Convolutional Neural Encoder for the 7th Dialogue System Technology Challenge

Mandy Korpusik, James Glass
MIT Computer Science and Artificial Intelligence Lab, Cambridge MA USA
(korpusik, glass)@mit.edu

2. DSTC7 Track 1

Goal: select the next best system response.

Focused on the Advising “Flex” dataset, subtask 1: select the next system response from among a list of 100 candidate responses.

Advisors: Hello Mingyang! Are you doing well?
Student: Hi advisor. I’m doing alright. I would like some advice on which courses to take next semester.
Advisor: My interested area is Software Development and Artificial Intelligence.
Student: How many different levels do these classes have?
Advisor: EECS381 is not easy.
Advisor: any thoughts about that?
Student: What time does the course occur?
I like afternoon classes and will find something else if it’s scheduled too early.
Advisor: EECS351 is after lunch. The others are before. EECS481 is from nine to ten thirty and EECS492 is from ten thirty to twelve.

Given the conversation history, the system must select the correct next response: “481 is the early morning and is quite similar to EECS381, so you might want to skip it.”

Students at the University of Michigan role played the student and advisor, recording 500 conversations for training data:

<table>
<thead>
<tr>
<th>min</th>
<th>max</th>
<th>mean</th>
<th>median</th>
</tr>
</thead>
<tbody>
<tr>
<td>History length</td>
<td>1</td>
<td>41</td>
<td>9.2</td>
</tr>
<tr>
<td>Utterance length</td>
<td>1</td>
<td>384</td>
<td>10.3</td>
</tr>
<tr>
<td>Candidate answer length</td>
<td>1</td>
<td>384</td>
<td>12.4</td>
</tr>
</tbody>
</table>

3. Neural Models

Baseline Dual LSTM Encoder:

Convolutional Neural Encoder:

Binary Verification: 1 (Match) / 0 (Not)

Convolutional (CNN) Ensemble: summed the ranked indices of seven randomly initialized models.

4. Validation Results

<table>
<thead>
<tr>
<th>Model</th>
<th>R@1</th>
<th>R@10</th>
<th>R@50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dual LSTM Encoder</td>
<td>6.20</td>
<td>36.0</td>
<td>80.0</td>
</tr>
<tr>
<td>Single CNN (meanpool)</td>
<td>3.43</td>
<td>27.2</td>
<td>97.9</td>
</tr>
<tr>
<td>Single CNN (maxpool)</td>
<td>10.9</td>
<td>46.3</td>
<td>97.2</td>
</tr>
<tr>
<td>2-CNN Ensemble</td>
<td>11.1</td>
<td>48.0</td>
<td>97.2</td>
</tr>
<tr>
<td>3-CNN Ensemble</td>
<td>11.8</td>
<td>47.3</td>
<td>97.4</td>
</tr>
<tr>
<td>4-CNN Ensemble</td>
<td>12.0</td>
<td>46.7</td>
<td>97.0</td>
</tr>
<tr>
<td>5-CNN Ensemble</td>
<td>12.2</td>
<td>46.3</td>
<td>97.0</td>
</tr>
<tr>
<td>6-CNN Ensemble</td>
<td>12.4</td>
<td>46.9</td>
<td>97.0</td>
</tr>
<tr>
<td>7-CNN Ensemble</td>
<td>12.6</td>
<td>46.9</td>
<td>97.0</td>
</tr>
<tr>
<td>8-CNN Ensemble</td>
<td>12.4</td>
<td>46.9</td>
<td>97.0</td>
</tr>
<tr>
<td>Single CNN (Glove)</td>
<td>12.2</td>
<td>50.3</td>
<td>98.3</td>
</tr>
<tr>
<td>Single CNN (word2vec)</td>
<td>14.8</td>
<td>47.5</td>
<td>97.0</td>
</tr>
</tbody>
</table>

5. Test Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Data</th>
<th>R@1</th>
<th>R@10</th>
<th>R@50</th>
<th>MRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN Ensemble</td>
<td>Test 1</td>
<td>20.6</td>
<td>54.8</td>
<td>82.4</td>
<td>32.3</td>
</tr>
<tr>
<td>CNN Ensemble</td>
<td>Test 2</td>
<td>8.8</td>
<td>32.0</td>
<td>72.8</td>
<td>16.9</td>
</tr>
</tbody>
</table>

6. Analysis

Errors due to unknown words and courses:

Student: does eecs370 have <-UNK>?
Correct: no, it’s not in eecs370.
Top-4 Predicted:
1) yes
2) doesn’t have a good score
3) no
4) i think i might take eecs370

Student: many thanks, and i hope to never see you again.
Correct: i wish for you that you graduate quickly.
Top-4 Predicted:
1) you’re welcome
2) i believe that is a good decision.
3) i wish you luck.
4) what time of day do you need it?

Student: do they pair well together?
Correct: it depends on your schedule. eecs370 is an easy class whereas, according to statistics, eecs376 is an easy class.
Top-4 Predicted:
1) yes
2) they do not like eecs370
3) i suggest strongly you take eecs280
4) you have met all the requirements, so it is possible.

Student: as a professor how is he?
Correct: if you want, you can see for yourself by registering for the class.
Top-4 Predicted:
1) well
2) that depends
3) 270 has a score of <-UNK> for being relatively <-UNK>.
4) it has a high degree of easiness, helpfulness, and clarity.
5) it’s not a very heavy load

CNN filters learn to focus on different entities:

4. Validation Results

7. Future Work

Leverage more contextual information from the full dialogue history and knowledge base.

Character embeddings for unknown words.

Apply the models to the Ubuntu dataset and other subtasks.