An Analysis of Prerequisite Skills for Reading Comprehension

Saku Sugawara (University of Tokyo), Akiko Aizawa (National Institute of Informatics, Japan)
sakus@is.s.u-tokyo.ac.jp

Overview
- Challenge: Evaluation method for reading comprehension (RC)
- Our goal: Construct a general evaluation methodology that decomposes the RC process and elucidates:
  * the detailed performance of RC systems
  * the characteristics of RC tasks from multiple points of view: prerequisite skills.
- Our approach:
  1. Define a set of prerequisite skills that are required for understanding documents
  2. Annotate questions of an RC task with the skills
  3. Analyze the performances of RC systems for the annotated questions to grasp the differences and limitations of their individual performances

Prerequisite Skills and Annotation
- 10 prerequisite basic skills were defined by investigating NLU tasks (WSC, COPA, CoNLL 15st, bAbI, and so on).
- We manually annotated questions with the RC skills that are required to answer each question (multi-labeling).
- We assume that when RC systems use RC skill, they already have the capability to recognize the facts described in the clauses that the skill pertained to.
- An example requires no skills:
  - Context: Todd lived in a town.
  - Question: Where did Todd live? — Answer: In a town

Reading Comprehension Skills and Annotation Results: accuracies and frequencies in RC tasks

<table>
<thead>
<tr>
<th>RC skills (*: “understanding of”)</th>
<th>MCTest accuracy</th>
<th>SQuAD frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>List/Enumeration</td>
<td>51.1% 65.1% 61.9% 14.7% 5.0%</td>
<td>0.9% 0.0%</td>
</tr>
<tr>
<td>Mathematical operations</td>
<td>20.0% 30.0% 30.0% 1.6% 0.0%</td>
<td>0.9% 0.0%</td>
</tr>
<tr>
<td>Coreference resolution</td>
<td>52.5% 63.6% 62.1% 63.8% 6.2%</td>
<td>0.0% 0.0%</td>
</tr>
<tr>
<td>Logical reasoning</td>
<td>100.0% 75.0% 66.7% 0.9% 0.0%</td>
<td>0.0% 0.0%</td>
</tr>
<tr>
<td>Analogical reasoning</td>
<td>- - - 0.0% 0.0%</td>
<td>0.0% 0.0%</td>
</tr>
<tr>
<td>Spatiotemporal relations*</td>
<td>48.9% 66.9% 67.1% 27.5% 2.5%</td>
<td>0.0% 0.0%</td>
</tr>
<tr>
<td>Causal relations*</td>
<td>45.7% 62.0% 60.9% 14.4% 6.2%</td>
<td>0.0% 0.0%</td>
</tr>
<tr>
<td>Commonsense reasoning*</td>
<td>44.0% 61.3% 59.6% 41.9% 86.2%</td>
<td>20.0% 25.0%</td>
</tr>
<tr>
<td>Complex sentences*</td>
<td>50.0% 65.9% 64.0% 20.6% 20.0%</td>
<td>0.0% 0.0%</td>
</tr>
<tr>
<td>Special sentence structure*</td>
<td>46.2% 69.2% 73.1% 8.1% 25.0%</td>
<td>0.0% 0.0%</td>
</tr>
<tr>
<td>(Overall accuracy)</td>
<td>50.9% 66.2% 65.9% 20.6% 20.0%</td>
<td>0.0% 0.0%</td>
</tr>
</tbody>
</table>

Description or Examples

- Annotate systems and datasets

Challenge
- Evaluation method for reading comprehension (RC)
- Our goal: Construct a general evaluation methodology that decomposes the RC process and elucidates:
  * the detailed performance of RC systems
  * the characteristics of RC tasks from multiple points of view: prerequisite skills.

Our approach
1. Define a set of prerequisite skills that are required for understanding documents
2. Annotate questions of an RC task with the skills
3. Analyze the performances of RC systems for the annotated questions to grasp the differences and limitations of their individual performances

System analysis
A. All systems are still not good at coreference resolution and commonsense reasoning (ideally the weakness is derived from the difference between accuracies of skill combinations...)
B. We could not observe that adding RTE significantly increased accuracy by small annotations :(

Dataset analysis
C. These scores reflect the difficulty of the datasets (SQuAD: Wikipedia (for adults); MCTest: tales (for children))
D. SQuAD has simple questions

Annotate example in MCTest (required 5 skills)

ID: MC160.dev.29 (1) multiple:
C1: The princess climbed out the window of the high tower and climbed down the south wall when her mother was sleeping.
C2: She wandered out a good ways.
C3: Finally she went into the forest where there are no electric poles but where there are some caves.
Q: Where did the princess wander to after escaping?
A: Forest

Analyze systems and conclusions
- Need more systems, datasets, and annotations!!
- Are the skills sufficient? e.g. commonsense reasoning...