END-TO-END DYNAMIC QUERY MEMORY NETWORK FOR ENTITY-VALUE INDEPENDENT TASK-ORIENTED DIALOG

Chien-Sheng Wu, Andrea Madotto, Genta Winata, Pascale Fung
Human Language Technology Center, Department of Electronic & Computer Engineering
Hong Kong University of Science and Technology

Introduction

- Task-oriented dialog systems require machines to understand the human intent and generate proper responses for an assigned task via natural language.
- Traditionally, these dialog systems have been built as a pipeline, with modules for language understanding, state tracking, action selection, and language generation. However, modeling the dependencies between modules is complex and expensive.
- Recently, end-to-end approaches using recurrent neural networks [1, 2, 3] are attractive solutions. In particular, end-to-end memory networks (MemNNs), which are recurrent attention models over a possibly large external memory [4], can achieve promising results.

Contribution

In this paper, we propose an end-to-end Dynamic Query Memory Network (DQMemNN) for task-oriented dialog systems with a Recorded Delexicalization (RDL) mechanism.

1. DQMemNN is designed to address one major drawback of MemNN: be insensitive to represented temporal dependencies between memories, which may influence the conversational semantics.
2. RDL is performed to reduce the learning complexity and alleviate the out-of-vocabulary (OOV) entity problems.
3. Our framework achieves a promising average per-response accuracy of 99.7% and per-dialog accuracy of 97.8% on bAbI Dialog dataset.

Dialog Example

Usr: Hi, what can I help you with today?
Sys: Chinese cuisine in a moderate price range for four
Usr: <Silence>
Sys: Where should it be?
Usr: Paris, please
Sys: Ok let me look into some options for you
Usr: <Silence>
Sys: api_call Chinese Paris four moderate
Usr: Instead could it be in Taiwan?
Sys: Sure is there anything else to update?
Usr: No
Sys: Ok let me look into some options for you
Usr: <Silence>
Sys: api_call Chinese Taiwan four moderate
KB1: DimSum, rating ..., address ..., phone ..., ...
KB2: DinTaiFung, rating ..., address ..., phone ..., ...
KB3: SiChuan, rating ..., address ..., phone ..., ...
Usr: <Silence>
Sys: What do you think of this option: DinTaiFung?
Usr: <Silence>
Sys: Let’s do it
Usr: Great let me do the reservation
Usr: <Silence>
Sys: Can you provide the address?
Usr: Here it is 101_Taipei_Street
Usr: Thank you
Sys: You are welcome

Recorded Delexicalization

The intuition of RDL is to replace entities in the raw dialog history with a simplified ordered template. Moreover, the number of system utterances can be reduced to a smaller size of action template candidates.

<table>
<thead>
<tr>
<th>Task</th>
<th>Test Avg.</th>
<th>MemNN</th>
<th>DQMemNN</th>
<th>HCN</th>
<th>DQMemNN+RDL</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>99.7 (97.8)</td>
<td>99.7 (98.9)</td>
<td>100 (100)</td>
<td>99.7 (98.9)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>T2</td>
<td>99.5 (99.5)</td>
<td>100 (100)</td>
<td>100 (100)</td>
<td>99.5 (99.5)</td>
<td>100 (100)</td>
</tr>
<tr>
<td>T3</td>
<td>74.9 (74.9)</td>
<td>74.9 (74.9)</td>
<td>74.9 (74.9)</td>
<td>74.9 (74.9)</td>
<td>74.9 (74.9)</td>
</tr>
<tr>
<td>T4</td>
<td>57.2 (57.2)</td>
<td>57.2 (57.2)</td>
<td>57.2 (57.2)</td>
<td>57.2 (57.2)</td>
<td>57.2 (57.2)</td>
</tr>
<tr>
<td>T5</td>
<td>99.6 (52.5)</td>
<td>96.1 (49.4)</td>
<td>96.3 (52.5)</td>
<td>99.2 (88.7)</td>
<td>99.9 (98.3)</td>
</tr>
<tr>
<td>Test Avg.</td>
<td>86.1 (58.1)</td>
<td>86.1 (58.1)</td>
<td>85.7 (50.5)</td>
<td>86.3 (58.1)</td>
<td>99.7 (98.7)</td>
</tr>
</tbody>
</table>

References


Framework Description

The bAbI dialog dataset [2] is used to evaluate the performance of our model.
- Task 1-5 are issuing API call, refining API call, recommending options, providing additional information, and full-dialog, respectively.
- One test set follows the same distribution as the training set, and another has OOV words from a different knowledge base.
- All experiments used a 3-hop model with the adjacent weight sharing scheme. During training, all embedding matrices are jointly learned by minimizing a standard cross-entropy loss.

Results and Visualization