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

Running example.

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.

[...]

On **CNN.com** in October 2019.

Abstractive Summary

Chilean President announced his country will not host the APEC forum and the COP25 anymore, due to protests in Santiago.

https://www.cnn.com/2019/10/30/americas/chile-protests-apec-cop25-hosting-canceled-intl/index.html

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 atBahrain. 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.

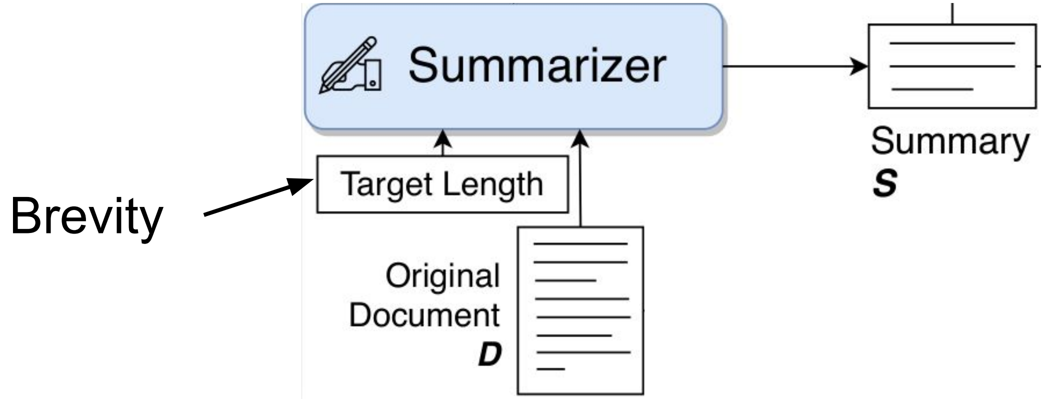
Three pillars of summarization:

brevity

fluency

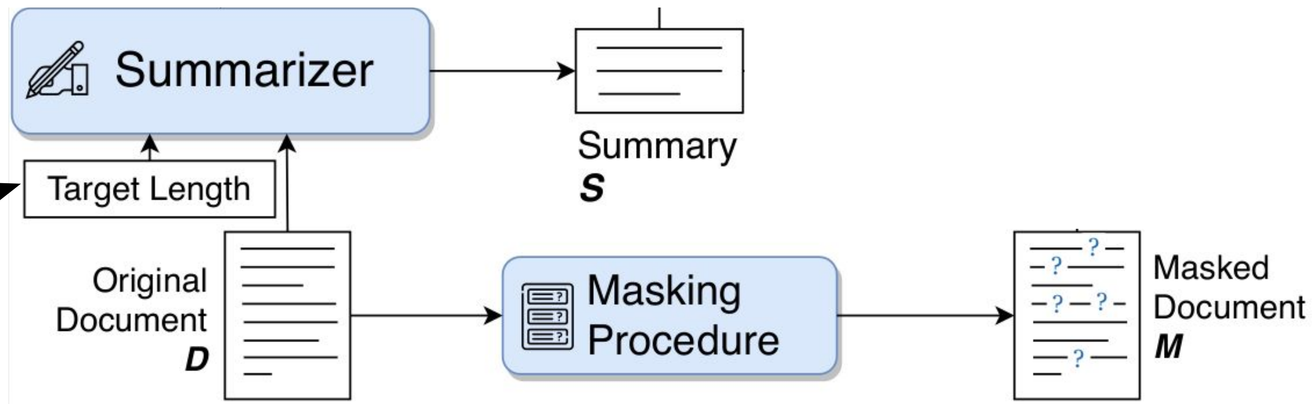
coverage

Summary Loop Diagram



Unsupervised and abstractive summarization technique

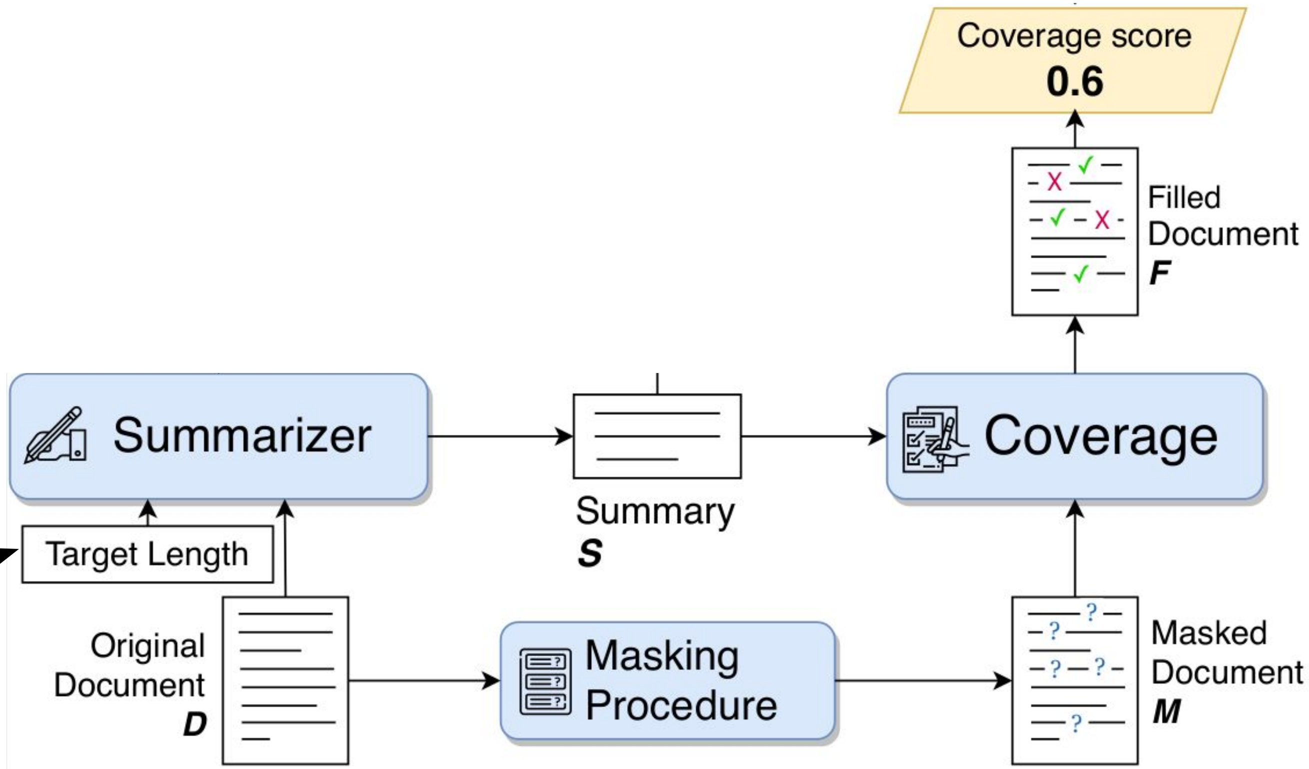
Summary Loop Diagram



Brevity

Unsupervised and abstractive summarization technique

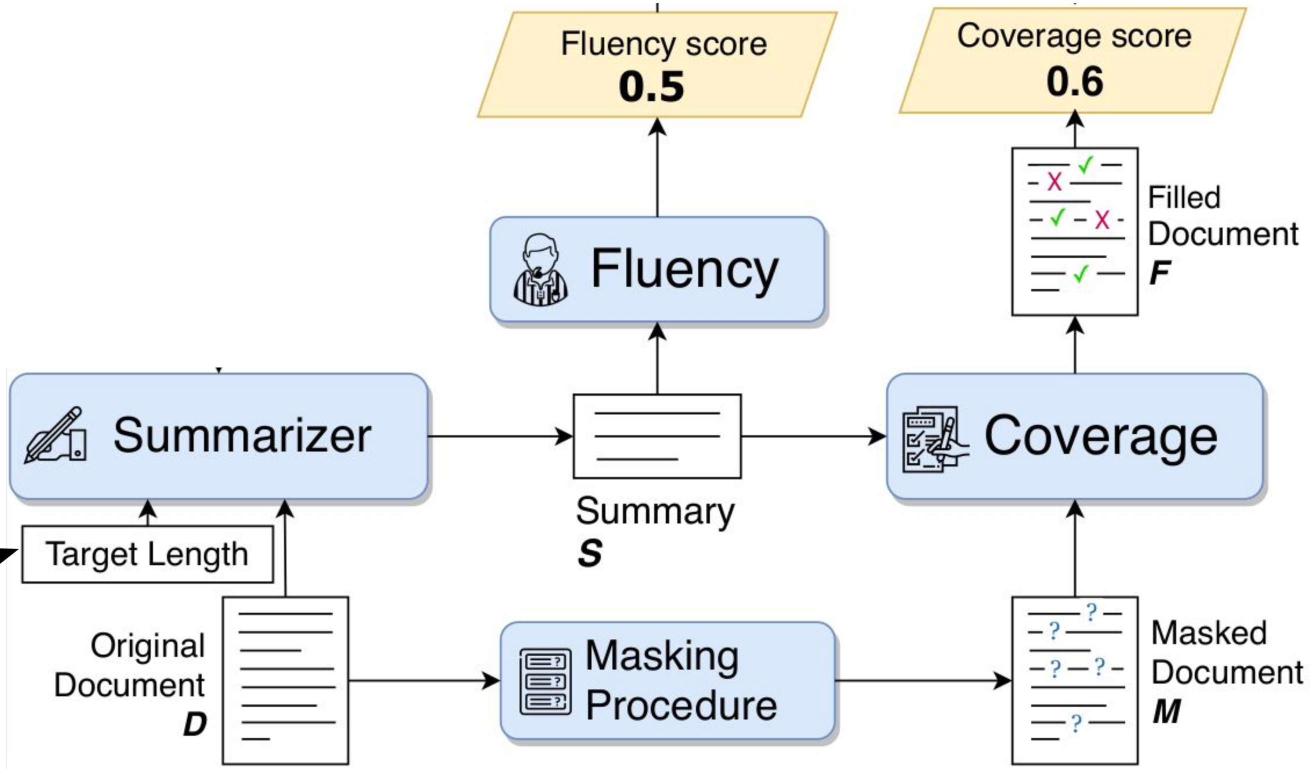
Summary Loop Diagram



Brevity

Unsupervised and abstractive summarization technique

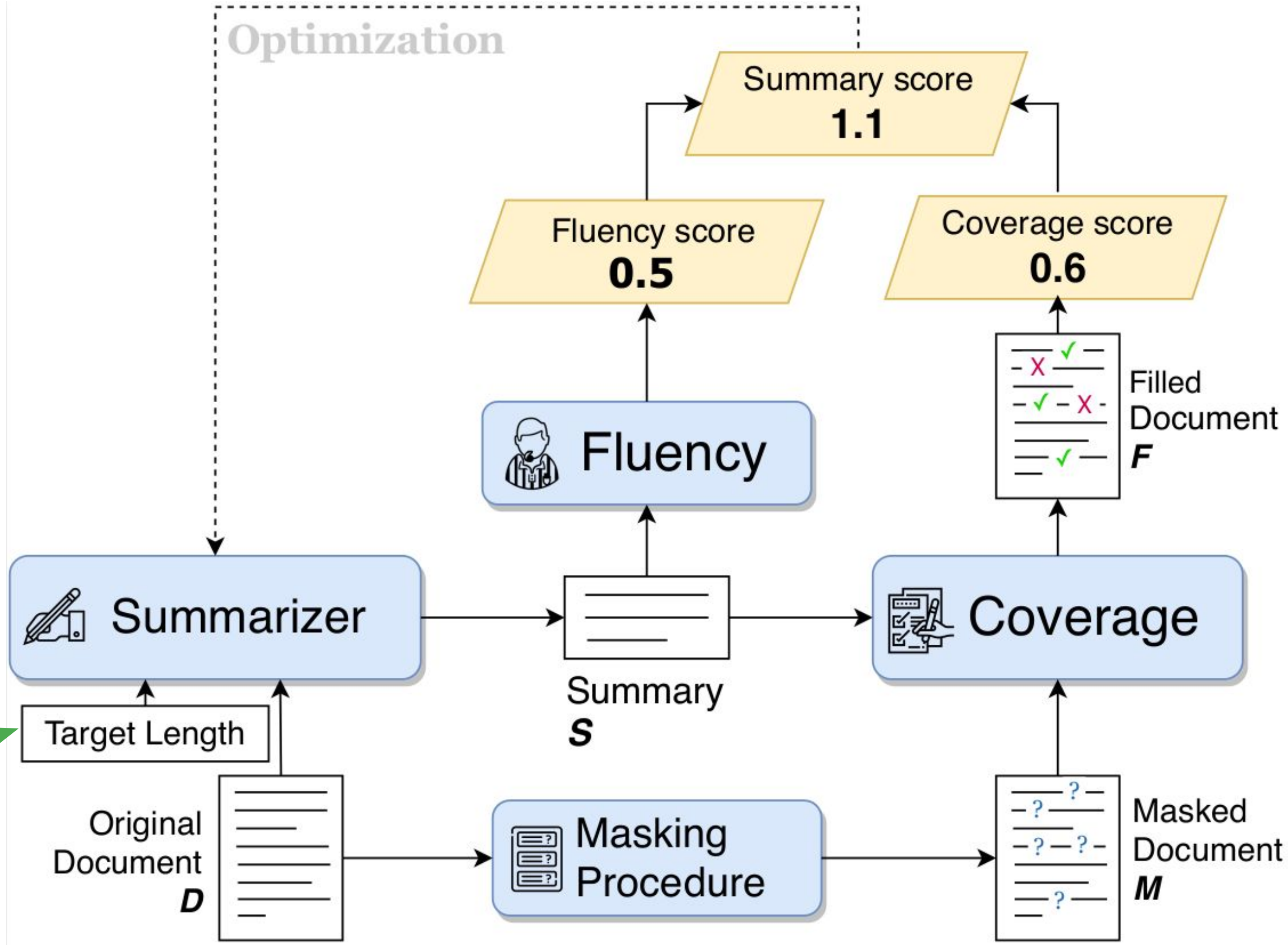
Summary Loop Diagram



Brevity

Unsupervised and abstractive summarization technique

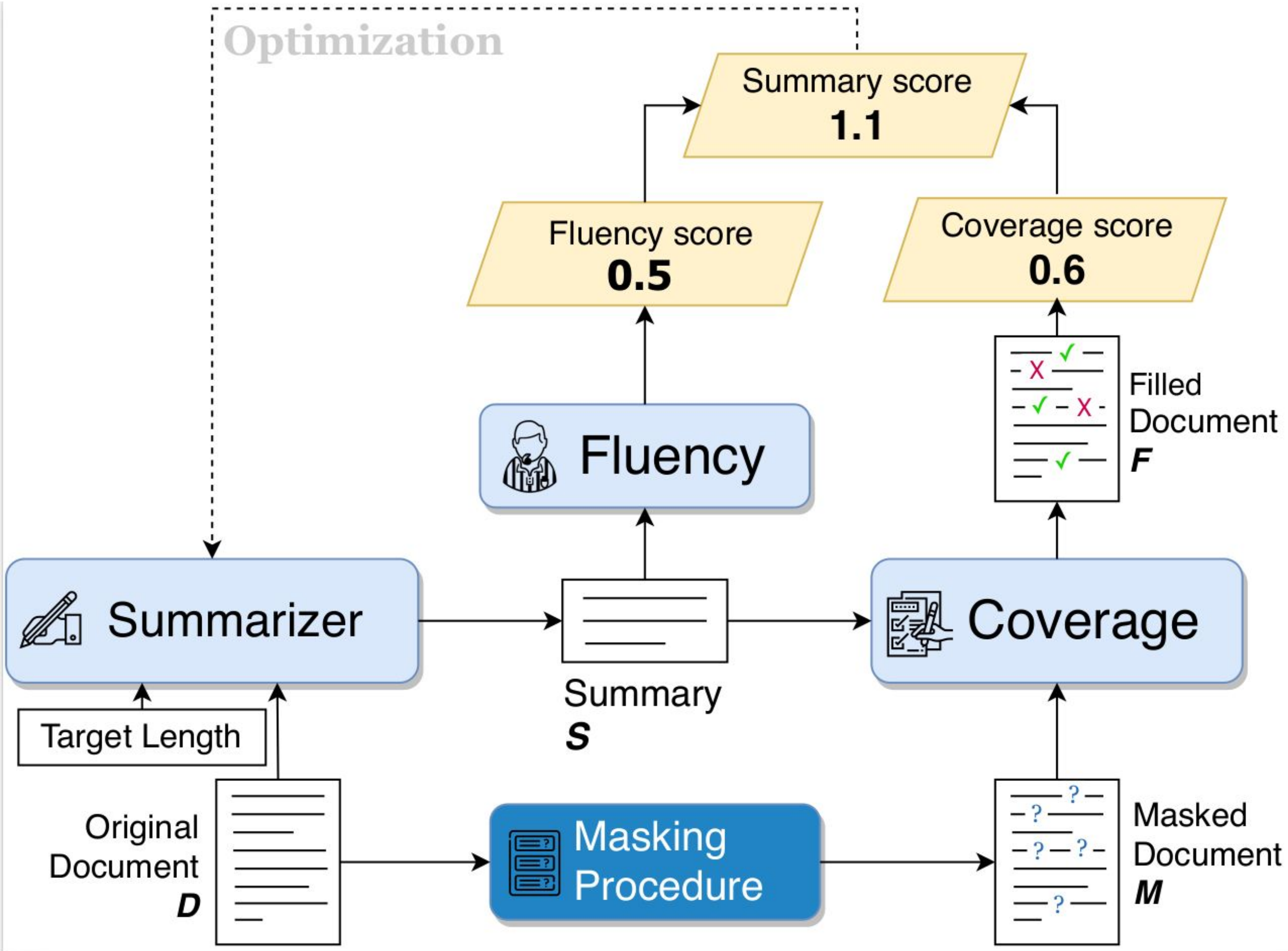
Summary Loop Diagram



Brevity

Unsupervised and abstractive summarization technique

Summary Loop Diagram



First, the masking procedure.

Masking Procedure

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**.

[...]

Compute keywords (unsupervised)

["chile", "president", "protests", "summits", "canceled", …]

Masking Procedure

News article

(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 _____.

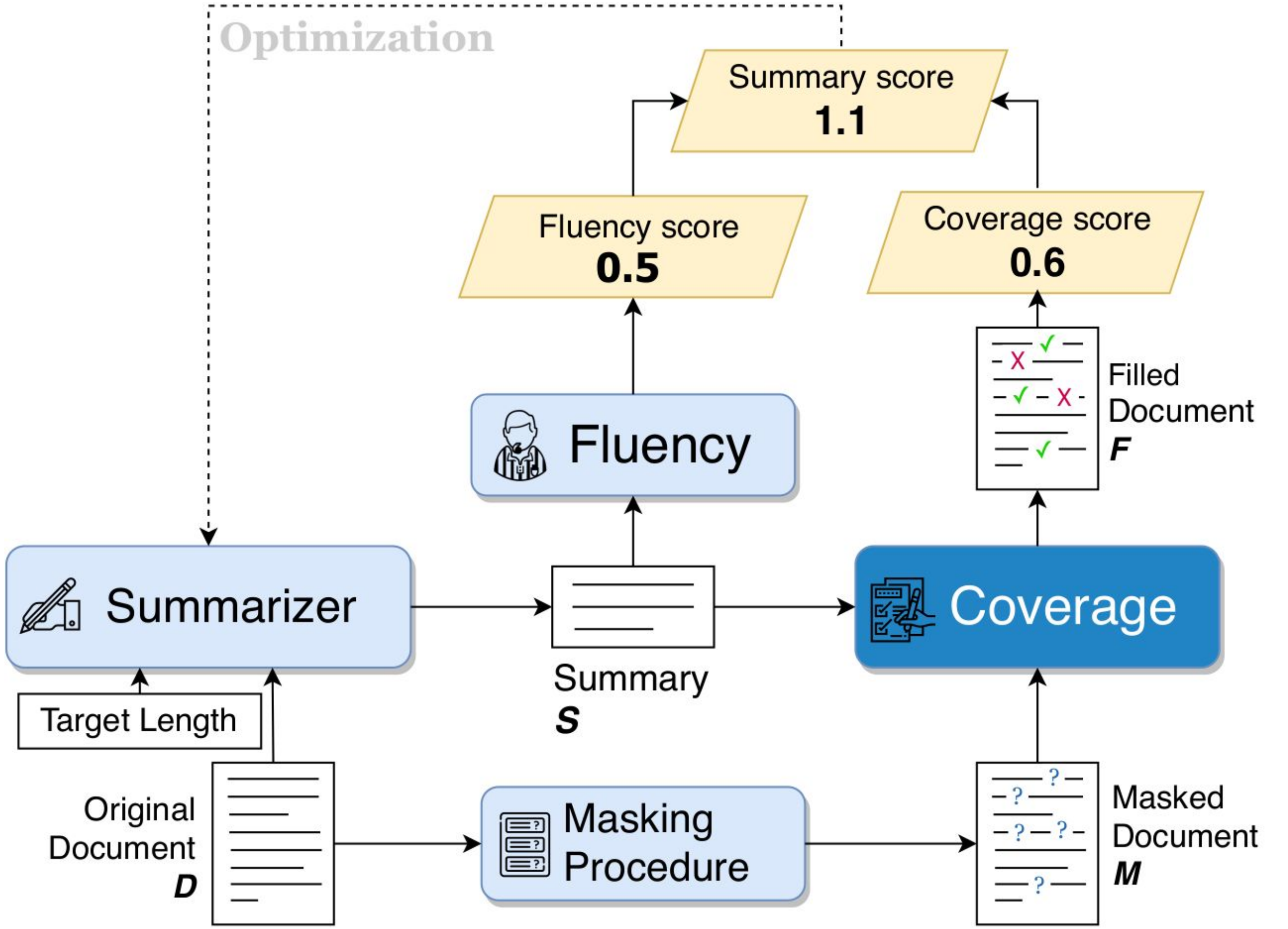
[...]

Blank keywords out.

Important to blank all occurrences.

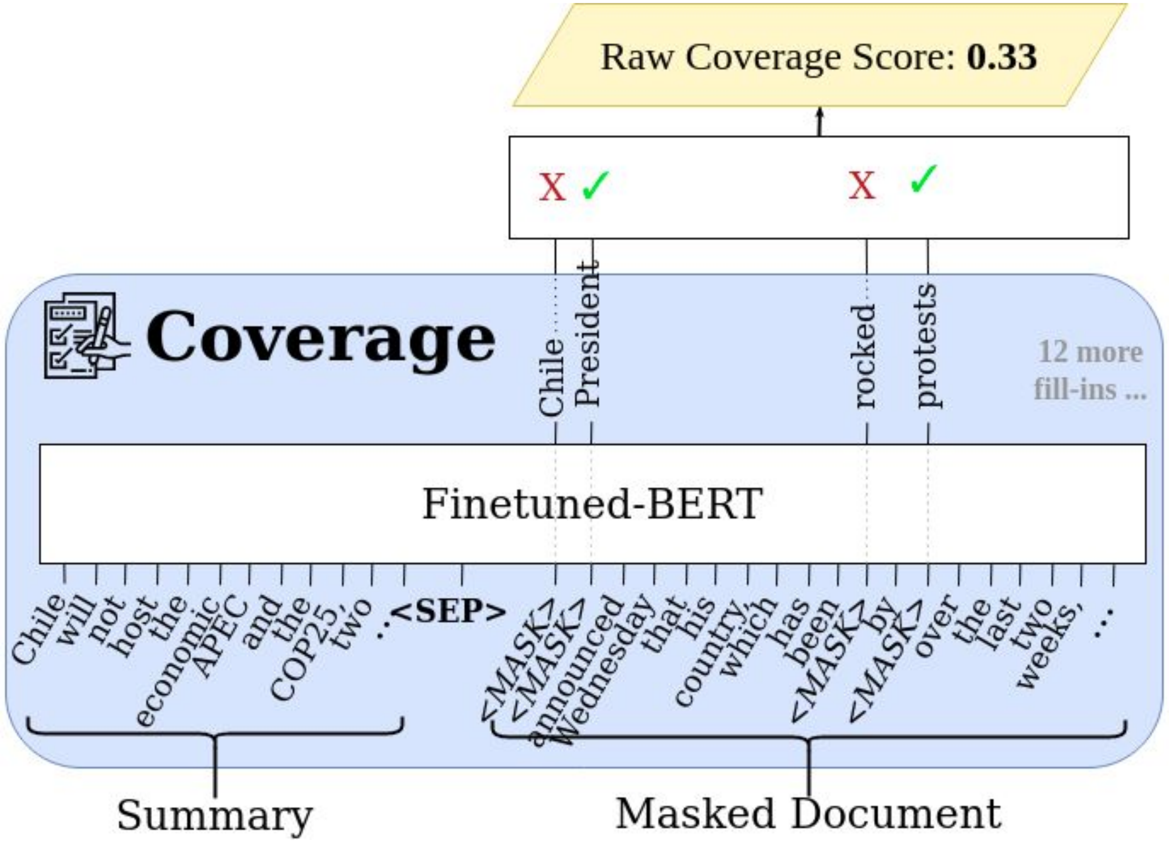
Number of keywords is a hyper-parameter.

Summary Loop Diagram



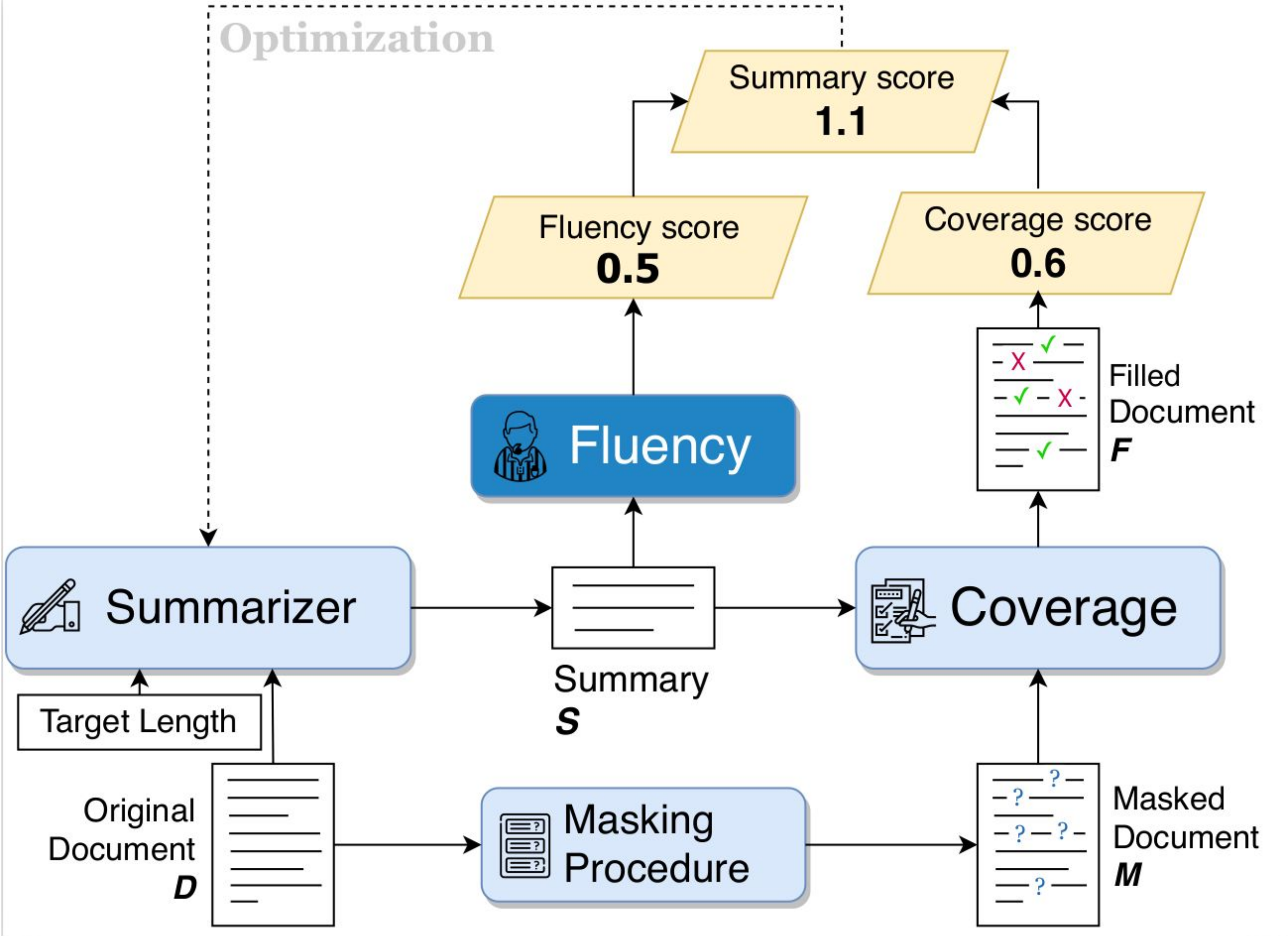
"A summary is a brief, fluent text that **covers** the main points of an original document."

Coverage Model


Raw Coverage Score: **0.33**
X ✓
X ✓
Coverage
Chile
President
rocked
protests
12 more fill-ins ...
Finetuned-BERT
Chile will not host the economic APEC and the COP25, two ..
<SEP>
<MASK> <MASK> announced Wednesday that his country, which has been <MASK> by <MASK> over the last two weeks, ...
Summary
Masked Document

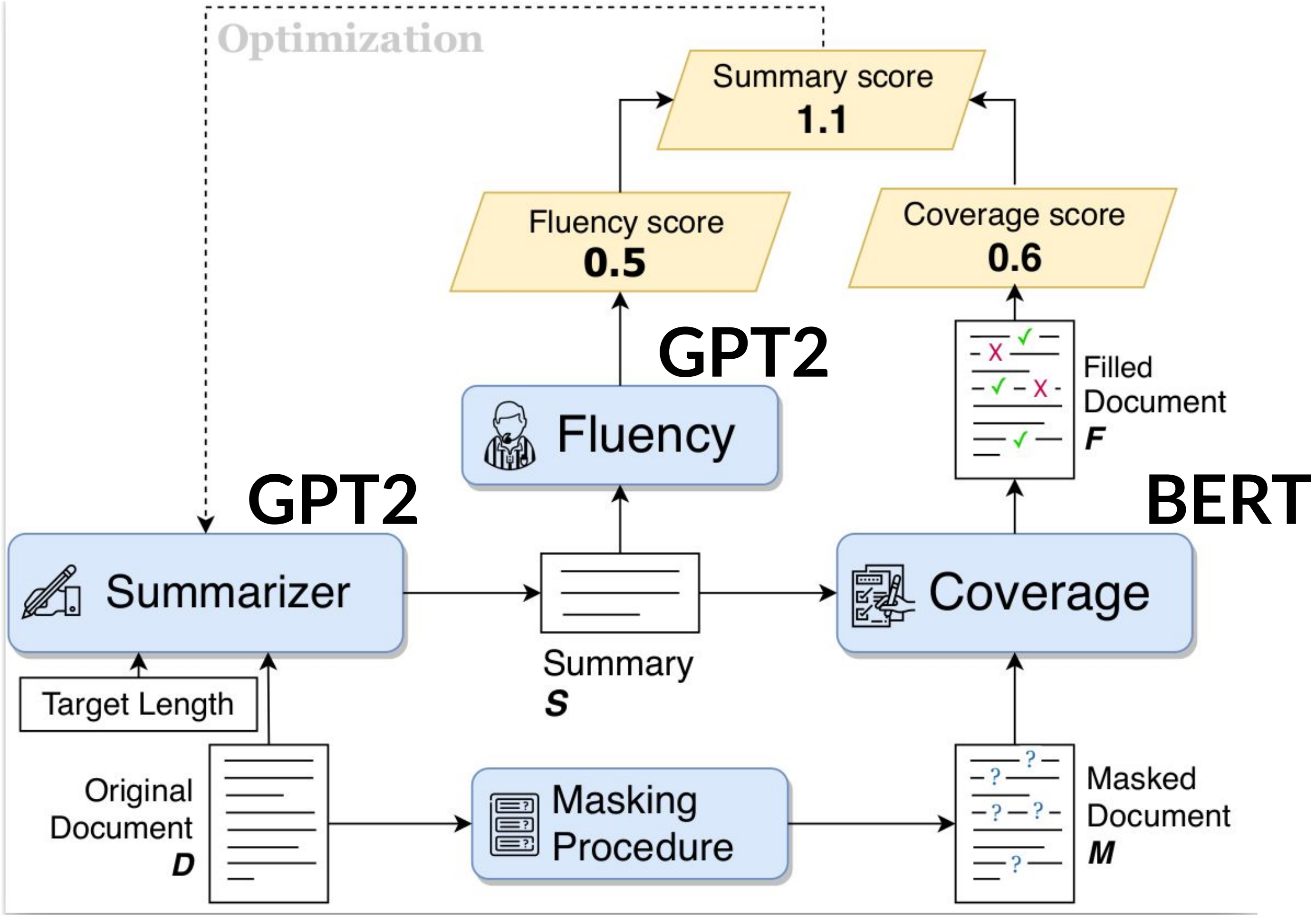
Coverage is the accuracy at recovering the masked keywords, using the summary.

Summary Loop Diagram



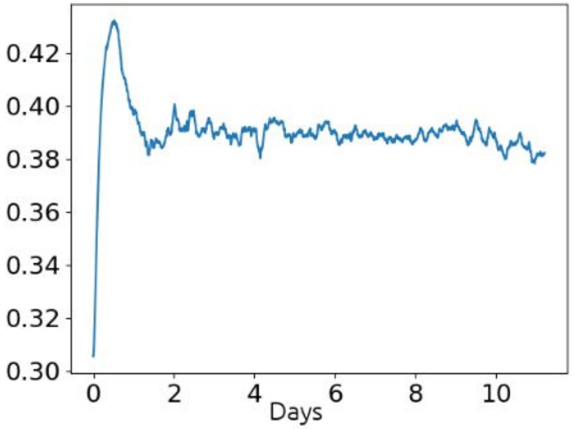
"A summary is a brief, **fluent** text that covers the main points of an original document."

Summary Loop Diagram

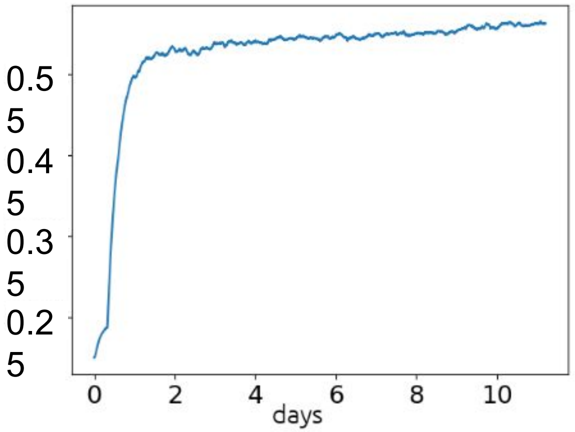


The Summary Loop involves 2 GPT-2 and a BERT model

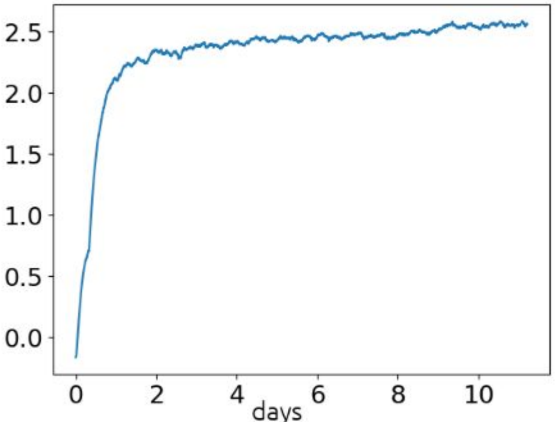
Example Training Run



Fluency Score



Coverage Score



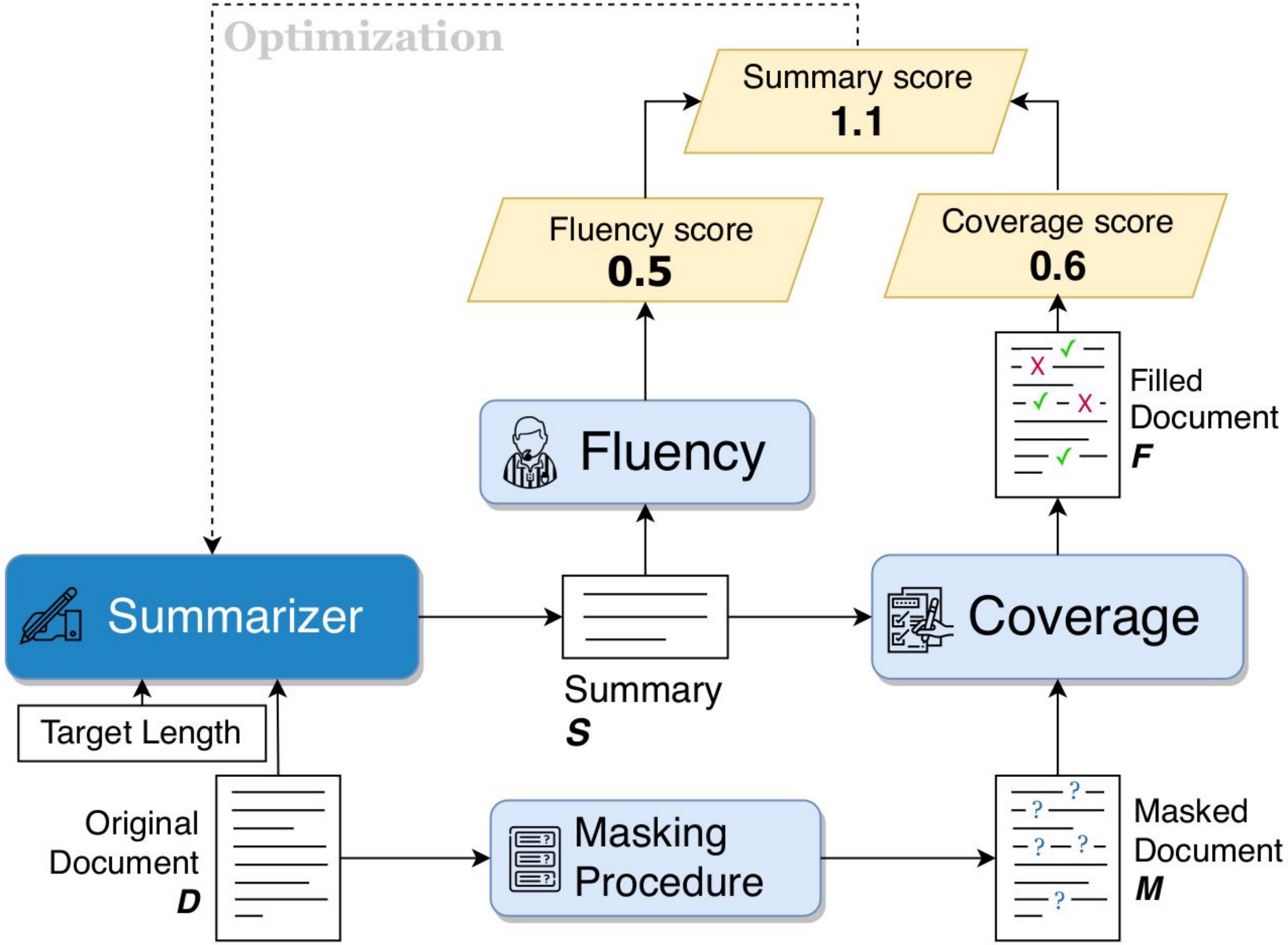
Summary Score

Trained with Self-Critical Sequence Training (SCST)

Target Length: 10 words

Trained on single Titan X GPU

Summary Loop Diagram



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.

[...]

Target Length = 10 words

Pinera cancelled the APEC summit at Santiago.

Coverage Score: 0.22

Target Length = 24 words

Pinera said Chileans have been canceled the hosting of the APEC summit, which was scheduled to take place in November.

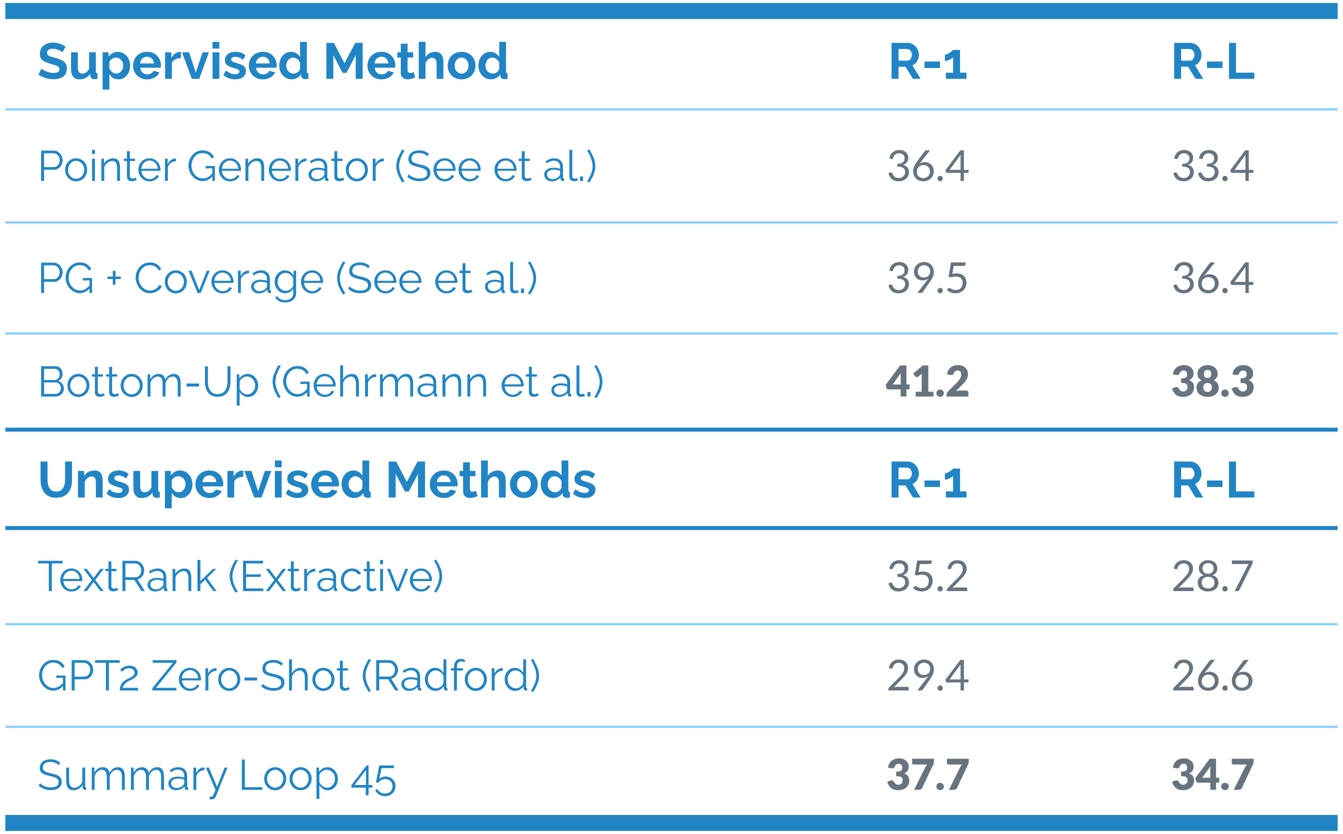
Coverage Score: 0.33

Target Length = 45 words

Sebastian Pinera announced Wednesday that his country will not hold the APEC summit, which was scheduled to take place in Santiago. Pinera said that Chileans had been paralyzed by protests over the last two weeks.

Coverage Score: 0.39

ROUGE Results



Supervised Method	R-1	R-L
Pointer Generator (See et al.)	36.4	33.4
PG + Coverage (See et al.)	39.5	36.4
Bottom-Up (Gehrmann et al.)	**41.2**	**38.3**
Unsupervised Methods	**R-1**	**R-L**
TextRank (Extractive)	35.2	28.7
GPT2 Zero-Shot (Radford)	29.4	26.6
Summary Loop 45	**37.7**	**34.7**

On standard test-set of CNN/DM. Approaching ROUGE of supervised methods **without seeing a single summary example**.

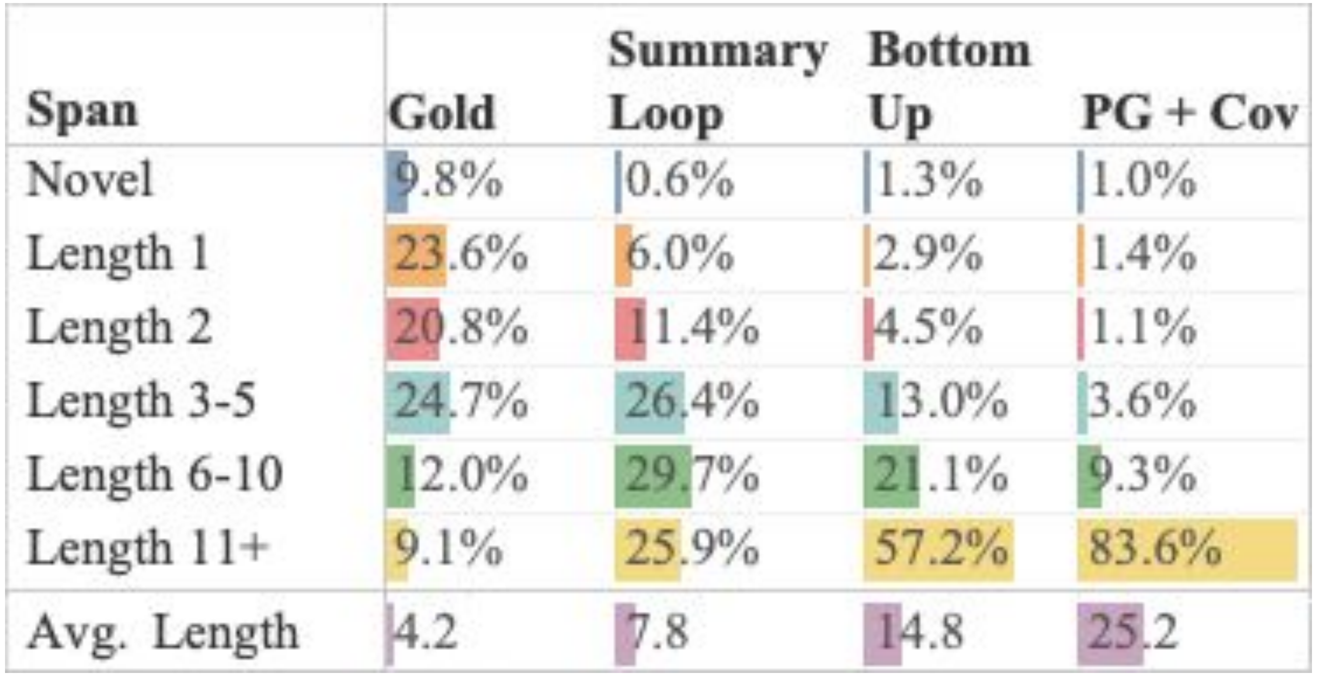
Technique & Error Analysis

Model Type	Point-Gen + Coverage	Bottom Up	Summary Loop
Inaccurate (%)	13%	32%	25%
Ungrammatical (%)	6%	16%	18%
Total Technique Used	148	287	425
Technique Application Success Rate (%) *	52%	47%	57%

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.



Span	Gold	Summary Loop	Bottom Up	PG + Cov
Novel	9.8%	0.6%	1.3%	1.0%
Length 1	23.6%	6.0%	2.9%	1.4%
Length 2	20.8%	11.4%	4.5%	1.1%
Length 3-5	24.7%	26.4%	13.0%	3.6%
Length 6-10	12.0%	29.7%	21.1%	9.3%
Length 11+	9.1%	25.9%	57.2%	83.6%
Avg. Length	4.2	7.8	14.8	25.2

Distribution of lengths of copied spans.
Gold summaries (handwritten) copy shorter passages and use novel words.

Take-Home Message

- 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:

Summary score = αFluency Score + βCoverage

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)

Rennie, S. J., Marcheret, E., Mroueh, Y., Ross, J., & Goel, V. (2017). Self-critical sequence training for image captioning. In *Proceedings of the IEEE CVPR*.

Optimization Procedure: SCST

Directly optimize:

Summary score = αFluency Score + βCoverage

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

Rennie, S. J., Marcheret, E., Mroueh, Y., Ross, J., & Goel, V. (2017). Self-critical sequence training for image captioning. In *Proceedings of the IEEE CVPR.*

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)

<u>Essentially:</u> Increasing likelihood of summary with higher reward, increasing expected reward.

Rennie, S. J., Marcheret, E., Mroueh, Y., Ross, J., & Goel, V. (2017). Self-critical sequence training for image captioning. In *Proceedings of the IEEE CVPR.*