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

Jiahuan Pei\textsuperscript{1}, Arent Stienstra\textsuperscript{1}, Julia Kiseleva\textsuperscript{2}, Maarten de Rijke\textsuperscript{1}

\textsuperscript{1} University of Amsterdam
\textsuperscript{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 \( C_t = (u_1, s_1, \ldots, u_t, s_t, b_1^t, b_2^t, \ldots, b_\lambda^t) \)

- **Goal**: select a response \( s_t \) from candidates by

\[
\psi_\Theta(C_t) \rightarrow s_t. \tag{1}
\]

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

- **Given**: a dialogue context $(U_t, S_{t-1}, B_t)$:
  - $U_t = (u_1, u_2, \ldots, u_t)$ are user utterances;
  - $S_{t-1} = (s_1, s_2, \ldots, s_{t-1})$ are system responses; and
  - $B_t = (b_1^t, b_2^t, \ldots, b_\lambda^t)$ is $\lambda$-best retrieved results from an external knowledge base (KB).

- **Goal**:
  \[ \psi_\Theta(U_t, S_{t-1}, B_t) \rightarrow s_t. \quad (2) \]

- **Solution**. Source-specific memories for different usage of words and syntactic structure.
2.1 Source-aware Recurrent Entity Network (SEntNet)

User history
1: can you book a table for six
2: I am looking for a romantic atmosphere

System history
1: hello what can i help you with today
2: are you looking for a specific atmosphere
3: any preference on a type of cuisine

Result history
1: restaurant_alimentum R_cuisine italian
2: restaurant_alimentum R_location london
3: restaurant_alimentum R_rating 10

Query
I'd like italian cuisine

E

Attention

Query
I'd like italian cuisine

H

Response
2.2 SEntNet – Input module

The embedding of the \( i \)-th utterance \( e_i(S) \) for source \( S \) is:

\[
e_i(S) = \sum_x f_x \odot w_x^i + l_x^i \in \mathbb{R}^d.
\]
For the $i$-th utterance from $S$ in the dialogue, the memory block for the $j$-th entity is updated as:

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

$$\tilde{h}^i_{j(S)} = \phi(G_S h^{i-1}_{j(S)} + V_S k^{i-1}_{j(S)} + W_S e_i(S)) \in \mathbb{R}^d$$ (5)
2.2 SEntNet – Dynamic memory module (2)

User input
2 can you book a table for six
4 I am looking for a romantic atmosphere

Bot input
1 hello what can i help you with today
3 are you looking for a specific atmosphere
5 any preference on a type of cuisine

Result input
restaurant_alimentum R_cuisine italian
restaurant_alimentum R_location london
restaurant_alimentum R_rating 10

\[
\begin{align*}
    h^i_j(S) &= \frac{h^{i-1}_j(S) + g^i_j(S) \odot \tilde{h}^i_j(S)}{\|h^{i-1}_j(S) + g^i_j(S) \odot \tilde{h}^i_j(S)\|} \in \mathbb{R}^d \\
    h_j(S) &= h^1_j(S) \oplus h^2_j(S) \oplus \cdots \oplus h^n_j(S).
\end{align*}
\]
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:

$$p_j(S) = \text{softmax}(q^T h_j(S)) \quad (8)$$

$$z_S = \Sigma_j h_j(S)p_j(S) \in \mathbb{R}^d \quad (9)$$
2.3 SEntNet – Output module (2)

$$z = z_{S_U} \oplus z_{S_S} \oplus z_{S_B} \in \mathbb{R}^{3d}$$  \hspace{1cm} (10)

$$y = L_\phi(q + Hz) \in \mathbb{R}^r$$  \hspace{1cm} (11)

$$y = \text{softmax}(\tilde{y}_j).$$  \hspace{1cm} (12)
3.1 Experimental setup: Datasets & Evaluation

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

<table>
<thead>
<tr>
<th></th>
<th># dialogues</th>
<th># words</th>
<th># responses</th>
<th>Partitioning</th>
</tr>
</thead>
<tbody>
<tr>
<td>bAbI</td>
<td>3,000</td>
<td>3,747</td>
<td>4,212</td>
<td>1000/1000/1000</td>
</tr>
<tr>
<td>DSTC2</td>
<td>2,785</td>
<td>1,229</td>
<td>2,406</td>
<td>1,168/500/1,117</td>
</tr>
</tbody>
</table>

- **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?

<table>
<thead>
<tr>
<th>Model</th>
<th>bAbl</th>
<th>DSTC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF-IDF</td>
<td>0.040</td>
<td>0.030</td>
</tr>
<tr>
<td>Q2A</td>
<td>0.570</td>
<td>0.220</td>
</tr>
<tr>
<td>EntNet</td>
<td>0.850</td>
<td>0.388</td>
</tr>
<tr>
<td>DQMemNN</td>
<td>0.863</td>
<td>–</td>
</tr>
<tr>
<td>HHCN</td>
<td>–</td>
<td>0.661</td>
</tr>
<tr>
<td>SEntNet</td>
<td><strong>0.910</strong></td>
<td>0.412</td>
</tr>
</tbody>
</table>

**Table**: Comparison with baselines on the bAbl and DSTC2 datasets.
4 Results

**RQ2**: How do different embeddings affect SEntNet’s performance?

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?

![Graphs showing turn-level accuracy of SEntNet on both datasets.](image)

*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.