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

Overview

- **Goal.** Select an appropriate response from candidates given a dialogue context for Task-oriented Dialogue Systems (TDSs).
- **Problem.** Obtaining key information from a complex, long dialogue context is challenging, especially when different sources of information are available.
- **Solution.** Employ source-specific memories to exploit differences in the usage of words and syntactic structure from different information sources, i.e., user, system, and knowledge base (KB).

System Response Selection in TDSs

- **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
  - \(B_t = \{b_1, b_2, \ldots, b_t\}\) is \(\lambda\)-best retrieved results from an external KB.
- **Goal:** select a response \(s_t\) from candidates by
  \[
  \nu_{\theta}(U_t, S_{t-1}, B_t) \rightarrow s_t \tag{1}
  \]

Figure: An example of response selection for booking a restaurant. The top box contains the input for response selection; the bottom box shows the selected response.

Source-aware Recurrent Entity Network (SEntNet)

- **Input module.** The embedding of the \(i\)-th utterance \(e_{ui}\) for source \(S\) is:
  \[
  e_{ui} = \sum_j f_j \odot w_{ij}^e + b_i^e \in \mathbb{R}^{d_e} \tag{2}
  \]
- **Dynamic memory module.** For the \(i\)-th utterance from \(S\) in the dialogue, the memory block for the \(j\)-th entity is updated as:
  \[
  g_{ij} = \sigma(e_{ui} h_{ij}^g + e_{uj} h_{ij}^{g_{uj}}) \in \mathbb{R}^{d_g} \tag{3}
  \]
  \[
  h_{ij}^g = \sigma(g_{ij} h_{ij} + V_h \delta_{ij} + W_{e_{ui}} e_{ui} + W_{e_{uj}} e_{uj}) \in \mathbb{R}^{d_h} \tag{4}
  \]
  \[
  h_{ij} = [h_{ij}^g \odot \hat{h}_{ij}^g] \in \mathbb{R}^{d_h} \tag{5}
  \]
  \[
  h_{ij} = \hat{h}_{ij}^g \odot \hat{h}_{ij}^g \odot \cdots \odot \hat{h}_{ij}^g \in \mathbb{R}^{d_h} \tag{6}
  \]
- **Output module.** Let \(q \in \mathbb{R}^{d_h}\) be the embedding of the user utterance \(u_t\) for the current turn \(t\). The output module is defined as:
  \[
  p_{1:t} = \text{softmax}(q^T h_{ij}) \tag{7}
  \]
  \[
  z_s = \sum_j h_{ij}^g p_{ij} \in \mathbb{R}^{d_h} \tag{8}
  \]
  \[
  z = z_b \odot z_g \odot z_b \in \mathbb{R}^{d_h} \tag{9}
  \]
  \[
  y = L_0(q + z \odot z) \in \mathbb{R}^{d_h} \tag{10}
  \]
  \[
  y = \text{softmax}(y) \tag{11}
  \]

Figure: Schematic representation of SEntNet architecture with separate source-specific memory modules.

SEntNet’s functions depend on three modules described below.

Experimental Setup

- **Research questions.**
  - RQ1: How well does SEntNet predict appropriate responses?
  - RQ2: How do different embeddings affect SEntNet’s performance?
  - RQ3: How well does SEntNet perform in the case of limited data? And
  - RQ4: How does lexical diversity affect SEntNet’s performance?

- **Datasets.** Dialog bAbI (Bordes&Weston, 2017); DSTC2 (Henderson et al., 2014).
- **Evaluation.** Turn-level accuracy – the fraction of correct responses out of all.

Results

<table>
<thead>
<tr>
<th>Model</th>
<th>bAbI</th>
<th>DSTC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>TF-IDF</td>
<td>0.400</td>
<td>0.300</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>0.661</td>
</tr>
<tr>
<td>HHCN</td>
<td></td>
<td>0.661</td>
</tr>
<tr>
<td>SEntNet</td>
<td>0.910</td>
<td>0.412</td>
</tr>
<tr>
<td>SEntNet+POS</td>
<td>0.890</td>
<td>0.409</td>
</tr>
</tbody>
</table>

Table: Comparison with baselines on the bAbI and DSTC2 datasets (RQ2).

Figure: Turn-level accuracy of SEntNet for different embedding spaces on both datasets. (RQ2).

Figure: Turn-level accuracy of SEntNet on both datasets, when trained with different volumes of training dialogues (RQ3).

Conclusion

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.

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.