

Text Summarization

By

Emerald Olorunsaiye

2024

Table of Contents

Table of Contents.....	1
1 - Introduction.....	2
1.1 - What is text summarisation?.....	2
1.2 - How can text-summarisation be used?.....	2
2 - Model.....	3
2.1 - Falconsai/text-summarisation.....	3
2.2 - Facebook/bart-large-cnn.....	3
2.3 - philschmid/bart-large-cnn-samsum.....	3
2.4 - google/pegasus-large.....	3
3 - Evaluation.....	4
3.1 - Metrics Used.....	4
3.1.1 - ROUGE.....	4
3.1.2 - BERTScore.....	4
3.1.3 - METEOR.....	4
4 - Experiment.....	5
4.1 - Sample texts.....	5
5 - Results.....	6
5.1 - Falconsai/text-summarisation.....	6
5.2 - Facebook/bart-large-cnn.....	7
5.3 - philschmid/bart-large-cnn-samsum.....	8
5.4 - google/pegasus-large.....	9
Final Evaluation Metric Results.....	10
Results Analysis.....	11
6 - Conclusion.....	12
7 - References.....	12

1 - Introduction

In this experiment, I intend to investigate text summarisation as an Artificial Intelligence tool. I am interested in observing how the models approach different bodies of text and what variables affect the outcome. This includes the dataset used to train the model, who developed it, and how much information a model is trained to include in their final summary. Besides the conclusions that I will draw with my analysis, I will also be using various evaluation metrics for a more quantitative analysis.

1.1 - What is text summarisation?

The dictionary defines a summary as “a short, clear description that gives the main facts or ideas about something”¹ The most important words I take from this definition are “*short*”, “*clear*” and “*main facts*”. These are the three most important factors when looking at a summary.

For a generative artificial intelligence model, there are two ways it can approach this task, Extractive or Abstractive. Both use Natural Language Processing (NLP) to achieve a final product. The extractive approach uses sentence ranking, this method is the more direct of the 2 ‘extracting’ sentences from the text and ranking them based on the words in the sentences. Each word in a sentence is assigned a value depending on how many times that word appears in the text² The abstractive approach utilizes Machine Learning and focuses on crafting a brand-new piece of text that best summarises the original. It crafts a new summary based on what the NLP feels is best for the summary³. This approach however does take up more effort from the model but produces results far better than extracting.

1.2 - How can text-summarisation be used?

Text summarization has many use cases, examples include summarising long articles to get a brief and precise overview, or summarising transcripts of a missed lecture or meeting to quickly catch up on key information, it can also be used to summarise personal notes so a user can easily get an understanding of older notes they hadn’t looked at in a while.

¹ Cambridge Dictionary, n.d. *Summary*. Retrieved May 27, 2024, from <https://dictionary.cambridge.org/dictionary/english/summary>

² Growth Bar, n.d. *How do NLP algorithms work in AI tools to summarize text?* Retrieved June 03, 2024, from <https://www.growthbarseo.com/blog/how-do-nlp-algorithms-work-in-ai-tools-to-summarize-text/>

³ Growth Bar, n.d. *How do NLP algorithms work in AI tools to summarize text?* Retrieved June 03, 2024, from <https://www.growthbarseo.com/blog/how-do-nlp-algorithms-work-in-ai-tools-to-summarize-text/>

2 - Model

2.1 - Falconsai/text-summarisation

2.2 - Facebook/bart-large-cnn

2.3 - philschmid/bart-large-cnn-samsum

2.4 - google/pegasus-large

Have a diagram of how each model works and approaches summarisation

Add a section called model and properly define each model and how they operate

3 - Evaluation

3.1 - Metrics Used

3.1.1 - ROUGE

Recall-Oriented Understudy for Gisting Evaluation (ROUGE). This evaluation metric measures accuracy, precision and recall. There are 4 results you receive when using ROUGE. Rouge 1 (unigram), measures the overlap of individual words, Rouge 2 (bigram) measures the overlap of word pairs, and RougeL and RougeLsum measure the longest matching word sequences. Rouge L focuses on averages while Lsum is computed over a whole summary⁴.

3.1.2 - BERTScore

BERTSCORE is used to measure the quality of text summarisation. This metric looks at how similar the text summary is to the original text. It compares pairs and chunks of texts, combines the precision and recall values and measures the semantic similarity of the two texts, identifying sentences with similar meaning but different phrasing⁵.

3.1.3 - METEOR

The metric for the Evaluation of Translation with Explicit Ordering is based on the harmonic mean of precision and recall. METEOR puts more weight on recall for a summary than precision. METEOR uses unigram-matching, predictions and references to make its comparisons and reach a numerical result. The metric measures how well-ordered the matched words in the summary are in relation to the reference⁶.

⁴ Zilliz, n.d. *Semantic similarity*. Retrieved May 2, 2024, from <https://zilliz.com/glossary/semantic-similarity#:~:text=Semantic%20similarity%20is%20used%20to,merely%20rephrased%20the%20source%20text>.

⁵ Hugging Face, n.d. *Chapter 7: Text Summarization*. Retrieved May 2, 2024, from <https://huggingface.co/learn/nlp-course/chapter7/5>

⁶ Hugging Face, n.d.) *METEOR metric*. Retrieved May 2, 2024, from <https://huggingface.co/spaces/evaluate-metric/meteor>

4 - Experiment

4.1 - Sample texts

All sample texts are within 120-135 words

1: The tower is 324 metres(1,063 ft) tall, about the same height as an 81-storey building, and the tallest structure in Paris. Its base is square, measuring 125 metres (410 ft) on each side. During its construction, the Eiffel Tower surpassed the Washington Monument to become the tallest man-made structure in the world, a title it held for 41 year until the Chrysler Building in New York City was finished in 1930. It was the first structure to reach a height of 300 metres. Due to the addition of a broadcasting aerial at the top of the tower in 1957, it is now taller than the Chrysler Building by 5.2 metres (17ft). Excluding transmitters, the Eiffel Tower is the second tallest free-standing structure in France after the Millau Viaduct.

2: Big Ben is the nickname of the Great Bell inside the Elizabeth Tower, a clock tower at the Palace for Westminster in London. It is one of the most famous landmarks and symbols of the United Kingdom and is parliamentary democracy. The tower was completed in 1859 and the bell was installed in the same year. The clock is very accurate and can be adjusted by adding or removing pennies from the pendulum. The tower is 96 metres tall and has four clock faces, each 7 metres in diameter. The bell weighs 13.7 tonnes and may have been named after Sir Benjamin Hall or Benjamin Caunt, a boxer. The tower was renamed in 2012 to honour Queen Elizabeth II on her Diamond Jubilee.

3: The Colosseum is a famous amphitheatre in Rome, Italy. It was built by the Flavian emperors between 72 and 80 AD, using stone, concrete, and tens of thousands of Jewish slaves. The Colosseum could seat about 5,000 spectators, who came to watch various events, such as gladiator fights, animal hunts and mock naval battles. The Colosseum had complex system of vaults, arches and trap doors, as well as an awning to provide shade. The Colosseum was named after a colossal statue of Emperor Nero that stood nearby. The Colosseum is a symbol of ancient Roman civilization and its culture of entertainment and violence. It is also a testament to the engineering and architectural skills of the Romans. Today, the Colosseum is a popular tourist attraction and a UNESCO World Heritage Site.

5 - Results

Plot a graph showing the different scores

X-axis - models

Y - axis - scores

Diff colours = metric

OR

X-axis - metric

Y - axis - scores

Diff colours - models

5.1 - Falconsai/text-summarisation

Input Text	Summarised Text	Pieces of Text Picked/Directly referenced	Inferences? (added information)
1	Eiffel Tower is 324 metres(1,063 ft) tall, about the same height as an 81-storey building . Its base is square, measuring 125 metres on each side . This is the tallest man-made structure in Paris .	“324 metres (1,063 ft) tall, about the same height as an 81-storey building.” “Its base is square, measuring 125 metres (410 ft) on each side.” “This is the tallest structure in Paris” “man-made”	<i>This is the tallest man-made structure in Paris.</i> — 2 different sections of the text combined by the model
2	Big Ben is the nickname of the Great Bell inside the Elizabeth Tower . It is one of the most famous landmarks and symbols of the United Kingdom . The tower	“Big Ben is the nickname of the Great Bell inside the Elizabeth Tower.” “It is one of the most famous landmarks and symbols of the United	N/A

	was completed in 1859 and the bell was installed in same year .	Kingdom“ “The tower was completed in 1859 and the bell was installed in the same year”	
3	The Colosseum was built by the Flavian emperors between 72 and 80 AD . It was built using stone, concrete, and tens of thousands of Jewish slaves . The 5,000 spectators came to watch various events such as gladiator fights, animal hunts and mock naval battles.	“The Colosseum is a famous amphitheatre in Rome, Italy. It was built by the Flavian emperors between 72 and 80 AD using stone, concrete, and tens of thousands of Jewish slaves "The Colosseum could seat about 5,000 spectators, who came to watch various events, such as gladiator fights, animal hunts and mock naval battles”	N/A

5.2 - Facebook/bart-large-cnn

Input Text	Summarised Text	Pieces of Text Picked/Directly referenced	Inferences? (added information)
1	The tower is 324 metres(1,063 ft) tall, about the same height as an 81-storey building. Its base is square, measuring 125 metres (410 ft) on each side. During its construction the Eiffel Tower surpassed the Washington Monument to become the tallest	"the tower is 324 metres (1,063 ft) tall, about the same height as an 81-storey building” “Its base is square, measuring 125 metres (410 ft) on each side” “During its construction, the Eiffel Tower surpassed the Washington Monument to become the tallest	N/A

	man-made structure in the world.	man-made structure in the world”	
2	The Elizabeth Tower is a clock tower at the Palace for Westminster in London. The tower was completed in 1859 and the bell was installed in the same year. The clock is very accurate and can be adjusted by adding or removing pennies from the pendulum.	“The Elizabeth Tower...A clock tower at the Palace for Westminster in London” “The tower was completed in 1859 and the bell was installed in the same year” “The clock is very accurate and can be adjusted by adding or removing pennies from the pendulum.”	N/A
3	The Colosseum is a famous amphitheatre in Rome, Italy. It was built by the Flavian emperors between 72 and 80 AD. It is a popular tourist attraction and a UNESCO World Heritage Site.	“The Colosseum is a famous amphitheatre in Rome, Italy” “It was built by Flavian Emperors between 72-80 AD” “...is a popular tourist attraction and a UNESCO World Heritage Site”	N/A

5.3 - philschmid/bart-large-cnn-samsum

Input Text	Summarised Text	Pieces of Text Picked/Directly referenced	Inferences? (added information)
1	The Eiffel Tower is 324 metres tall and the second tallest free-standing	“The ... tower is 324 metres tall” “... the second tallest	N/A

	structure in France after the Millau Viaduct. It was the first structure to reach a height of 300 metres.	free-standing structure in France after the Millau Viaduct.” “It was the first structure to reach a height of 300 metres”	
2	Big Ben is the nickname of the Great Bell inside the Elizabeth Tower, a clock tower at the Palace for Westminster in London. The tower was completed in 1859 and the bell was installed in the same year. It was renamed in 2012 to honour Queen Elizabeth II on her Diamond Jubilee.	“Big Ben is the nickname of the Great Bell inside the Elizabeth Tower, a clock tower at the Palace for Westminster in London” “The tower was completed in 1859” “...was renamed in 2012 to honour Queen Elizabeth II on her Diamond Jubilee”	N/A
3	The Colosseum is a famous amphitheatre in Rome, Italy. It was built by the Flavian emperors between 72 and 80 AD. It is a UNESCO World Heritage Site.	“The Colosseum is a famous amphitheatre in Rome, Italy. It was built by the Flavian emperors between 72 and 80 AD” “...a UNESCO World Heritage Site. ”	N/A

5.4 - google/pegasus-large

Input Text	Summarised Text	Pieces of Text Picked/Directly referenced	Inferences? (added information)

1	the Eiffel Tower surpassed the Washington Monument to become the tallest man-made structure in the world, a title it held for 41 year until the Chrysler Building in New York City was finished in 1930	“the Eiffel Tower surpassed the Washington Monument to become the tallest man-made structure in the world, a title it held for 41 year until the Chrysler Building in New York City was finished in 1930.”	N/A
2	Big Ben is the nickname of the Great Bell inside the Elizabeth Tower, a clock tower at the Palace for Westminster in London.	“Big Ben is the nickname of the Great Bell inside the Elizabeth Tower, a clock tower at the Palace for Westminster in London.”	N/A
3	The Colosseum could seat about 5,000 spectators, who came to watch various events, such as gladiator fights, animal hunts and mock naval battles.	“The Colosseum could seat about 5,000 spectators, who came to watch various events, such as gladiator fights, animal hunts and mock naval battles.”	N/A

Final Evaluation Metric Results

METRIC/MODEL	falconsai/text-summarization	facebook/bart-large-cnn	philshmid/bart-large-cnn-samsum	google/pegasus-large
ROUGE-1	'rouge1': 0.234656323610 15176,18830620 68}	'rouge1': 0.24598930481283 424	'rouge1': 0.21993455262686 032	'rouge1': 0.17280040814310 013
ROUGE-2	'rouge2': 0.211965839838	'rouge2': 0.23519862983053	'rouge2': 0.20269456721650	'rouge2': 0.16858438328492

	5749	166,	544	775
ROUGE-L	'rougeL': 0.225142718830 62068,	'rougeL': 0.24402852049910 875,	'rougeL': 0.20226152053075 13	'rougeL': 0.17280040814310 013
ROUG-LSUM	'rougeLsum': 0.225142718830 62068	'rougeLsum': 0.24402852049910 875	'rougeLsum': 0.20226152053075 13	'rougeL': 0.17280040814310 013
BERTSCORE	Similarity between the texts: 0.1147	Similarity between the texts: 0.1368	Similarity between the texts: 0.1242	Similarity between the texts: 0.1153
METEOR	'meteor': 0.125171862264 38953	'meteor': 0.14324070675631 537	'meteor': 0.12886516443388 407	'meteor': 0.08866870712468 576

Results Analysis

Overall, all the models received low scores from each metric for their summarisation performances.

When it came to the ROUGE tests they all averaged a ROUGE-1 score of 0.2183451472982366, a ROUGE-2 score of 0.20461085504263493 an average ROUGE L of 0.21105829175089522 and an average ROUGE LSUM of 0.21105829175089522, all very low scores showing low accuracy, precision and recall in the summaries presented by the models. For BERTSCORE the average score was 0.1228 and for METEOR it was 0.12148661014481868 once again showing the models were unable to produce similar summaries to the original text with precision and especially with recall, indicating a disparity between the summaries and original text in overall quality.

Out of all the models, the Google/pegasus-large model performed the worst while Facebook/bart-large-cnn performed the best. From what I observed with the models I believe this is mainly because of the type of summarisation approach the models took. Facebook/bart-large-cnn took the abstract approach, trying to form brand-new sentences with semantic similarity while Google/pegasus-large took the extraction approach, picking and displaying the sentence it deemed to be the most important.

6 - Conclusion

Artificial Intelligence Summarisation still has some way to go for certain models and approaches, although the abstraction worked the best during this experiment it still produced poor evaluation scores at the end.

The tool has a lot of potential because of the many areas it can be applied and used to improve productivity and make certain tasks easier, however, from my results, I can see that the models are prone to easily making mistakes and producing inaccurate and unreliable pieces of text.

Overall, text summarisation is an extremely useful tool that with development could open many doors and make way for efficient information processing, streamlined communication, faster decision making etc. Moving forward, AI summarization could still transform how we view and extract information.

7 - References

1. **Cambridge Dictionary**. (n.d.). Summary. In *Cambridge Dictionary*. Retrieved May 27, 2024, from <https://dictionary.cambridge.org/dictionary/english/summary>
2. **Growth Bar**. (n.d.). How do NLP algorithms work in AI tools to summarize text? In *Growth Bar*. Retrieved June 03, 2024, from <https://www.growthbarseo.com/blog/how-do-nlp-algorithms-work-in-ai-tools-to-summarize-text/>
4. Zilliz. (n.d.). Semantic similarity. In *Zilliz*. Retrieved May 2, 2024, from <https://zilliz.com/glossary/semantic-similarity#:~:text=Semantic%20similarity%20is%20used%20to,merely%20rephrased%20the%20source%20text.>
5. Hugging Face. (n.d.). Chapter 7: Text Summarization. In *Hugging Face*. Retrieved May 2, 2024, from <https://huggingface.co/learn/nlp-course/chapter7/5>
6. Hugging Face. (n.d.). METEOR metric. In *Hugging Face*. Retrieved May 2, 2024, from <https://huggingface.co/spaces/evaluate-metric/meteor>
7. Hugging Face. (n.d.). pegasus-large. In *Hugging Face*. Retrieved February 27, 2024, from <https://huggingface.co/google/pegasus-large>
8. Hugging Face. (n.d.). text_summarization. In *Hugging Face*. Retrieved February 27, 2024, from https://huggingface.co/Falconsai/text_summarization?library=true
9. Hugging Face. (n.d.). bart-large-cnn. In *Hugging Face*. Retrieved February 27, 2024, from <https://huggingface.co/facebook/bart-large-cnn>
10. Hugging Face. (n.d.). bart-large-cnn-samsum. In *Hugging Face*. Retrieved February 27, 2024, from <https://huggingface.co/philschmid/bart-large-cnn-samsum>