Learning to Describe Solutions for Bug Reports Based on Developer Discussions

Sheena Panthaplackel(spantha@cs.utexas.edu), Junyi Jessy Li, Milos Gligoric, Raymond J. Mooney
The University of Texas at Austin

Motivation
- When a bug is reported, developers engage in a dialogue to collaboratively understand it and ultimately resolve it with the necessary code changes.
- To expedite bug resolution, we aim to guide developers in better absorbing content relevant towards implementing the solution from long bug report discussions.

Task
Generate concise natural language description of the solution by synthesizing relevant content in the discussion when it emerges in real-time

Data
~12.3K bug reports for open-source Java projects that are linked to a commit/PR

After filtering noisy examples with generic descriptions, uninformative descriptions, and discussions with insufficient context: ~5.9K

Benchmarking Generating Solution Descriptions

Automated Metrics

<table>
<thead>
<tr>
<th>Model</th>
<th>BLEU</th>
<th>METER</th>
<th>ROUGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copy Title</td>
<td>14.4</td>
<td>12.62</td>
<td>14.2</td>
</tr>
<tr>
<td>Seq2Seq + Pointer</td>
<td>16.6</td>
<td>14.2</td>
<td>14.5</td>
</tr>
<tr>
<td>Hierarchical Seq2Seq + Pointer</td>
<td>24.4</td>
<td>25.0</td>
<td>24.1</td>
</tr>
<tr>
<td>PLBART (Full)</td>
<td>28.3</td>
<td>25.1</td>
<td>28.3</td>
</tr>
<tr>
<td>PLBART (Filtered)</td>
<td>10.0</td>
<td>9.9</td>
<td>12.3</td>
</tr>
</tbody>
</table>

Human Evaluation
- Annotators select the most informative model prediction(s)
- 160 annotations from 8 annotators
- Selected "All options are poor" 20.0%–22.1% of the time
- Selected "Insufficient context" 25.6%–31.9% of the time

Takeaway #1: PLBART outperforms baselines for generating solution descriptions.
Takeaway #2: Fine-tuning PLBART on filtered data leads to more informative solution descriptions.

Open Challenges:
- Disentangling content related to the solution from elaboration of the problem
- Synthesizing content when information related to the solution spans multiple utterances
- Bimodal reasoning across language and in-lined code

Generating Solution Descriptions in Real-Time

Title and Utterances #1…k → PLBART Encoder → Classification Head → Pipelined System

Enough context to generate at time step k?

Trained Model for Generating Solution Descriptions

Generated solution description

Jointly Trained System

Takeaway #3: When sufficient context is available, system output is useful.
Takeaway #4: Balancing the trade-off between generating too early and deferring to later time steps is an open challenge.