Summary Loop: Unsupervised Abstractive Summarization

The 58th Annual Meeting of the Association for Computational Linguistics

Philippe Laban, John Canny, Marti Hearst, UC Berkeley
Andrew Hsi, Bloomberg

TechAtBloomberg.com
(CNN) - Chilean President Sebastian Pinera announced Wednesday that his country, which has been paralyzed by protests over the last two weeks, will no longer host two major international summits.

Clashes at demonstrations in the capital of Santiago have left at least 20 people dead and led to the resignation of eight key ministers from Pinera’s cabinet.

The President has now canceled the hosting of the economic APEC forum and the COP25 environmental summit, which were both due to take place later this year.

[...]

On CNN.com in October 2019.

What is a good summary?

Most common automatic evaluation: **ROUGE**.

ROUGE is based on n-gram overlap between the evaluated summary and a reference (human written).
What is a good summary?

GREAT! Can we directly optimize ROUGE score? Paulus et. al 2017 tried it.
What is a good summary?

GREAT! Can we directly optimize ROUGE score?
Paulus et. al 2017 tried it.

Good news. Trained a model with RL that achieved very high ROUGE score.

Bad news. The summaries are poorly rated by humans.

Example summary with high ROUGE score:
Button was denied his 100th race for McLaren after an ERS prevented him from making it to the start-line. It capped a miserable weekend for the Briton. Button has out-qualified. Finished ahead of Nico Rosberg at Bahrain. Lewis Hamilton has. In 11 races. . The race. To lead 2,000 laps. . In. . . And.
What is a good summary?

Let's try with a definition.
What is a good summary?

Let's try with a definition.

A summary is a brief, fluent text that covers the main points of an original document.
What is a good summary?

Let's try with a definition.

A summary is a brief, fluent text that covers the main points of an original document.

Three pillars of summarization:

brevity  fluency  coverage
Unsupervised and abstractive summarization technique
Summary Loop Diagram

Unsupervised and abstractive summarization technique
Summary Loop Diagram

Unsupervised and abstractive summarization technique
Unsupervised and abstractive summarization technique
Unsupervised and abstractive summarization technique
First, the masking procedure.
CNN - Chilean President Sebastian Pinera announced Wednesday that his country, which has been paralyzed by protests over the last two weeks, will no longer host two major international summits.

Clashes at demonstrations in the capital of Santiago have left at least 20 people dead and led to the resignation of eight key ministers from Pinera’s cabinet.

The President has now canceled the hosting of the economic APEC forum and COP25 environmental summit, which were both due to take place later this year.

[...]
(CNN) - __________ __________ __________ ________
announced Wednesday that his country, which has been
__________ by __________ over the last two weeks, will no
longer ____ two major international ________.

Clashes at demonstrations in the capital of Santiago have left
at least 20 people dead and led to the resignation of eight
key ministers from Pinera's cabinet.

The __________ has now __________ the ________ of the
economic ____ forum and _____ environmental ________,
which were both due to take place later this ____.

[...]
"A summary is a brief, fluent text that covers the main points of an original document."
Coverage is the accuracy at recovering the masked keywords, using the summary.
"A summary is a brief, fluent text that covers the main points of an original document."
The Summary Loop involves 2 GPT-2 and a BERT model.
Example Training Run

Trained with Self-Critical Sequence Training (SCST)

Target Length: 10 words

Trained on single Titan X GPU
What about brevity?
Effect of the Target Length

News article

(CNN) - Chilean President Sebastian Pinera announced Wednesday that his country, which has been paralyzed by protests over the last two weeks, will no longer host two major international summits.

Clashes at demonstrations in the capital of Santiago have left at least 20 people dead and led to the resignation of eight key ministers from Pinera's cabinet.

The President has now canceled the hosting of the economic APEC forum and COP25 environmental summit, which were both due to take place later this year.

[...]
## ROUGE Results

<table>
<thead>
<tr>
<th>Supervised Method</th>
<th>R-1</th>
<th>R-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pointer Generator (See et al.)</td>
<td>36.4</td>
<td>33.4</td>
</tr>
<tr>
<td>PG + Coverage (See et al.)</td>
<td>39.5</td>
<td>36.4</td>
</tr>
<tr>
<td>Bottom-Up (Gehrmann et al.)</td>
<td>41.2</td>
<td>38.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unsupervised Methods</th>
<th>R-1</th>
<th>R-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>TextRank (Extractive)</td>
<td>35.2</td>
<td>28.7</td>
</tr>
<tr>
<td>GPT2 Zero-Shot (Radford)</td>
<td>29.4</td>
<td>26.6</td>
</tr>
<tr>
<td>Summary Loop 45</td>
<td>37.7</td>
<td>34.7</td>
</tr>
</tbody>
</table>

On standard test-set of CNN/DM. Approaching ROUGE of supervised methods **without seeing a single summary example**.
### Manual Analysis of 200 random samples of CNN/DM test set (errors, techniques).

Each summary is labeled for the presence of 4 summarization techniques. Unsuccessful technique application can lead to an error (inaccuracy or ungrammaticality).

* Computed on the 3 most challenging techniques (not considering Sentence Compression).
Abstractive? How abstractive.

Distribution of lengths of copied spans.
Gold summaries (handwritten) copy shorter passages and use novel words.
The Summary Loop is an unsupervised, abstractive summarization method.

You can try it on your domain/language if you have:
- A large corpus of documents (100K minimum)
- A desired summary length (e.g., 30 words)
- A BERT model in your target language (for Coverage)
- A GPT2 model in your target language (for Summarizer & Fluency)
Questions?
Come ask at the Live Q&A
Tuesday July 7th, Session 9A & 10B
Code on GitHub:
https://github.com/CannyLab/summary_loop
Contact:
phillab@berkeley.edu
Optimization Procedure: SCST

Directly optimize:

\[ \text{Summary score} = \alpha \text{Fluency Score} + \beta \text{Coverage} \]
Optimization Procedure: SCST

Directly optimize:

Summary score = $\alpha$ Fluency Score + $\beta$ Coverage

Self-Critical Sequence Training originally applied to Image Captioning:

1) Generate two candidate summaries $S_1$ and $S_2$ (different sampling methods)
Optimization Procedure: SCST

Directly optimize:

Summary score = $\alpha$Fluency Score + $\beta$Coverage

Self-Critical Sequence Training originally applied to Image Captioning:

1) Generate two candidate summaries $S_1$ and $S_2$ (different sampling methods)
2) Compute Summary Score for each: $R_1$ and $R_2$

Optimization Procedure: SCST

Directly optimize:

\[ \text{Summary score} = \alpha \text{Fluency Score} + \beta \text{Coverage} \]

Self-Critical Sequence Training originally applied to Image Captioning:

1) Generate two candidate summaries $S_1$ and $S_2$ (different sampling methods)
2) Compute Summary Score for each: $R_1$ and $R_2$
3) Gradients updates using REINFORCE, based on the difference between scores: $(R_1 - R_2)$

Essentially: Increasing likelihood of summary with higher reward, increasing expected reward.