DeepSQA: Understanding Sensor Data via a Deep Learning Approach to Sensory Question Answering

Tianwei Xing, Marc Roig Vilamala, Luis Garcia, Federico Cerutti, Lance Kaplan, Alun Preece and Mani Srivastava

MOTIVATION
- Current machine learning systems can only perform one or a few limited tasks on sensory data.
- Introducing new tasks requires re-training of the system.
- User cannot get other information except for predefined high-level labels.

SQA DATASET GENERATION
- SQA dataset are generated automatically using a template-based approach. Each element in SQA has its own semantic representation.

Semantic Representation: Scene List $S = \{A_1, A_2, ... A_n\}$, $A_i = \{t_i, d_i, s_i\}$

Semantic Representation: Functional Program
What did the user do [before] <opening the fridge> and [after] <closing the drawer>? query_action_type( AND( relate( before, open the fridge ), relate( after, close the drawer )))

Answer obtained from question engine: Apply the functional program above on the scene list

PRELIMINARY RESULT
- Evaluate SQA models on OppQA dataset:
  - Source dataset: OPPORTUNITY, ~8h, 4 users, 24 runs
  - Question family: 16, (query, existence, counting, comparison...) 110+ templates,
  - Avg length: 21 words per question
  - Answers: 28 unique answers for classification task

<table>
<thead>
<tr>
<th>Question Types</th>
<th>Prior</th>
<th>Prior-Q</th>
<th>Neuro-Symbolic</th>
<th>CNN</th>
<th>LSTM</th>
<th>CRNN</th>
<th>SAN</th>
<th>DeepSQA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>44.18</td>
<td>55.30</td>
<td>45.61</td>
<td>44.23</td>
<td>63.81</td>
<td>72.09</td>
<td>57.16</td>
<td>74.77</td>
</tr>
<tr>
<td>Existence</td>
<td>60.08</td>
<td>41.69</td>
<td>42.62</td>
<td>59.88</td>
<td>62.77</td>
<td>73.12</td>
<td>62.13</td>
<td>72.02</td>
</tr>
<tr>
<td>Counting</td>
<td>0</td>
<td>55.89</td>
<td>28.40</td>
<td>0</td>
<td>57.30</td>
<td>65.07</td>
<td>65.96</td>
<td>67.56</td>
</tr>
<tr>
<td>Query</td>
<td>0</td>
<td>5.67</td>
<td>0</td>
<td>0</td>
<td>40.25</td>
<td>46.13</td>
<td>39.43</td>
<td>52.77</td>
</tr>
<tr>
<td>Compare</td>
<td>50.87</td>
<td>75.27</td>
<td>66.43</td>
<td>59.81</td>
<td>72.15</td>
<td>76.34</td>
<td>71.27</td>
<td>79.75</td>
</tr>
<tr>
<td>Compare-act</td>
<td>45.45</td>
<td>45.45</td>
<td>0</td>
<td>46.17</td>
<td>56.10</td>
<td>50.00</td>
<td>53.18</td>
<td>75.00</td>
</tr>
</tbody>
</table>

- Adopt a representative subset of methods from VQA task:
  - Prior and Prior-Q: predict answer based on training answer statistics
  - LSTM/ CNN : baselines using only question/ sensory data as input
  - CRNN: LSTM + CNN with easy information fusion before classification
  - SAN: Stacked attention network
  - DeepSQA: proposed method using compositional attention model.

- DeepSQA model shows better performance than other baselines

FUTURE WORK
- Study the SQA model robustness to linguistic variations, and train robust SQA model using the cycle-consistency framework.
- Investigate the neural-symbolic approach for sensory QA, explicitly parse the input question and construct task-dependent model during the runtime.

RELATED WORK