oplus h^n_{j(\mathcal{S})}. \tag{7}$$

2.3 SEntNet – Output module (1)

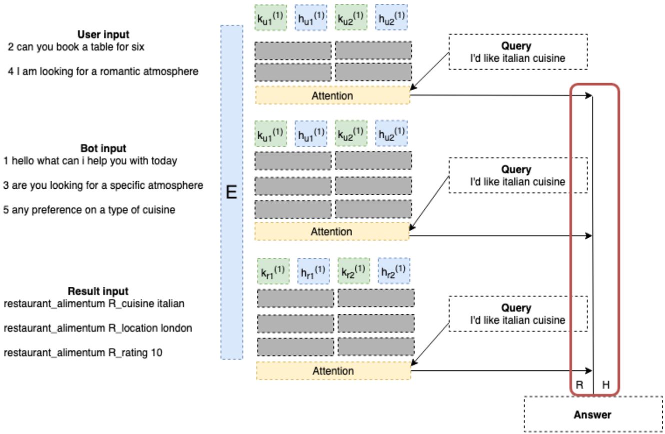


- Let $q \in \mathbb{R}^d$ be the embedding of the user utterance u_t for the current turn t. The output module is defined as:

$$p_{j(\mathcal{S})} = \mathrm{softmax}(q^T h_{j(\mathcal{S})}) \tag{8}$$

$$z_{\mathcal{S}} = \Sigma_j h_{j(\mathcal{S})} p_{j(\mathcal{S})} \in \mathbb{R}^d \tag{9}$$

2.3 SEntNet – Output module (2)



$$z = z_{\mathcal{S}_{\mathbb{U}}} \oplus z_{\mathcal{S}_{\mathbb{S}}} \oplus z_{\mathcal{S}_{\mathbb{B}}} \in \mathbb{R}^{3d} \tag{10}$$

$$y = L\phi(q + Hz) \in \mathbb{R}^{r} \tag{11}$$

$$y = \mathrm{softmax}(\tilde{y}_j). \tag{12}$$

3.1 Experimental setup: Datasets & Evaluation

- **Datasets**.
  - Dialog bAbI (Bordes&Weston,2017)
  - DSTC2 (Henderson et al.,2014).

Table: Statistics of the two datasets

	# dialogues	# words	# responses	Partitioning
bAbI	3,000	3,747	4,212	1000/1000/1000
DSTC2	2,785	1,229	2,406	1,168/500/1,117

- **Evaluation**. Turn-level accuracy – the fraction of correct responses out of all.

3.2 Experimental setup: Baselines

- **TF-IDF**. This model ranks candidate responses by TF-IDF weighted cosine similarity between one-hot vectors of input and candidate responses.
- **Query-to-answer (Q2A)**. Given a query, it finds the most common response in the train set (Weston et al., 2015).
- **DQMemNN**. This is the state-of-the-art for response selection on dialog bAbI dataset (Wu et al., 2018); for a fair comparison, we used DQMemNN without exact matching and delexicalization.
- **HHCN**. This is the state-of-the-art for response selection on the DSTC2 dataset (Liang and Yang, 2018).
- **EntNet**. We reproduced EntNet, which was originally introduced for question answering and is reported to have strong reasoning abilities (Henaff et al., 2017).

4 Results

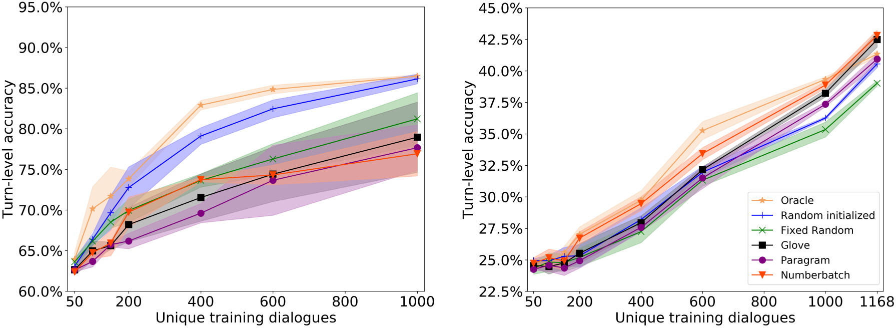
RQ1: How well does SEntNet predict appropriate responses?

Model	bAbI	DSTC2
TF-IDF	0.040	0.030
Q2A	0.570	0.220
EntNet	0.850	0.388
DQMemNN	0.863	–
HHCN	–	**0.661**
SEntNet	**0.910**	0.412

Table: Comparison with baselines on the bAbI and DSTC2 datasets.

4 Results

RQ2: How do different embeddings affect SEntNet's performance?



(a) bAbI

(b) DSTC2

Figure: Turn-level accuracy of SEntNet for different embedding spaces on bAbI and DSTC2 datasets. (Please note that the scales on the x-axes and y-axes differ.)

4 Results

RQ3: How well does SEntNet perform in the case of limited data?

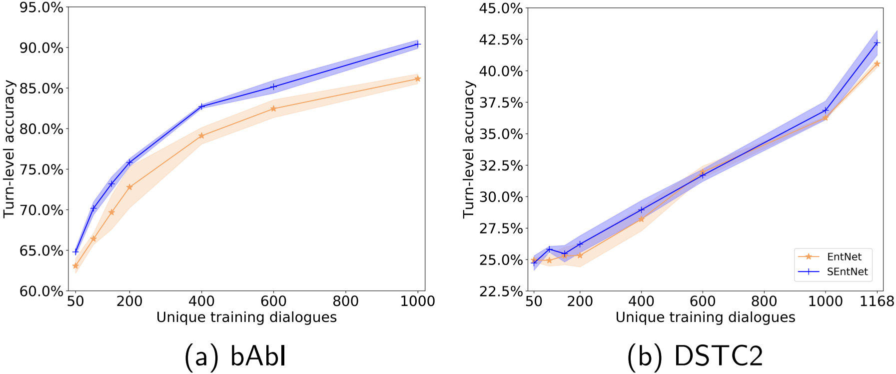


Figure: Turn-level accuracy of SEntNet on both datasets, when trained with different volumes of training dialogues. (Please note that the scales on the x-axes and y-axes differ.)

5 Conclusion & Future work

We propose **SEntNet**, a dialogue response selection model in memory network architecture:

- Select responses aware of source-specific history and consistently outperforms the baselines for end-to-end TDSs.
- Optimizing embeddings while training is useful for the performance.
- Tolerant of sparse data and able to handle different degrees of lexical diversity.
- Increase of learnable parameters by introducing extra memory modules can be addressed with parallel update mechanism design inherited from EntNet.

In the future work, we plan to apply the source-aware context idea that underlies SEntNet to other variant memory networks.

Thanks for your attention!
Q&A

Acknowledgments. This research was partially supported by Ahold Delhaize, the Association of Universities in the Netherlands (VSNU), the China Scholarship Council (CSC), and the Innovation Center for Artificial Intelligence (ICAI). We would like to thank Huawei, Microsoft, Naver and Google for their generous travel support.