Get To The Point:
Summarization with Pointer-Generator Networks

Abigail See*  Peter J. Liu†  Christopher Manning*

*Stanford NLP  †Google Brain

1st August 2017
Two approaches to summarization

Extractive Summarization

*Select parts* (typically sentences) of the original text to form a summary.

- Easier
- Too restrictive (no paraphrasing)
- Most past work is extractive

Abstractive Summarization

*Generate novel sentences* using natural language generation techniques.

- More difficult
- More flexible and human
- Necessary for future progress
The papal bill Pope Francis insisted on paying himself...before catching the bus home after winning the election

Pope Francis insisted on returning to his hotel to settle the bill himself.
- The pontiff also chose to use a bus instead of a chauffeur driven car.
- The 76-year-old has eschewed ceremonial traditions for a more humble approach.

With the spiritual wellbeing of the world's 1.2 billion Catholics on his shoulders he must have quite a to-do list. But despite his new responsibilities, Francis did not forget to stop off - between engagements - to

say his HOPE. He jokingly said: 'I need to set a good example.'

Staff at the central Rome priests' residence where Bergoglio was staying before the conclave, were astonished when the newly elected Pope strolled in to collect his luggage and settle the bill.

He wanted to come here because he wanted to thank the personnel, people who work in this house he said. He greeted them one by one, no rush, the whole staff, one by one. Mr Rytle-Bergoglio added that Francis apparently knew everyone by name.

A Vatican spokesman said: "He wanted to get his luggage and the bags. He had left everything there."

He then stopped in the office, greeted everyone and decided to pay the bill for the room...because he was concerned about giving a good example of what priests and bishops should do.

He is already winning plaudits for his "down-to-earth manner."

He has so far defied a motorcade and the official papal Jag for official business. And even on the ride of the papa, instead of the traditional car...the other cardinals back to their lodgings by minibus, saying: I came on the bus, so I go home on the bus.

Meeting cardinals yesterday on his second day of Papal business he eschewed protocol in favour of kissing on two cheeks, shaking hands and hugging.

He told his deputies that people like himself are like good wine, getting better with age, before urging them to impart their wisdom to the young.

Francis began his reign in unorthodox fashion as he shunned public events in order to pray to the Virgin Mary.

During his first Mass since being elected as supreme pontiff, Pope Francis and his cardinals were dressed in simple yellow robes over their cassocks, rather than the formal ceremonial outfits they would normally wear on such a high occasion.

Speaking in Italian without notes, he said: "We can walk all we want, we can build many things, but if we don't proclaim Jesus Christ, something is wrong. We would become a compassionate NGO and not a Church which is the bride of Christ."

"He who does not pray to the Lord prays to the devil. When we don't proclaim Jesus Christ, we proclaim the worldliness of the devil, the worldliness of the demon."

"We must always walk in the presence of the Lord, in the light of the lord, always trying to live in an irreproachable way," he said in a heartfelt homily of a parish priest, loaded with biblical references and simple imagery.

"When we walk without the cross, when we build without the cross and when we proclaim Christ without the cross, we are not disciples of the Lord. We are worldly," he said.

"We may be bishops, priests, cardinals, popes, all of this, but we are not disciples of the Lord," he said.

It was a far simpler message than the dense, three-page discourse Benedict delivered in Latin during his first Mass as pope in 2005.

The difference in style was a sign of Francis' belief that the Catholic Church needs to be one with the people...the one who serves and not impose its message on a society that often doesn't want to hear it, Francis' authorised biographer, Sergio Rubin, said.

Francis took the helm of the 1.2 billion-member Church at a time of style and intrigue, with the Vatican rocked by a string of sex abuse scandals, accusations of infighting within its central government and by allegations of financial wrongdoing.

But many within the church believe he could change it for the better.

"It seems to me for now what is certain is it's a great change of style, which for us isn't a small thing," Mr Ruben said, recalling how the former Cardinal Jorge Bergoglio would celebrate Masses with homeless people and prostitutes in Buenos Aires.

"He believes the church has to go to the streets," he said, "to express this closeness of the church and this accompanying with those who are suffering."
Sequence-to-sequence + attention model

Context Vector

Germany

Vocabulary Distribution

a

Attention Distribution

zoo

Encoder Hidden States

weighted sum

weighted sum

Decoder Hidden States

“beat”

Source Text

Partial Summary

Germany emerge victorious in 2-0 win against Argentina on Saturday...
Germany emerge victorious in 2-0 win against Argentina on Saturday.

Source Text: Germany emerge victorious in 2-0 win against Argentina on Saturday...

Partial Summary: <START> Germany beat...
Sequence-to-sequence + attention model

Source Text:

Germany emerge victorious in 2-0 win against Argentina on Saturday

Argentina 2-0 beat Germany
Two Problems

Problem 1: The summaries sometimes reproduce factual details inaccurately.

   e.g. Germany beat Argentina 3-2

Problem 2: The summaries sometimes repeat themselves.

   e.g. Germany beat Germany beat Germany beat...
Two Problems

**Problem 1:** The summaries sometimes reproduce factual details inaccurately.

  e.g. Germany beat Argentina 3-2

  Incorrect rare or out-of-vocabulary word

**Solution:** Use a pointer to copy words.

**Problem 2:** The summaries sometimes repeat themselves.

  e.g. Germany beat Germany beat Germany beat...
Get to the point!

Source Text

Germany emerge victorious in 2-0 win against Argentina on Saturday.
<table>
<thead>
<tr>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UNK UNK</strong> was expelled from the <em>dubai open chess tournament</em></td>
<td><strong>gaioz nigalidze</strong> was expelled from the <em>dubai open chess tournament</em></td>
</tr>
<tr>
<td><em>the 2015 rio olympic games</em></td>
<td><em>the 2016 rio olympic games</em></td>
</tr>
</tbody>
</table>
Two Problems

**Problem 1:** The summaries sometimes reproduce factual details inaccurately.

  e.g. *Germany beat Argentina 3-2*

**Solution:** Use a *pointer* to copy words.

**Problem 2:** The summaries sometimes repeat themselves.

  e.g. *Germany beat Germany beat Germany beat...*
Two Problems

**Problem 1:** The summaries sometimes reproduce factual details inaccurately.

* e.g. *Germany beat Argentina 3-2*

**Solution:** Use a pointer to copy words.

**Problem 2:** The summaries sometimes repeat themselves.

* e.g. *Germany beat Germany beat Germany beat...*

**Solution:** Penalize repeatedly attending to same parts of the source text.
Reducing repetition with coverage

Coverage = cumulative attention = what has been covered so far

Source Text: Germany emerge victorious in 2-0 win against Argentina on Saturday

Summary: Germany beat _____
Reducing repetition with coverage

Coverage = cumulative attention = what has been covered so far

1. Use coverage as extra input to attention mechanism.

Source Text: Germany emerge victorious in 2-0 win against Argentina on Saturday
Summary: Germany beat _____

Reducing repetition with coverage

Coverage = cumulative attention = what has been covered so far

1. Use coverage as **extra input to attention mechanism**.
2. **Penalize** attending to things that have already been covered.

---

[4] *Modeling coverage for neural machine translation*. Tu et al., 2016,


Reducing repetition with coverage

Coverage = cumulative attention = what has been covered so far

1. Use coverage as extra input to attention mechanism.
2. Penalize attending to things that have already been covered.

Result: repetition rate reduced to level similar to human summaries

---

Summaries are still mostly extractive

Article (truncated): Lagos, Nigeria (CNN) a day after winning Nigeria’s presidency, Muhammadu Buhari told CNN’s Christiane Amanpour that he plans to aggressively fight corruption that has long plagued Nigeria and go after the root of the nation’s unrest. Buhari said he’ll “rapidly give attention” to curbing violence in the northeast part of Nigeria, where the terrorist group Boko Haram operates. By cooperating with neighboring nations Chad, Cameroon and Niger, he said his administration is confident it will be able to thwart criminals and others contributing to Nigeria’s instability. For the first time in Nigeria’s history, the opposition defeated the ruling party in democratic elections. Buhari defeated incumbent Goodluck Jonathan by about 2 million votes, according to Nigeria’s independent national electoral commission. The win comes after a long history of military rule, coups and botched attempts at democracy in Africa’s most populous nation.
Results

ROUGE compares the machine-generated summary to the human-written reference summary and counts co-occurrence of 1-grams, 2-grams, and longest common sequence.

<table>
<thead>
<tr>
<th></th>
<th>ROUGE-1</th>
<th>ROUGE-2</th>
<th>ROUGE-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nallapati et al. 2016</td>
<td>35.5</td>
<td>13.3</td>
<td>32.7</td>
</tr>
</tbody>
</table>

Previous best abstractive result
## Results

ROUGE compares the **machine-generated summary** to the **human-written reference summary** and counts co-occurrence of 1-grams, 2-grams, and **longest common sequence**.

<table>
<thead>
<tr>
<th></th>
<th>ROUGE-1</th>
<th>ROUGE-2</th>
<th>ROUGE-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nallapati et al. 2016</td>
<td>35.5</td>
<td>13.3</td>
<td>32.7</td>
</tr>
<tr>
<td>Ours (seq2seq baseline)</td>
<td>31.3</td>
<td>11.8</td>
<td>28.8</td>
</tr>
<tr>
<td>Ours (pointer-generator)</td>
<td>36.4</td>
<td>15.7</td>
<td>33.4</td>
</tr>
<tr>
<td>Ours (pointer-generator + coverage)</td>
<td><strong>39.5</strong></td>
<td><strong>17.3</strong></td>
<td><strong>36.4</strong></td>
</tr>
</tbody>
</table>

Previous best abstractive result: 35.5, 13.3, 32.7

Our improvements: 36.4, 15.7, 33.4
### Results

ROUGE compares the machine-generated summary to the human-written reference summary and counts co-occurrence of 1-grams, 2-grams, and longest common sequence.

<table>
<thead>
<tr>
<th></th>
<th>ROUGE-1</th>
<th>ROUGE-2</th>
<th>ROUGE-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nallapati et al. 2016</td>
<td>35.5</td>
<td>13.3</td>
<td>32.7</td>
</tr>
<tr>
<td>Ours (seq2seq baseline)</td>
<td>31.3</td>
<td>11.8</td>
<td>28.8</td>
</tr>
<tr>
<td>Ours (pointer-generator)</td>
<td>36.4</td>
<td>15.7</td>
<td>33.4</td>
</tr>
<tr>
<td>Ours (pointer-generator + coverage)</td>
<td>39.5</td>
<td><strong>17.3</strong></td>
<td>36.4</td>
</tr>
<tr>
<td>Paulus et al. 2017 (hybrid RL approach)</td>
<td>39.9</td>
<td>15.8</td>
<td>36.9</td>
</tr>
<tr>
<td>Paulus et al. 2017 (RL-only approach)</td>
<td><strong>41.2</strong></td>
<td>15.8</td>
<td><strong>39.1</strong></td>
</tr>
</tbody>
</table>

Previous best abstractive result:
- **worse** ROUGE; **better** human eval

Our improvements:
- **better** ROUGE; **worse** human eval
Results

ROUGE compares the **machine-generated summary** to the **human-written reference summary** and counts co-occurrence of 1-grams, 2-grams, and **longest common sequence**.

<table>
<thead>
<tr>
<th></th>
<th>ROUGE-1</th>
<th>ROUGE-2</th>
<th>ROUGE-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nallapati et al. 2016</td>
<td>35.5</td>
<td>13.3</td>
<td>32.7</td>
</tr>
<tr>
<td>Ours (seq2seq baseline)</td>
<td>31.3</td>
<td>11.8</td>
<td>28.8</td>
</tr>
<tr>
<td>Ours (pointer-generator)</td>
<td>36.4</td>
<td>17.3</td>
<td>33.4</td>
</tr>
<tr>
<td>Ours (pointer-generator + coverage)</td>
<td>39.5</td>
<td>17.3</td>
<td>36.4</td>
</tr>
<tr>
<td>Paulus et al. 2017 (hybrid RL approach)</td>
<td>39.9</td>
<td>15.8</td>
<td>36.9</td>
</tr>
<tr>
<td>Paulus et al. 2017 (RL-only approach)</td>
<td>41.2</td>
<td>15.8</td>
<td>39.1</td>
</tr>
</tbody>
</table>

Previous best abstractive result
- **worse** ROUGE; **better** human eval

Our improvements
- **better** ROUGE; **worse** human eval
The difficulty of evaluating summarization

- Summarization is **subjective**
  - There are many correct ways to summarize
The difficulty of evaluating summarization

- Summarization is **subjective**
  - There are many correct ways to summarize

- ROUGE is based on strict **comparison to a reference summary**
  - Intolerant to rephrasing
  - **Rewards extractive** strategies
The difficulty of evaluating summarization

- Summarization is subjective
  - There are many correct ways to summarize

- ROUGE is based on strict comparison to a reference summary
  - Intolerant to rephrasing
  - Rewards extractive strategies

- Take first 3 sentences as summary → higher ROUGE than (almost) any published system
  - Partially due to news article structure
A crowd gathers near the entrance of Tokyo’s upscale Mitsukoshi Department Store, which traces its roots to a kimono shop in the late 17th century.

Fitting with the store’s history, the new greeter wears a traditional Japanese kimono while delivering information to the growing crowd, whose expressions vary from amusement to bewilderment.

It’s hard to imagine the store’s founders in the late 1600’s could have imagined this kind of employee.

That’s because the greeter is not a human -- it’s a robot.

Aiko Chihira is an android manufactured by Toshiba, designed to look and move like a real person.
What next?
Extractive methods
Human-level summarization

MOUNT ABSTRACTION

long text understanding
paraphrasing

Extractive methods

SAFETY
Human-level summarization

MOUNT ABSTRACTION

long text understanding paraphrasing

SAFETY

Extractive methods

SWAMP OF BASIC ERRORS

repetition nonsense copying errors
Human-level summarization

Mount Abstraction

- More high-level understanding?
- More scalability?
- Better metrics?

Swamp of Basic Errors

- RNNs
- Repetition
- Nonsense
- Copying errors

Safety

Extractive methods
Thank you!

Blog post: www.abigailsee.com
Code: github.com/abisee/pointer-generator