
Survey of different text summarization methods

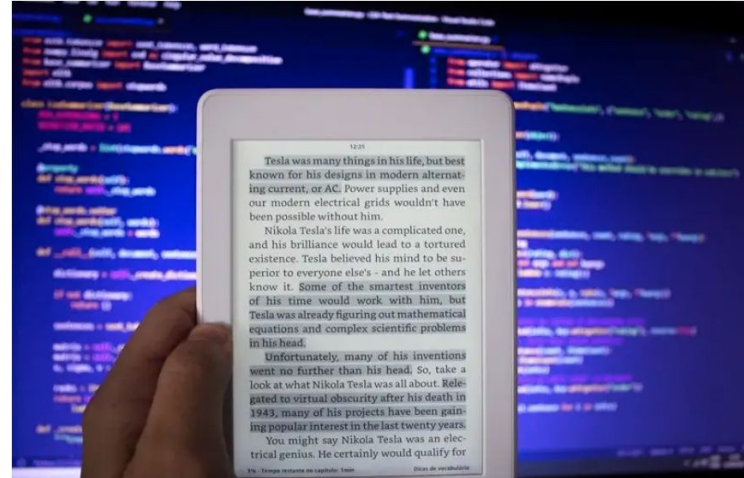
—— Yutao Chen, Ziyue Li, Yichen Liu, Haoxuan Weng ——

Introduction

- Background
- Text Rank
- GPT2
- LSTM
- Evaluation
- Conclusion

Background - What is text summarization?

- There are too many news and articles. We can hardly read all of them.
- Goal:
 - Shorten a long text into a short summary.
 - Keep the general idea of the long text.



Background - The data set we used

CNN-DailyMail News Data Set:

- It provides highlights as references for evaluation.
- The data set covers a wide range of areas.
- Over 300,000 rows of data .
 - We used the testing set, which includes over 10000 data.

index	id	article	highlights
0	92c514c913c0bdfc25341af9fd72b29db544099b	Ever noticed how plane seats appear to be getting smaller and smaller? With increasing numbers of people taking to the skies, some experts are questioning if having such packed out planes is putting passengers at risk. They say that the shrinking space on aeroplanes is not only uncomfortable - it's putting our health and safety in danger. More than squabbling over the arm rest, shrinking space on planes putting our health and safety in danger? This week, a U.S consumer advisory group set up by the Department of Transportation said at a public hearing that while the government is happy to set standards for animals flying on planes, it doesn't stipulate a minimum amount of space for humans. 'In a world where animals have more rights to space and food than humans,' said Charlie Leocha, consumer representative on the committee. 'It is time that the DOT and FAA take a stand for humane treatment of passengers.' But could crowding on planes lead to more serious issues than fighting for space in the overhead lockers, crashing elbows and seat back kicking? Tests conducted by the FAA use planes with a 31 inch pitch, a standard which on some airlines has decreased . Many economy seats on United Airlines have 30 inches of room, while some airlines offer as little as 28 inches . Cynthia Corbertt, a human factors researcher with the Federal Aviation Administration, that it conducts tests on how quickly passengers can leave a plane. But these tests are conducted using planes with 31 inches between each row of seats, a standard which on some airlines has decreased, reported the Detroit News. The distance between two seats from one point on a seat to the same point on the seat behind it is known as the pitch. While most airlines stick to a pitch of 31 inches or above, some fall below this. While United Airlines has 30 inches of space, Gulf Air economy seats have between 29 and 32 inches, Air Asia offers 29 inches and Spirit Airlines offers just 28 inches. British Airways has a seat pitch of 31 inches, while easyJet has 29 inches, Thomson's short haul seat pitch is 28 inches, and Virgin Atlantic's is 30-31.	Experts question if packed out planes are putting passengers at risk . U.S consumer advisory group says minimum space must be stipulated . Safety tests conducted on planes with more leg room than airlines offer .

Background - The general methods we used

- Extractive Text Summarization
 - Pick up the most important and meaningful sentences from the original text.
 - It will cost less computing power.
- Abstractive Text Summarization
 - The model might generate new words and phrases, so this is a harder approach.
 - It might cost more computing power as well.

Background - Evaluation methods

- Three metrics of Rouge library:
 - Rouge 1: Consider 1 - gram
 - Rouge 2: Consider 2 - gram
 - Rouge L: Based on the longest common subsequence (LCS)
- F1 Score:
 - Precision:
 - Recall: $\frac{\text{number_of_overlapping_words}}{\text{total_words_in_reference_summary}}$

$$\frac{2(P * R)}{P + R}$$

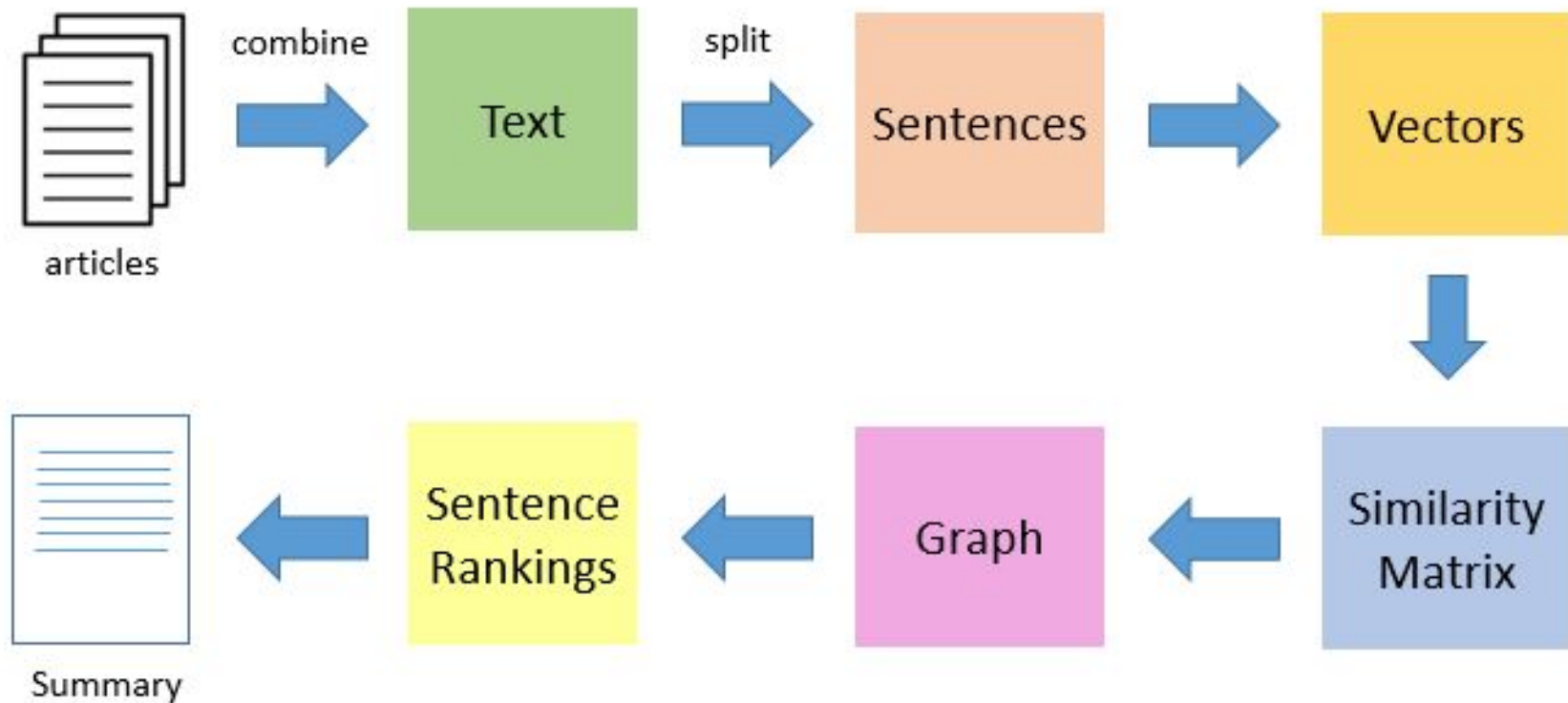
P = the precision

R = the recall of the classification model

Textrank algorithm

What is Textrank?

- Textrank is an **extractive** technique in text summarization, and it is also an **unsupervised learning** algorithm.
- Textrank is a ranking model for graph-based text processing.
- Textrank does **not** require **training corpora**.



Textrank algorithm - See a result of Textrank and compare with original highlights

Highlights	Textrank
<p>Drunk teenage boy climbed into lion enclosure at zoo in west India .</p> <p>Rahul Kumar, 17, ran towards animals shouting 'Today I kill a lion!'</p> <p>Fortunately he fell into a moat before reaching lions and was rescued .</p>	<p>Next level drunk: Intoxicated Rahul Kumar, 17, climbed into the lions' enclosure at a zoo in Ahmedabad and began running towards the animals shouting 'Today I kill a lion!' Mr Kumar had been sitting near the enclosure when he suddenly made a dash for the lions, surprising zoo security.</p>

The score of Textrank

	Rouge_1 (Unigram)	Rouge_2 (Bigram)	Rouge_L	Average
Recall (r)	0.52	0.22	0.48	0.41
Precision (p)	0.25	0.10	0.23	0.19
F-1 Score (f)	0.31	0.12	0.29	0.24

Abstractive Text Summarization

Abstractive Text Summarization is the task of generating a short and concise summary that captures the salient ideas of the source text. The generated summaries potentially contain new phrases and sentences that may not appear in the source text.

- GPT2
- Long Short-Term Memory (LSTM)

GPT2

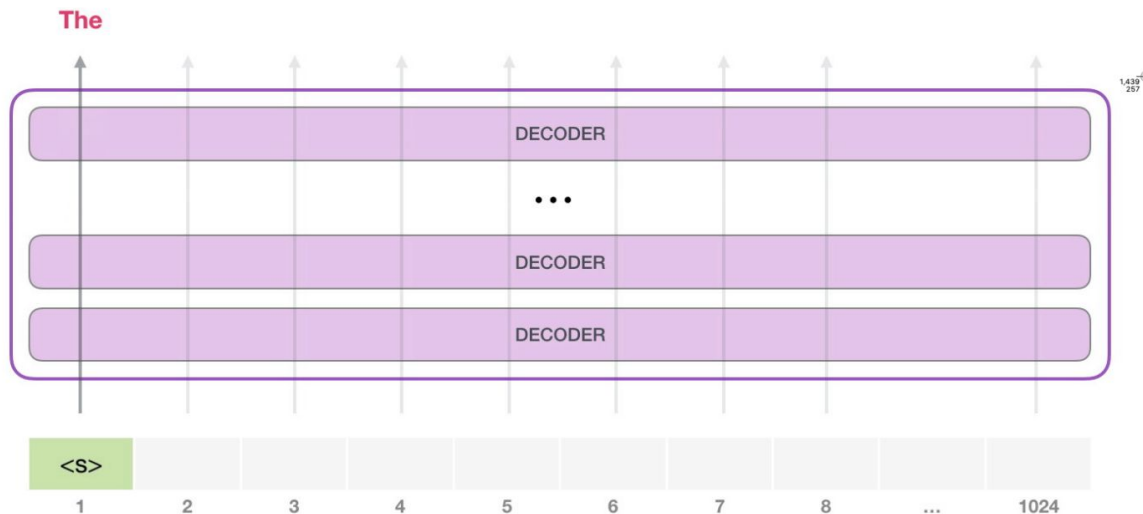
- GPT (Generative Pre-trained Transformer), GPT2 has around 1 billion parameters
- OpenAI, one of the pioneers in AI research, created and trained GPT models.
- GPT-3.5 is the most recent version in 12/1/2022.

GPT2

- The Seq2Seq architecture with Transformers is quite popular for difficult natural language processing tasks, like machine translation or text summarization.
- It uses multi-headed masked self-attention, which allows it to look at only the first i tokens at time step t , and enables them to work like traditional uni-directional language models.
- Decoder-only language models like GPT were not only trained to continue texts, but also they don't extract the idea of the given texts.

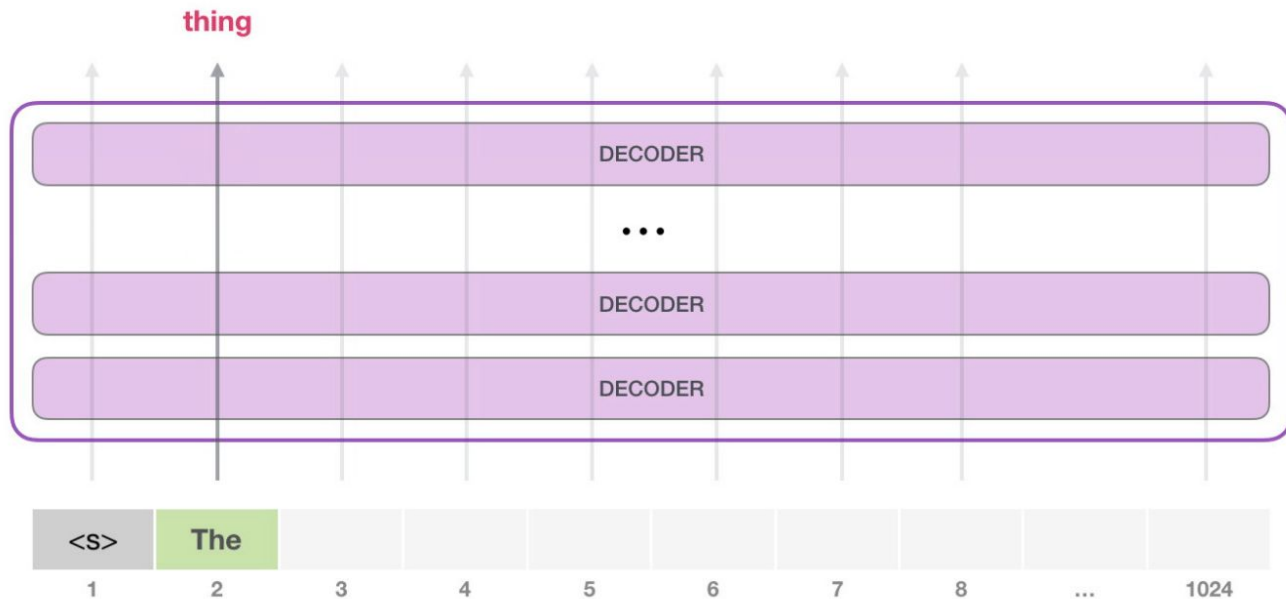
GPT2

- Using `<s>` as its start token.
- The model only has one input token, so that path would be the only active one. The token is processed successively through all the layers, then a vector is produced along that path. That vector can be scored against the model's vocabulary. In this case we selected the token with the highest probability, 'the'. GPT-2 has a parameter called top-k that we can use to choose for output word

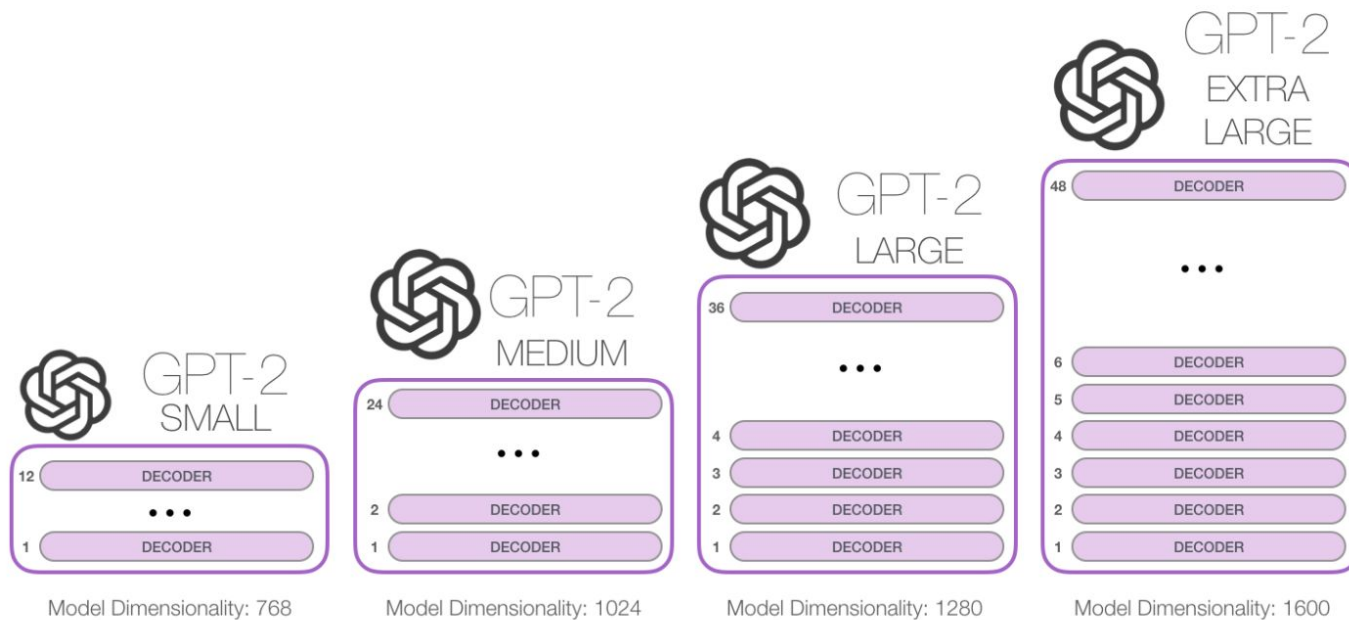


GPT2

- The way these models actually work is that after each token is produced, that token is added to the sequence of inputs. And that new sequence becomes the input to the model in its next step. The model will make the next prediction.



GPT2



The Score of GPT2

	Rouge_1 (Unigram)	Rouge_2 (Bigram)	Rouge_L
Recall (r)	0.25254	0.08338	0.209819
Precision (p)	0.18408	0.046636	0.15537
F-1 Score (f)	0.18789	0.05476	0.15707

Why not vanilla RNN?

Unidirectional LSTM V.S. Vanilla RNN

Vanilla RNN: Vanishing gradient problem!

Models weights are only updated with respect to near effects not long-term effects

Language model task: The writer of the books ____ (Is | Are)

Correct answer: The writer of the books **is** planning a sequel. Vanishing gradient problem

• **Syntactic recency:** The writer of the books is (correct)

• **Sequential recency:** The writer of the books are (incorrect)



Rnns are better at learning **sequential recency** than syntactic recency.

Long short term memory(LSTM)

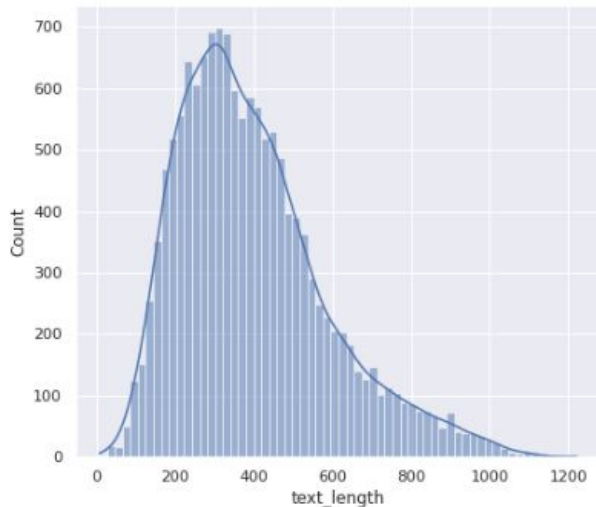
- A type of RNN proposed by Hochreiter and Schmidhuber in 1997 as a solution to the vanishing gradients problem.
 - On step t , there is a hidden state $h^{(t)}$ and a cell state $c^{(t)}$
 - Both are vectors length n
 - The cell stores long-term information
 - The LSTM can erase, write and read information from the cell
- ❖ Forget Gate
 - ❖ Input Gate
 - ❖ Output gate

Then why bidirectional?

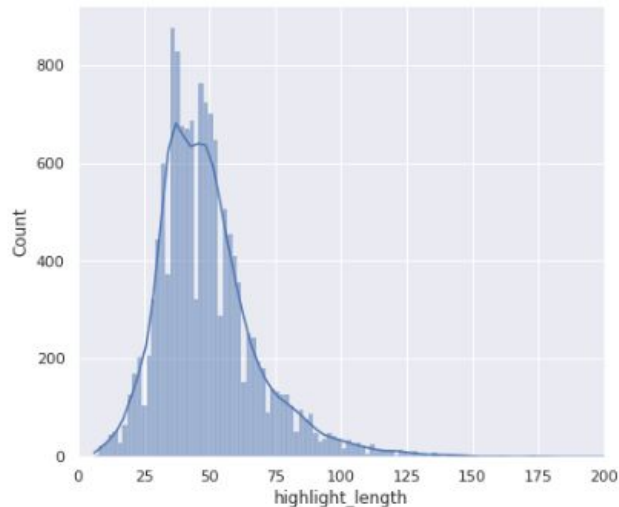
Hidden state in the timestamp t is computed by the previous hidden state $t-1$ and current input

E.g: I am terribly exciting

bidirectional- LSTM



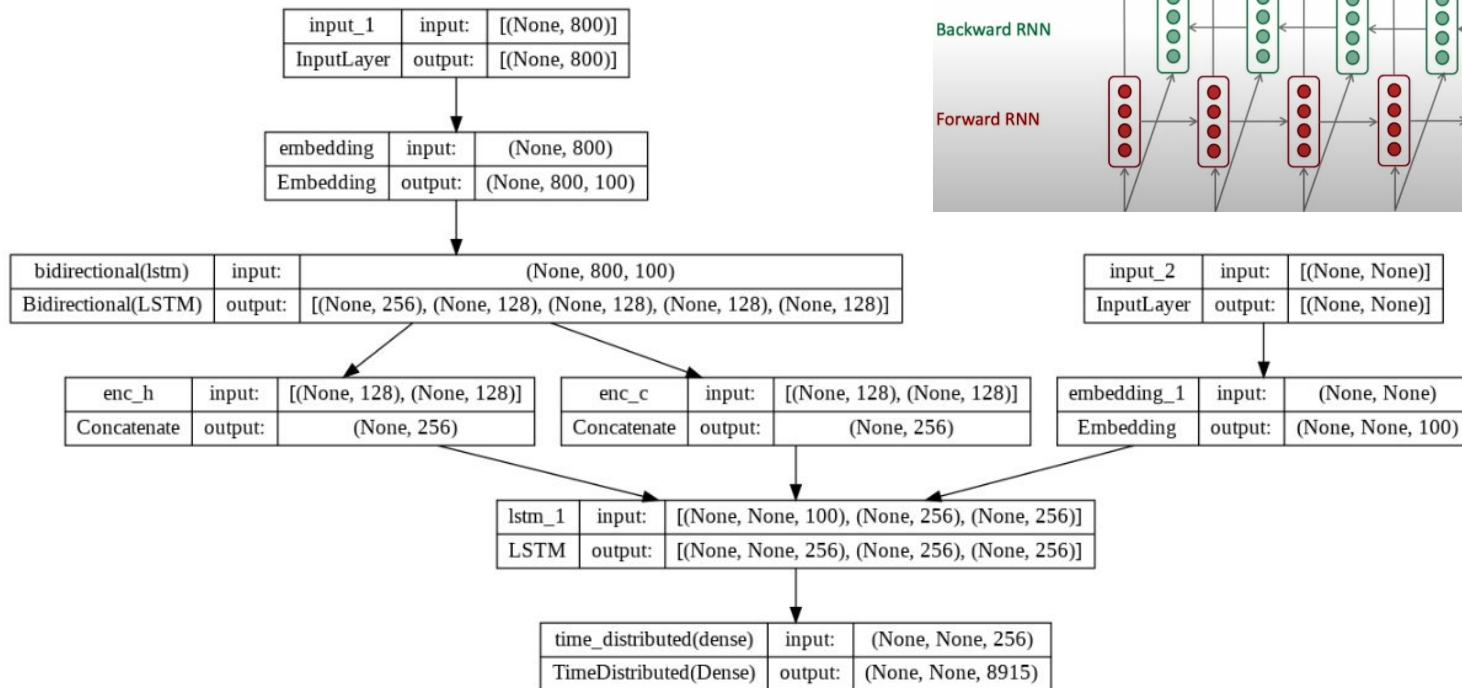
Max text words length: 800 - - - - > 95%
Max summary words length: 100 - - - - > 97%



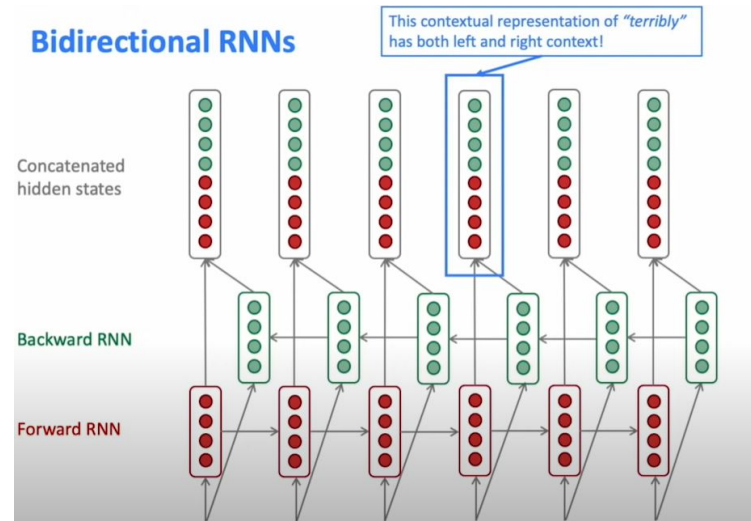
Limited computational capacity!

100 dimension pretrained Glove to generate a word embedding

Model Architecture

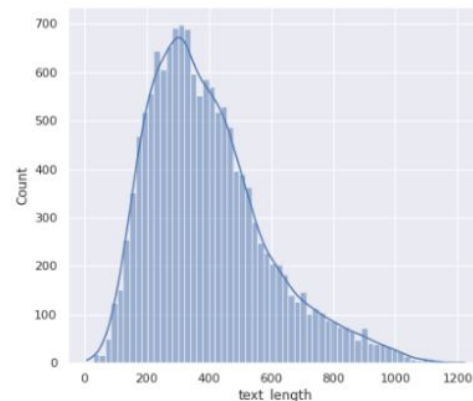
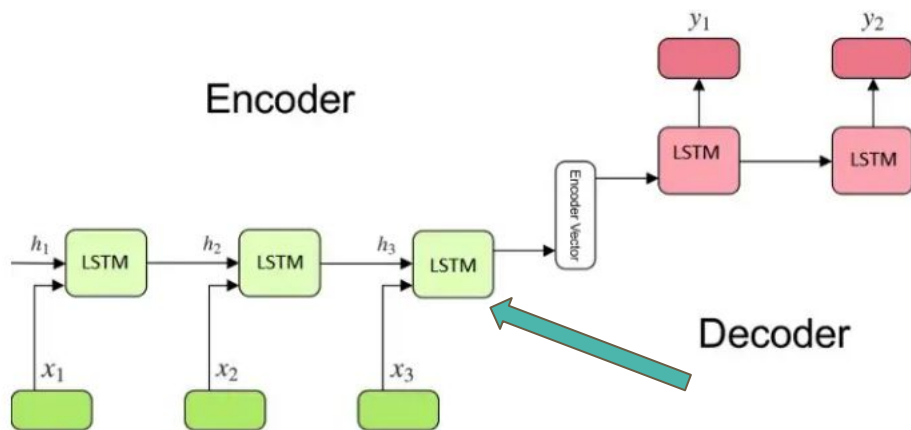


Bidirectional RNNs



Attention is what you need

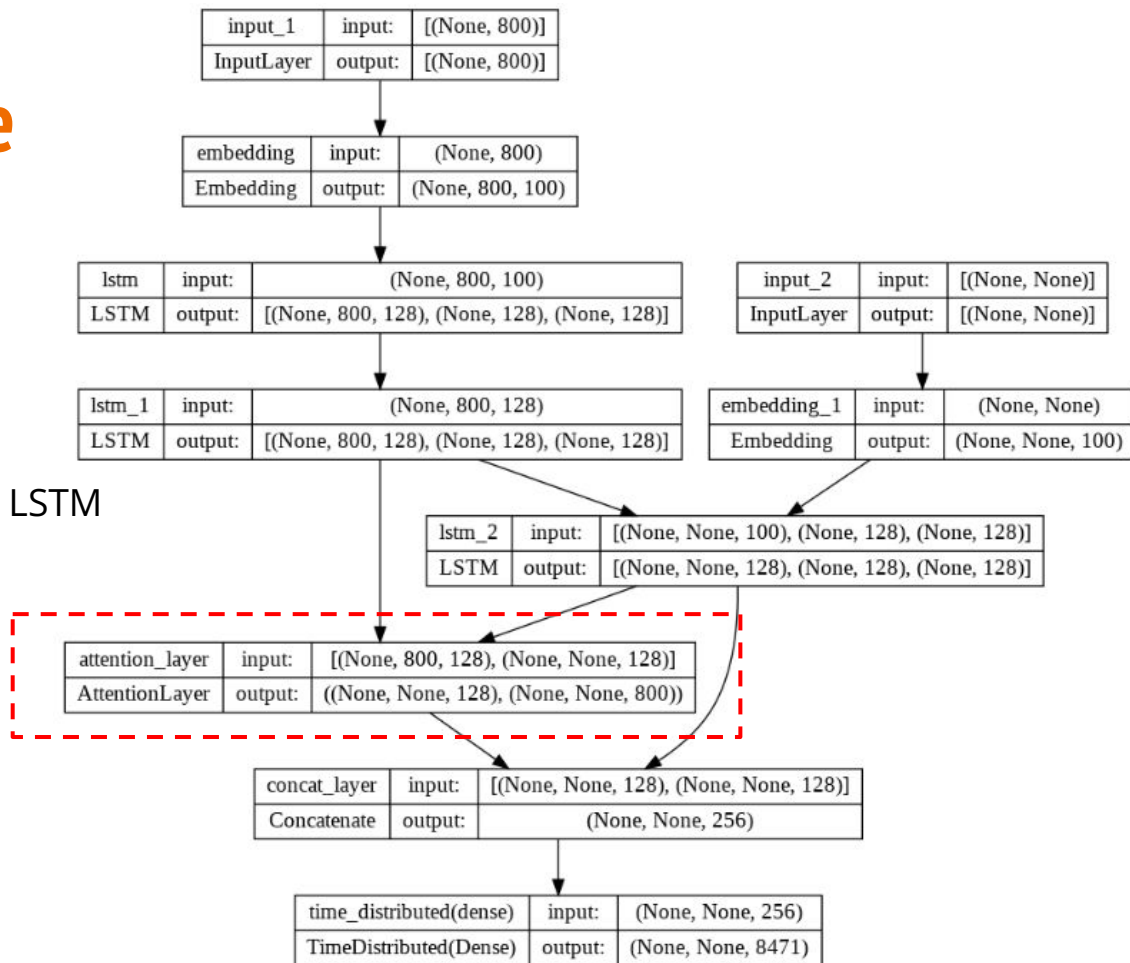
sequence - to-sequence information bottleneck problem



Add an attention layer !

Encoding of the source sentence. This final hidden state needs to capture all information about the source sentence.

Model architecture



- ❖ High performance multi-layer LSTM
- ❖ Attention layer

Conclusion

	Rouge_1 (Unigram)	Rouge_2 (Bigram)	Rouge_L
F-1 Score (f): TextRank	0.31	0.12	0.29
F-1 Score (f): GPT2	0.19	0.05	0.16
F-1 Score (f): LSTM	0.09	0.005	0.08
F-1 Score (f): LSTM+attention	0.12	0.012	0.09

Conclusion

- Overall, the extractive method out-perform abstractive summarizer in the ROUGE-score metric.
 - Since the ROUGE metric only look at identical n-grams; therefore it would ignore the paraphrasings.
 - The abstractive method requires more computing resources.
- In the future:
 - These models could be improved if we have more training data and more computing resources.
 - We could use more subjective method to evaluate the result.

THANK YOU !

References

<https://medium.com/analytics-vidhya/text-summarization-using-bert-gpt2-xlnet-5ee80608e961>

<https://towardsdatascience.com/understand-text-summarization-and-create-your-own-summarizer-in-python-b26a9f09fc70>

<https://www.kaggle.com/datasets/gowrishankarp/newspaper-text-summarization-cnn-dailymail?resource=download>

<https://www.analyticsvidhya.com/blog/2018/11/introduction-text-summarization-textrank-python/>

<https://blog.paperspace.com/generating-text-summaries-gpt-2/>

<https://jalammar.github.io/illustrated-gpt2/>

<https://www.youtube.com/watch?v=QEw0qEa0E50&list=PLoROMvodv4rOhcuXMZkNm7j3fVwBBY42z&index=8>

<https://www.freecodecamp.org/news/what-is-rouge-and-how-it-works-for-evaluation-of-summaries-e059fb8ac840/>

<https://vincentk1991.github.io/GPT2-summarizer/>

<https://web.eecs.umich.edu/~mihalcea/papers/mihalcea.emnlp04.pdf>

<https://towardsdatascience.com/text-summarization-with-gpt2-and-layer-ai-599625085d8e>