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

CNN / Daily Mail dataset

- Long news articles (average ~800 words)

- Multi-sentence summaries (usually 3 or 4 sentences, average 56 words)

- Summary contains information from throughout the article

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 is 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 pay his hotel bill.

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.

Pope Francis insisted on returning to the hotel to collect his luggage and greet the staff before settling the hotel bill himself

'I need to set a good example' he joked.

He was driven to the hotel in a simple car and The Rev. Pawel Rytel-Andrianek, who teaches at the nearby Pontifical Holy Cross University and is staying at the residence, said that workers at the hotel were touched by the Pope's decision to return and bid them farewell.

'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 Rytel-Andrianek 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.'

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

He has so far refused a motorcade and the official papal Jag for official business. And even on the night of the election he insisted on accompanying the other cardinals back to their lodgings by mini bus, saying: 'I came on the bus, so I'll 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 old 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 major 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 irreprehensible 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 at one with the people it serves and not impose its message on a society that often doesn't want to hear it, Francis' authorised biographer, Sergio Rubin, said.
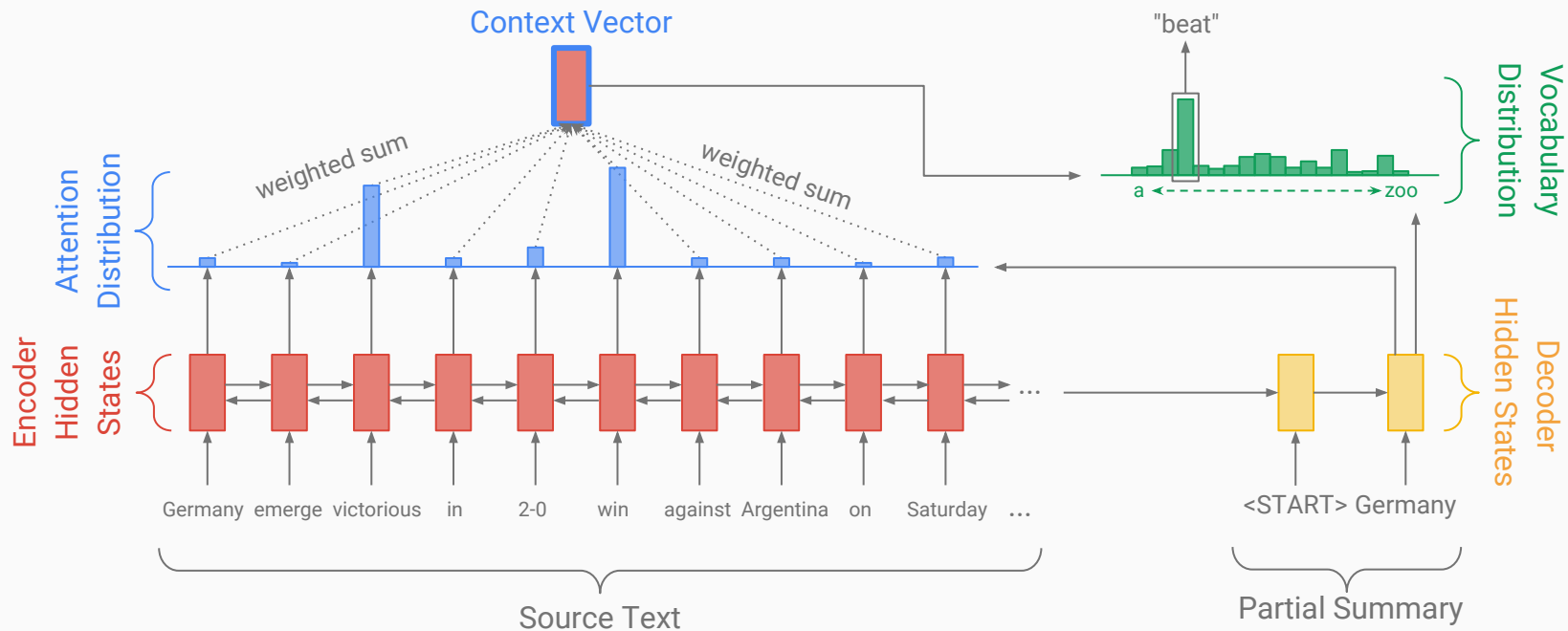
Francis took the helm of the 1.2 billion-member Church at a time of strife and intrigue, with the Vatican rocked by a string of sex abuse scandals, accusations of infighting within its central government and by allegations of financial wrongdoing.

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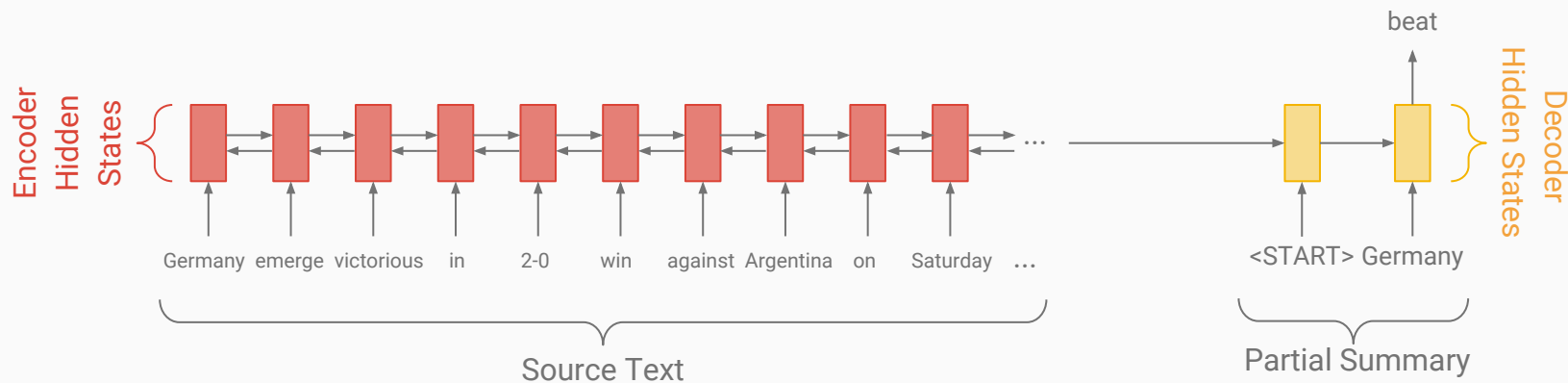
'It seems to me for now what is certain is it's a great change of style, which for us isn't a small thing,' Mr Rubin said, recalling how the former Cardinal Jorge Bergoglio would celebrate Masses with homeless people and prostitutes in Buenos Aires.

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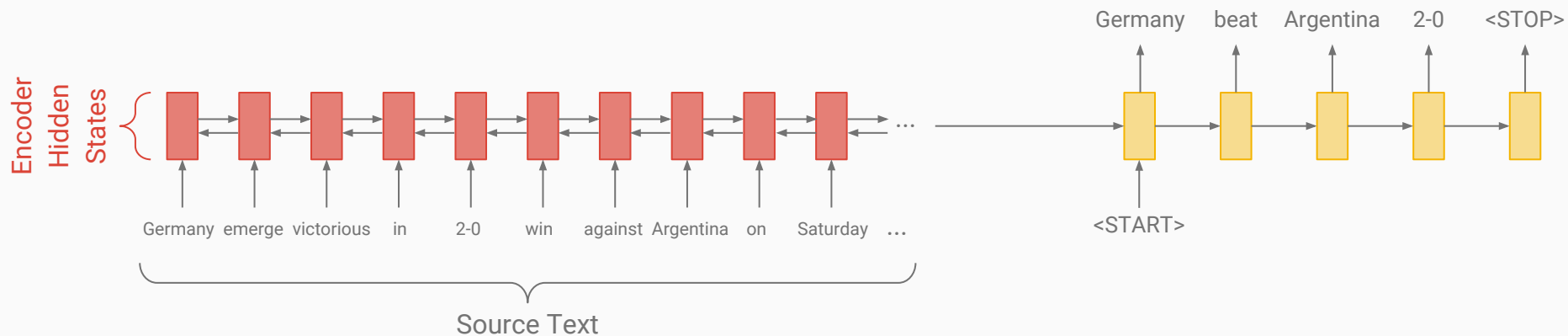
Sequence-to-sequence + attention model

Sequence-to-sequence + attention model

Sequence-to-sequence + attention model

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

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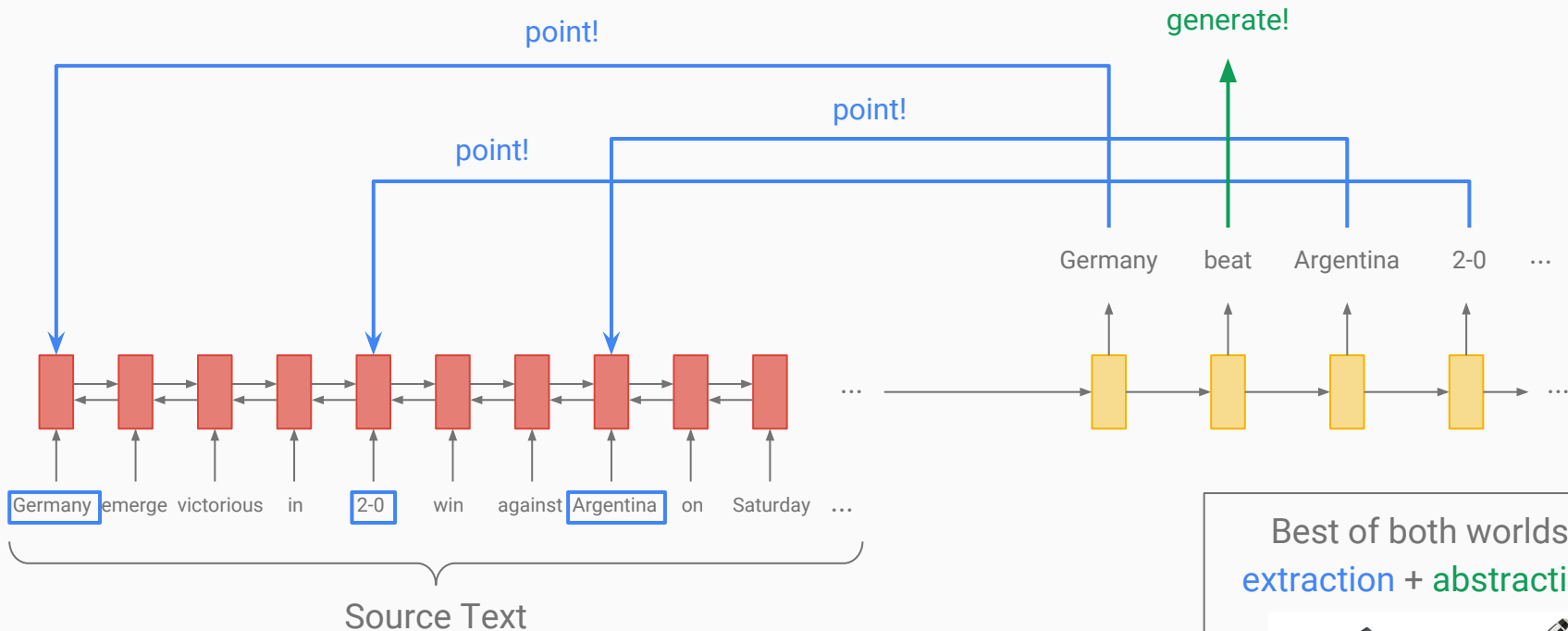
e.g. *Germany beat Argentina 3-2*

Incorrect **rare** or **out-of-vocabulary** word

Solution: Use a pointer to copy words.

Problem 2: The summaries sometimes repeat themselves.

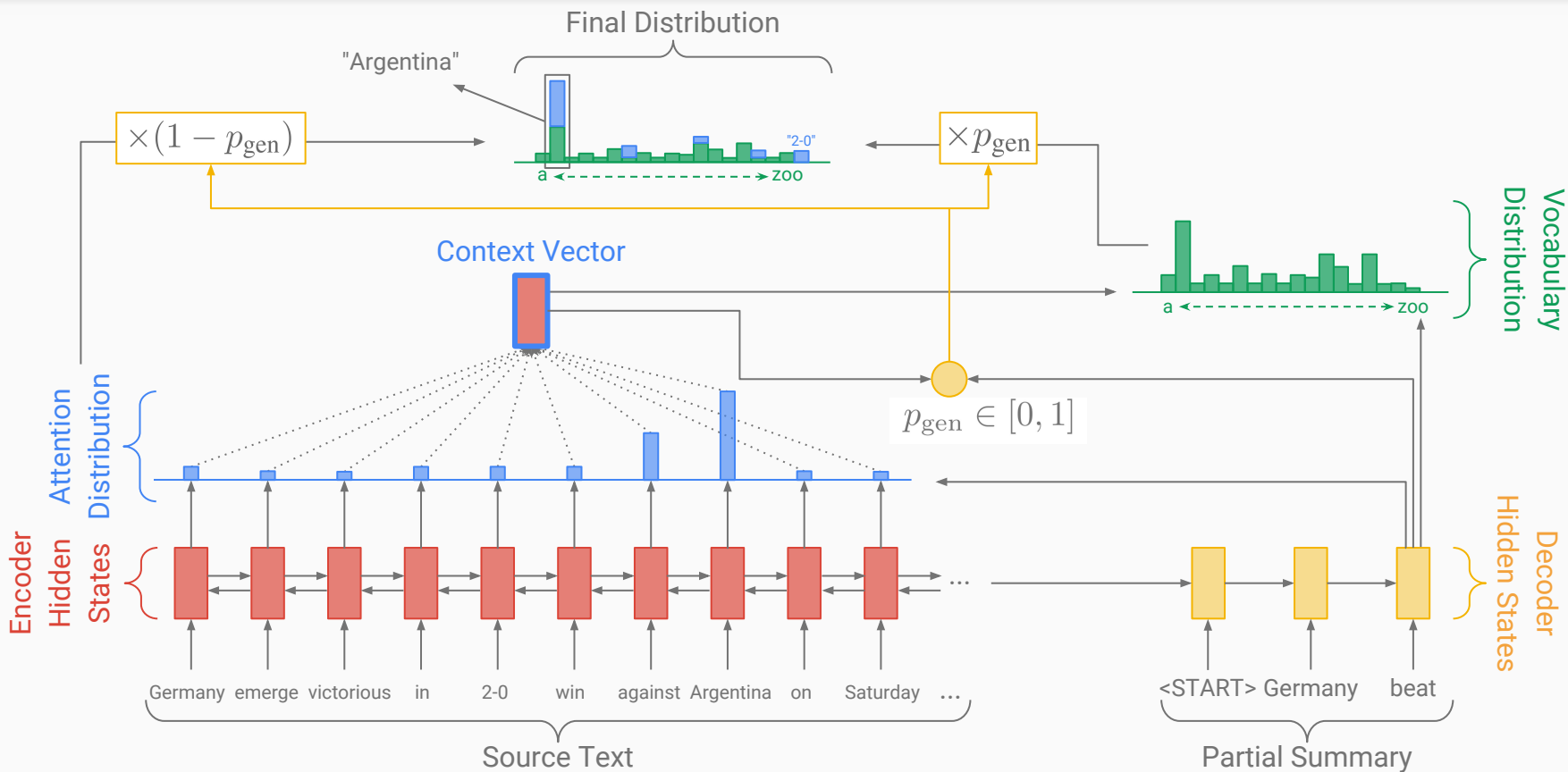e.g. *Germany beat Germany beat Germany beat…*

Get to the point!

point!

point!

point!

generate!

Germany   beat   Argentina   2-0   ...

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

Source Text

Best of both worlds:
extraction + abstraction

[1] *Incorporating copying mechanism in sequence-to-sequence learning.* Gu et al., 2016.
[2] *Language as a latent variable: Discrete generative models for sentence compression.* Miao and Blunsom, 2016.

Pointer-generator network

Improvements

Before	After
UNK UNK was expelled from the dubai open chess tournament	*gaioz nigalidze was expelled from the dubai open chess tournament*
the 2015 rio olympic games	*the 2016 rio olympic games*

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 _____

[4] **Modeling coverage for neural machine translation.** *Tu et al., 2016,*
[5] **Coverage embedding models for neural machine translation.** *Mi et al., 2016*
[6] **Distraction-based neural networks for modeling documents.** *Chen et al., 2016.*

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 _____

1. Use coverage as extra input to attention mechanism.

[4] **Modeling coverage for neural machine translation.** *Tu et al., 2016,*
[5] **Coverage embedding models for neural machine translation.** *Mi et al., 2016*
[6] **Distraction-based neural networks for modeling documents.** *Chen et al., 2016.*

Reducing repetition with coverage

Coverage = cumulative attention = what has been covered so far

Don't attend here

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

Summary: Germany beat _____
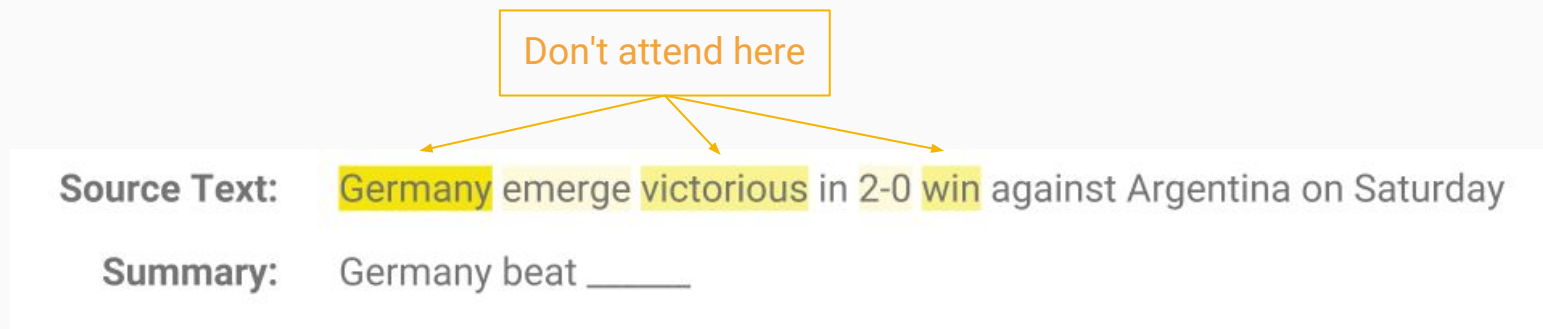
1. Use coverage as extra input to attention mechanism.

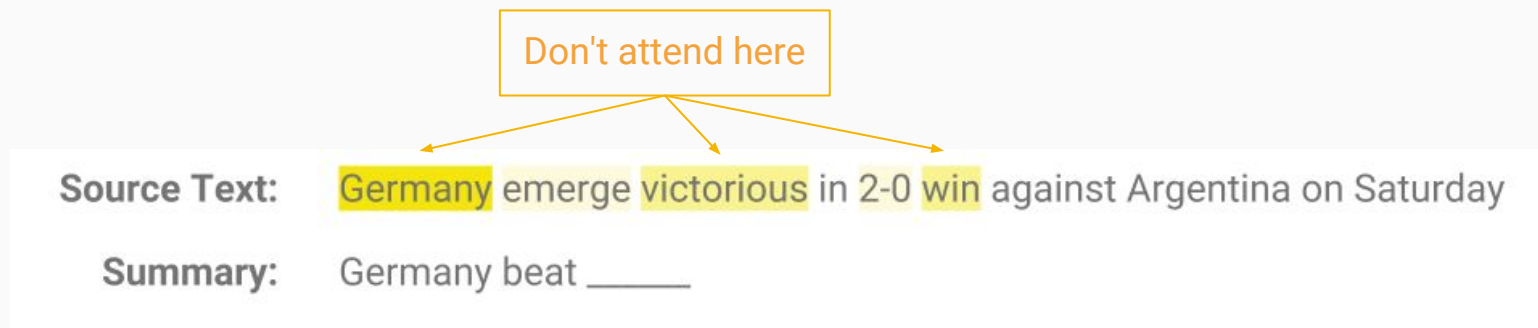2. Penalize attending to things that have already been covered.

[4] **Modeling coverage for neural machine translation.** Tu et al., 2016,
[5] **Coverage embedding models for neural machine translation.** Mi et al., 2016
[6] **Distraction-based neural networks for modeling documents.** Chen et al., 2016.

Reducing repetition with coverage

Coverage = cumulative attention = what has been covered so far

Don't attend here

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

Summary: Germany beat _____

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,*
[5] **Coverage embedding models for neural machine translation.** *Mi et al., 2016*
[6] **Distraction-based neural networks for modeling documents.** *Chen et al., 2016.*

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 .

Source Text

Final Coverage

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.

	ROUGE-1	ROUGE-2	ROUGE-L
Nallapati et al. 2016	**35.5**	**13.3**	**32.7**

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.

	ROUGE-1	ROUGE-2	ROUGE-L
Nallapati et al. 2016	35.5	13.3	32.7
Ours (seq2seq baseline)	31.3	11.8	28.8
Ours (pointer-generator)	36.4	15.7	33.4
Ours (pointer-generator + coverage)	**39.5**	**17.3**	**36.4**

Previous best abstractive result

Our improvements

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.

	ROUGE-1	ROUGE-2	ROUGE-L
Nallapati et al. 2016	35.5	13.3	32.7
Ours (seq2seq baseline)	31.3	11.8	28.8
Ours (pointer-generator)	36.4	15.7	33.4
Ours (pointer-generator + coverage)	39.5	**17.3**	36.4
Paulus et al. 2017 (hybrid RL approach)	39.9	15.8	36.9
Paulus et al. 2017 (RL-only approach)	**41.2**	15.8	**39.1**

Previous best abstractive result

Our improvements

worse ROUGE; better human eval

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.

	ROUGE-1	ROUGE-2	ROUGE-L	
Nallapati et al. 2016	35.5	13.3	32.7	Previous best abstractive result
Ours (seq2seq baseline)	31.3	11.8	28.8	
Ours (pointer-generator)	36.4	15.	33.4	Our improvements
Ours (pointer-generator + coverage)	39.5	17.3	36.4	
Paulus et al. 2017 (hybrid RL approach)	39.9	15.8	36.9	worse ROUGE; better human eval
Paulus et al. 2017 (RL-only approach)	41.2	15.8	39.1	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

First sentences not always a good summary

Robots tested in Japan companies

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.

Irrelevant

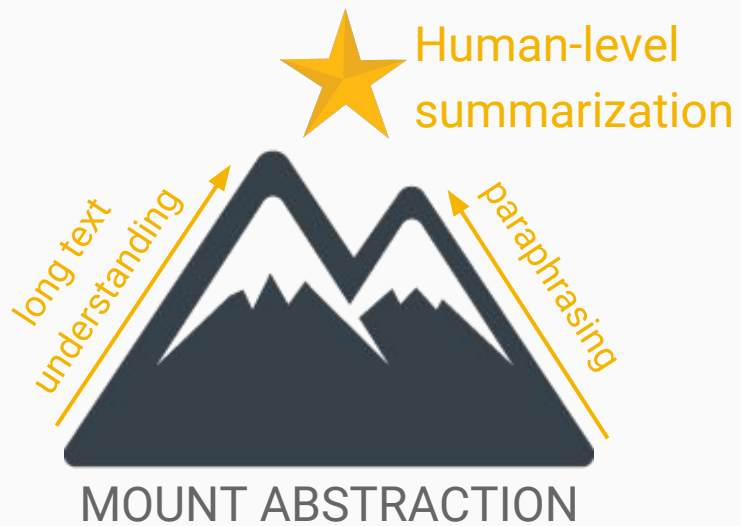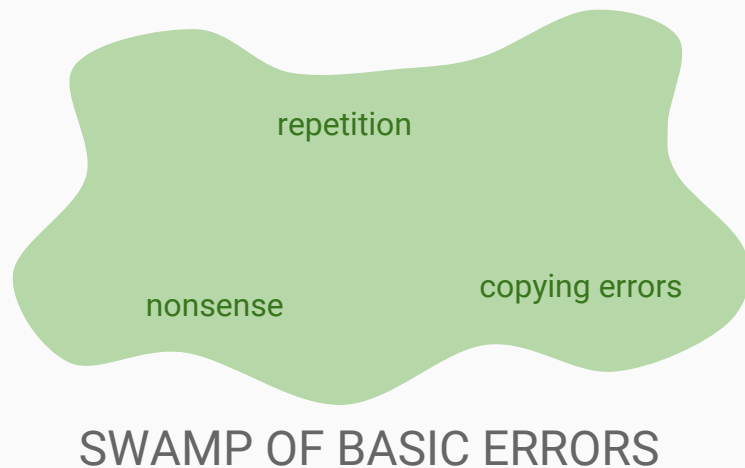*Aiko Chihira is an android manufactured by Toshiba, designed to look and move like a real person.*
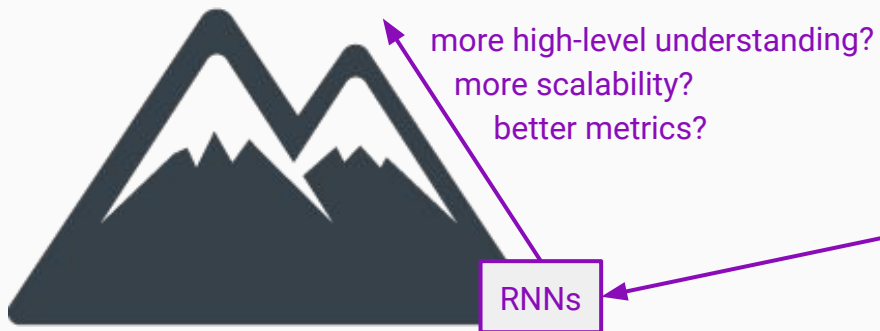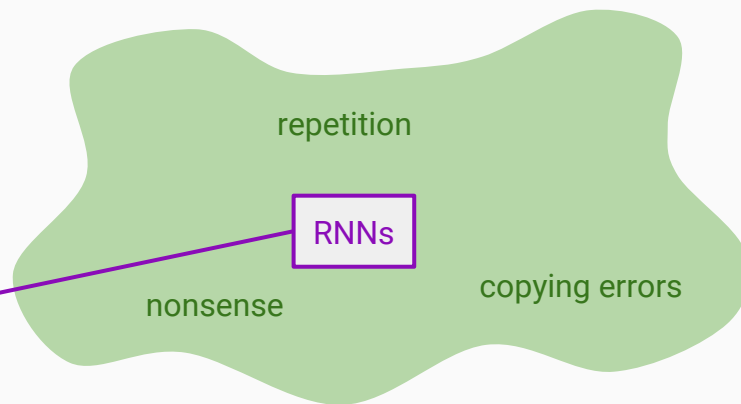
Our system starts here

...

What next?

Extractive
methods

SAFETY

Human-level summarization

long text understanding

paraphrasing

MOUNT ABSTRACTION

repetition

nonsense

copying errors

SWAMP OF BASIC ERRORS

Extractive methods

SAFETY

Human-level summarization

more high-level understanding?
more scalability?
better metrics?

MOUNT ABSTRACTION

repetition

RNNs

nonsense

copying errors

SWAMP OF BASIC ERRORS

RNNs

Extractive methods

SAFETY

Thank you!

⭐

🏔️

Blog post: www.abigailsee.com

Code: github.com/abisee/pointer-generator