Retrieve, Rerank and Rewrite: Soft Template Based Neural Summarization

Ziqiang Cao\textsuperscript{1}  Wenjie Li\textsuperscript{1}  Furu Wei\textsuperscript{2}  Sujian Li\textsuperscript{3}

\textsuperscript{1}Department of Computing, The Hong Kong Polytechnic University
\textsuperscript{2}Microsoft Research Asia
\textsuperscript{3}Key Laboratory of Computational Linguistics, Peking University

July 16, 2018
Outline

1. Introduction
2. Method
3. Experiments
4. Conclusion
Sentence Summarization

Definition
- Generate a shorter version of a given sentence
- Preserve its original meaning

Usage
- Design or refine appealing headlines
Seq2seq Summarization

- Require less human efforts
- Achieve the state-of-the-art performance
Problems of Seq2seq Summarization

Solely depend on the source text to generate summaries

- Encounter error propagation
- Lose control
  - 3% of summaries ≤ 3 words
  - 4 summaries repeat a word for 99 times
  - Focus on extraction rather than abstraction
### Template-based Summarization

- A traditional approach to abstractive summarization
- Fill an incomplete with the input text using the manually defined rules
- Be able to produce fluent and informative summaries

<table>
<thead>
<tr>
<th>Template</th>
<th>[REGION] shares [open/close] [NUMBER] percent [lower/higher]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>hong kong shares closed down #.# percent on friday due to an absence of buyers and fresh incentives.</td>
</tr>
<tr>
<td>Summary</td>
<td>hong kong shares close #.# percent lower</td>
</tr>
</tbody>
</table>
Problems of Template-based Summarization

- Template construction is extremely time-consuming and requires a plenty of domain knowledge
- It is impossible to develop all templates for summaries in various domains
Use actual summaries in the training datasets as **soft templates** to combine seq2seq and template-based summarization

**Seq2seq** Guide the generation of seq2seq

**Template-based** Automatically learn to rewrite from soft templates
Proposed Method

Re³Sum: consists of three modules: Retrieve, Rerank and Rewrite.

- Use Information Retrieval to find out candidate soft templates from the training dataset (Retrieve).
- Extend the seq2seq model to jointly learn template saliency measurement (Rerank) and final summary generation (Rewrite)
Contributions

1. Introduce soft templates to improve the readability and stability in seq2seq
2. Extend seq2seq to conduct template reranking and template-aware summary generation simultaneously
3. Fuse the IR-based ranking technique and seq2seq-based generation technique, utilizing both supervisions
4. Demonstrate potential to generate diversely
Outline

1. Introduction
2. Method
3. Experiments
4. Conclusion
**Flow Chat**

**Retrieve**  Search actual summaries as candidate soft templates

**Rerank**  Find out the most proper soft template from the candidates

**Rewrite**  Generate the summary based on source sentence and soft template
Assumption: Similar sentences, similar summary patterns

**Input**  A sentence

**Platform**  LUCENE

**Output**  30 actual summaries in the training dataset whose sources are the most similar to the input sentence
Jointly Rerank and Rewrite

Share encoders

Saliency  Bilinear  Rerank

Decoder  Summary  Rewrite

Sentence  Template
Rerank

- Retrieve ranks templates according to the text similarity between sentences
- Rerank finds out the soft template most similar to the actual output summary

Model: Bilinear network

\[ s(r, x) = \text{sigmoid}(h_r W_s h_x^T + b_s) \]
A soft template accords with the facts in the input sentences
Use Seq2seq to generate more faithful and informative summaries

Concatenate the encoders of sentence and template

\[ H_c = [h_1^x; \cdots ; h_{-1}; h_1^r; \cdots ; h_{-1}] \]

Use attentive RNN decoder to generate summaries

\[ s_t = \text{Att-RNN}(s_{t-1}, y_{t-1}, H_c), \]
Learning

- Cross Entropy (CE) for Rerank
- Negative Log-Likelihood (NLL) for Rewrite
- Add the above two costs as the final loss

\[ J_R(\theta) = \text{CE}(s(r, x), s^*(r, y^*)) \]
\[ = -s^* \log s - (1 - s^*) \log(1 - s) \]
\[ J_G(\theta) = -\log(p(y^* | x, r)) \]
\[ = -\sum_t \log(p_t[y_t^*]) \]
\[ J(\theta) = J_R(\theta) + J_G(\theta) \]
**Dataset**  Gigaword (sentence, headline) pairs

**Framework**  OpenNMT

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Train</th>
<th>Dev.</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>3.8M</td>
<td>189k</td>
<td>1951</td>
</tr>
<tr>
<td>AvgSourceLen</td>
<td>31.4</td>
<td>31.7</td>
<td>29.7</td>
</tr>
<tr>
<td>AvgTargetLen</td>
<td>8.3</td>
<td>8.3</td>
<td>8.8</td>
</tr>
<tr>
<td>COPY(%)</td>
<td>45</td>
<td>46</td>
<td>36</td>
</tr>
</tbody>
</table>
Re³Sum significantly outperforms other approaches

<table>
<thead>
<tr>
<th>Model</th>
<th>ROUGE-1</th>
<th>ROUGE-2</th>
<th>ROUGE-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABS†</td>
<td>29.55*</td>
<td>11.32*</td>
<td>26.42*</td>
</tr>
<tr>
<td>ABS+†</td>
<td>29.78*</td>
<td>11.89*</td>
<td>26.97*</td>
</tr>
<tr>
<td>Featseq2seq†</td>
<td>32.67*</td>
<td>15.59*</td>
<td>30.64*</td>
</tr>
<tr>
<td>RAS-Elman†</td>
<td>33.78*</td>
<td>15.97*</td>
<td>31.15*</td>
</tr>
<tr>
<td>Luong-NMT†</td>
<td>33.10*</td>
<td>14.45*</td>
<td>30.71*</td>
</tr>
<tr>
<td>OpenNMT</td>
<td>35.01*</td>
<td>16.55*</td>
<td>32.42*</td>
</tr>
<tr>
<td>Re³Sum</td>
<td>37.04</td>
<td>19.03</td>
<td>34.46</td>
</tr>
</tbody>
</table>
Linguistic Quality Performance

- Low LEN_DIF and LESS_3 $\rightarrow$ Stable
- Low COPY $\rightarrow$ Abstractive
- Low NEW_NE and NEW_UP $\rightarrow$ Faithful

<table>
<thead>
<tr>
<th>Item</th>
<th>Template</th>
<th>OpenNMT</th>
<th>Re$^3$Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEN_DIF</td>
<td>2.6±2.6</td>
<td>3.0±4.4</td>
<td>2.7±2.6</td>
</tr>
<tr>
<td>LESS_3</td>
<td>0</td>
<td>53</td>
<td>1</td>
</tr>
<tr>
<td>COPY(%)</td>
<td>31</td>
<td>80</td>
<td>74</td>
</tr>
<tr>
<td>NEW_NE</td>
<td>0.51</td>
<td>0.34</td>
<td>0.30</td>
</tr>
<tr>
<td>NEW_UP</td>
<td>0.38</td>
<td>0.19</td>
<td>0.11</td>
</tr>
</tbody>
</table>
Effects of Template

- Performance highly relies on templates
- The rewriting ability is strong

<table>
<thead>
<tr>
<th>Type</th>
<th>ROUGE-1</th>
<th>ROUGE-2</th>
<th>ROUGE-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>+Random</td>
<td>32.60</td>
<td>14.31</td>
<td>30.19</td>
</tr>
<tr>
<td>+First</td>
<td>36.01</td>
<td>17.06</td>
<td>33.21</td>
</tr>
<tr>
<td>+Max</td>
<td>41.50</td>
<td>21.97</td>
<td>38.80</td>
</tr>
<tr>
<td>+Optimal</td>
<td>46.21</td>
<td>26.71</td>
<td>43.19</td>
</tr>
<tr>
<td>+Rerank(Re^3Sum)</td>
<td>37.04</td>
<td>19.03</td>
<td>34.46</td>
</tr>
</tbody>
</table>
### Generation Diversity

**OpenNMT**  Beam search n-best outputs

**Re³Sum**  Provide different templates

<table>
<thead>
<tr>
<th>Source</th>
<th>anny ainge said thursday he had two one-hour meetings with the new owners of the boston celtics but no deal has been completed for him to return to the franchise .</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target</td>
<td>ainge says no deal completed with celtics</td>
</tr>
<tr>
<td>Templates</td>
<td>major says no deal with spain on gibraltar roush racing completes deal with red sox owner</td>
</tr>
<tr>
<td>Re³Sum</td>
<td>ainge says no deal done with celtics ainge talks with new owners</td>
</tr>
<tr>
<td>OpenNMT</td>
<td>ainge talks with celtics owners ainge talks with new owners</td>
</tr>
</tbody>
</table>
Outline

1 Introduction
2 Method
3 Experiments
4 Conclusion
Conclusion

- Introduce soft templates as additional input to guide seq2seq summarization
- Combine IR-based ranking techniques and seq2seq-based generation techniques to utilize both supervisions
- Improve informativeness, stability, readability and diversity
Thank you