**Objective**

To develop a QA system that can answer a subset of 4th grade questions involving recognizing instances of physical, biological, and other natural processes.

The questions present a short description of an instance and multiple process names as the answer choices.

**Question:** As water vapor rises in the atmosphere, it cools and changes back to liquid. Tiny drops of liquid form clouds in this process called

**Answer Choices:**
- condensation
- evaporation
- precipitation
- run-off

**Approach**

This work explores a knowledge-driven approach to answering such questions.

- We represent processes using a light-weight semantic role based representation.
- We answer a question by assessing how well the roles of the instance in the question align with the roles of the candidate answer processes.

**Question Answering Using Semantic Roles**

We score each candidate answer process based on how well the roles of the instance described in the question align with the roles of the process. We use the following procedure to answer questions:

1. **Identify the roles in the question statement** ($Q$).
2. **Collect the roles for all the answer processes** ($A^N_m$) from the Knowledge Base. ($M$ – # of answer choices, $N$ – # of frames)
3. **For every QA frame pair**, compute an alignment score by checking for the textual entailment of the corresponding roles.

   \[
   \text{alignment}(Q, A^N_m) = \sum_{\text{roles}_i} \text{entails}(\text{role}_i(A^N_m), \text{role}_i(Q))
   \]

   where, $R$ = (Input, Result, Enabler, Trigger). $\text{entails}(x, y)$ is computed as a textual entailment score that reflects how well the text $x$ entails text $y$ or the other way around.

4. **Compute the mean of the top 5 frame alignment scores** for each process and return the top scoring process as the answer.

**Results**

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>0.4323</td>
<td>0.3325</td>
<td>0.3758</td>
</tr>
<tr>
<td>Per Process</td>
<td>0.4225</td>
<td>0.2556</td>
<td>0.3185</td>
</tr>
<tr>
<td>Distant Supervision</td>
<td>0.5614</td>
<td>0.2642</td>
<td>0.3594</td>
</tr>
<tr>
<td>Dom. Adaptation</td>
<td>0.4386</td>
<td>0.3351</td>
<td>0.3799</td>
</tr>
</tbody>
</table>

**Error Analysis**

**Automatic SRL Failures**
- Issues that arise out of data sparsity
- Gap between verb-based role and our customized process-based role

**QA Failures**
- Knowledge Representation Issues (37%) 
- Entailment Issues (32%)
- Scoring Issues (31%)

**References**


**Acknowledgements**

This work is funded in part by Fulbright PhD Fellowship and by the Allen Institute for Artificial Intelligence. The findings and conclusions expressed herein are those of the authors alone.