Automatic Text Summarization

Abeer Ahmad (38)
Mohammed Deifallah (59)

Outline

- What Is Text Summarization?
- Text Summarization Techniques Classification
- Our Target
- Related Work
- Dataset
- Evaluation Metric
- Model Architecture
- Training
  - Hyperparameters
  - Loss Function
- Evaluation
- Best Result
- How to Extend?

What Is Text Summarization?

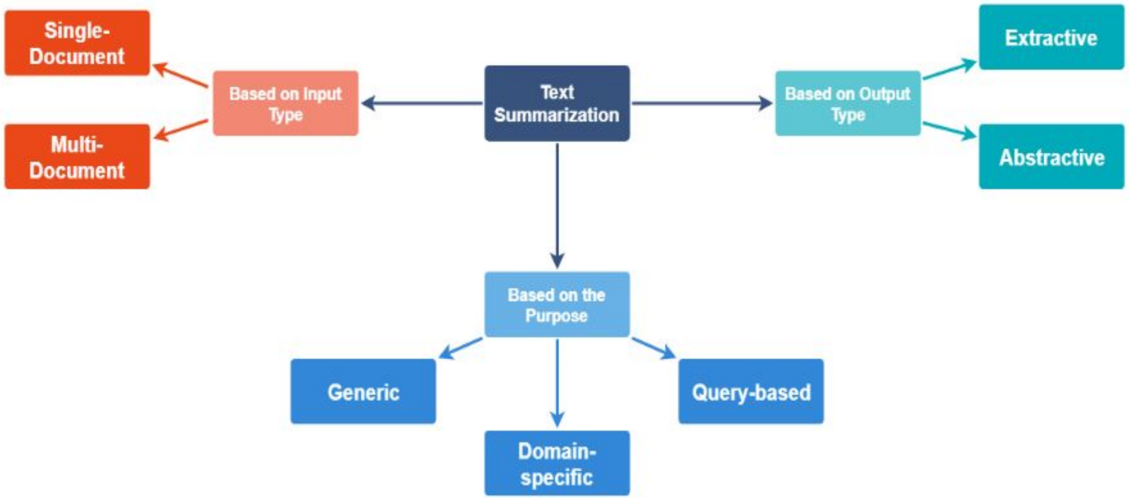
Text summarization is the process of creating a short and coherent version of a longer document; since reading the whole corpus or summarizing it manually is absolutely a tedious job and waste of both effort and time. The main idea of summarization is to capture a subset of data that contains the “information” of the entire set.

Outline

- What Is Text Summarization?
- Text Summarization Techniques Classification
- Our Target
- Related Work
- Dataset
- Evaluation Metric
- Model Architecture
- Training
  - Hyperparameters
  - Loss Function
- Evaluation
- Best Result
- How to Extend?

Text Summarization Techniques Classification

Text summarization methods can be classified into 3 different types:



Text Summarization Techniques Classification

1. Based on Input Type

1. ***Single Document*** _ input length is short.
2. ***Multi Document*** _ input can be arbitrary long.

Text Summarization Techniques Classification

2. Based on Purpose

1. ***Generic*** _ model makes no assumption about the domain of input text - most common so far.
2. ***Domain-Specific*** _ model uses domain-specific knowledge to form a more accurate summary.
3. ***Query-Based*** _ summary only contains information that answers natural language questions about input text.

Text Summarization Techniques Classification

3. Based on Output Type

1. ***Extractive*** _ important sentences are selected from input text to form a summary- most approaches today are extractive in nature.
2. ***Abstractive*** _ model forms its own phrases and sentences to offer a more coherent summary - more appealing, but much more difficult than extractive summarization.

Outline

- What Is Text Summarization?
- Text Summarization Techniques Classification
- Our Target
- Related Work
- Dataset
- Evaluation Metric
- Model Architecture
- Training
  - Hyperparameters
  - Loss Function
- Evaluation
- Best Result
- How to Extend?

Our Target

In our problem, we followed the single-document generic abstractive approach.

Outline

- What Is Text Summarization?
- Text Summarization Techniques Classification
- Our Target
- Related Work
- Dataset
- Evaluation Metric
- Model Architecture
- Training
  - Hyperparameters
  - Loss Function
- Evaluation
- Best Result
- How to Extend?

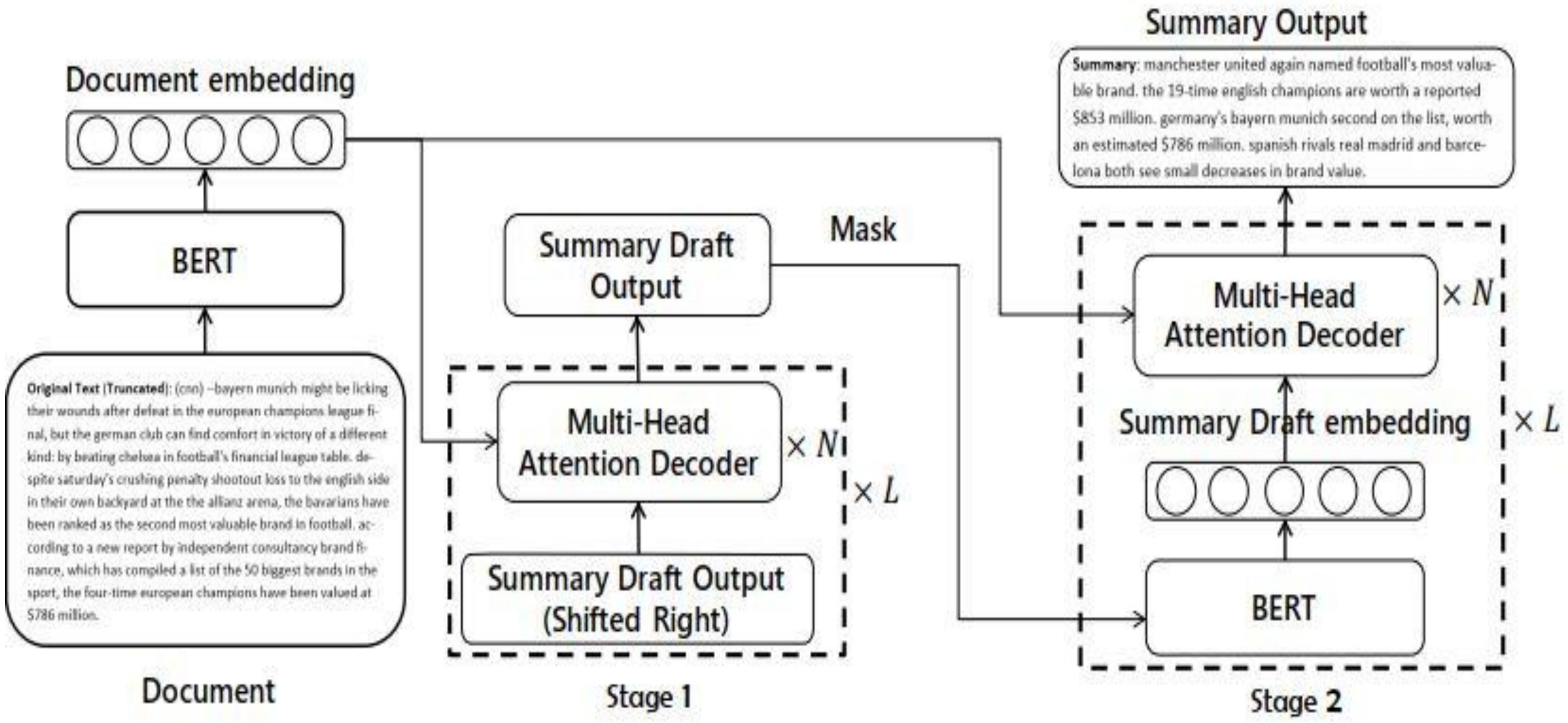
Related Work

Now, we showcase some state-of-the-art models trained to tackle such a problem...

Related Work

1. Pretraining-Based Natural Language Generation for Text Summarization

1. Published in February, 2019.
2. Used *DMQA* and *NYT* Datasets.
3. Following the **Abstractive** approach.
4. Tried to solve: **Left-context-only** decoder and not utilizing the **Pretrained contextualized language models.**
5. Evaluated the model with **ROUGE-1, ROUGE-2, ROUGE-L** and **ROUGE-AVG**.

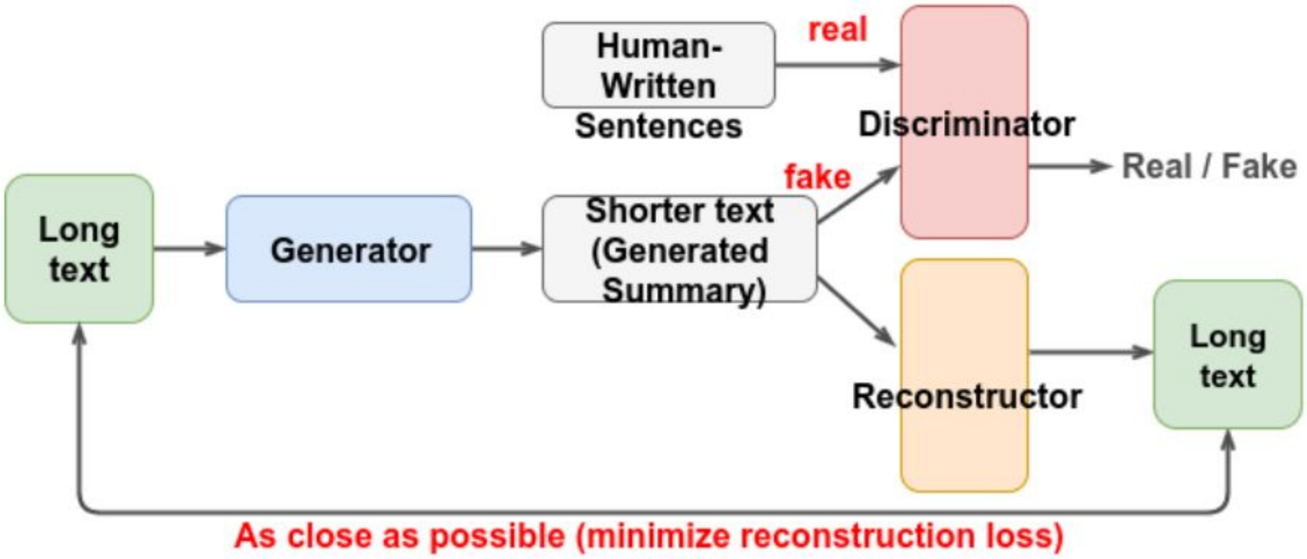


1. Pretraining-Based Model

Related Work

2. Generative Adversarial Network for Abstractive Text Summarization

1. Published in Nov, 2017.
2. Used *DMQA* Dataset.
3. Following the **Abstractive** approach.
4. Tried to solve: **Trivial & Generic summary, Limited grammaticality & readability** and **MLE Shortcomings**.
5. Evaluated the model with **ROUGE-1, ROUGE-2, ROUGE-L**.

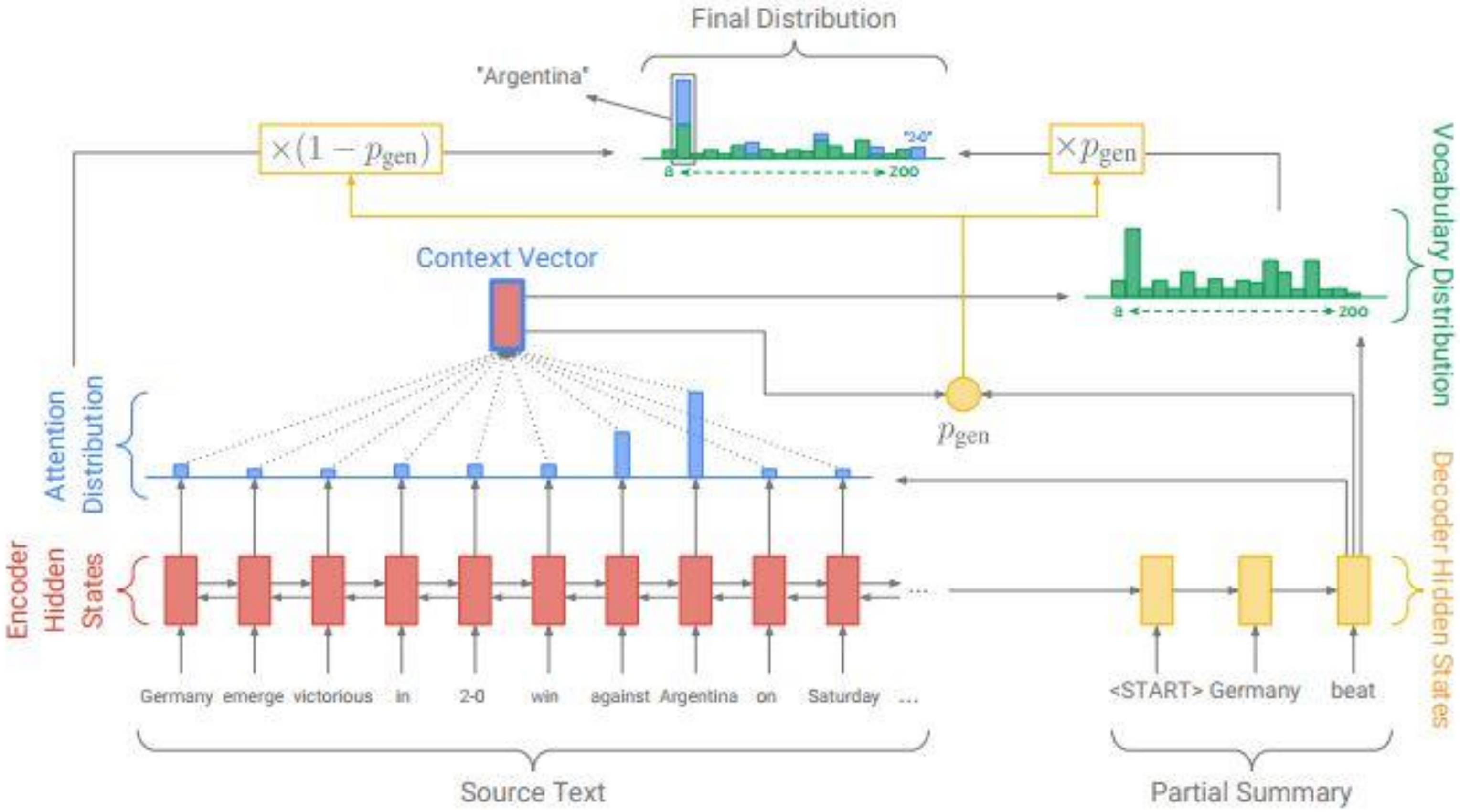


2. GAN Model - [source]

Related Work

3. Pointer-Generator Networks

1. Published in April, 2017.
2. Used *DMQA* Dataset.
3. Tried to solve: ***Factual errors***, ***OOV*** (Out-of-Vocab) words and ***Repeating***.
4. Evaluated the model with **ROUGE-1**, **ROUGE-2**, **ROUGE-L** and **METEOR**.



3. Pointer-Generator Model

State-of-the-Art Results

	ROUGE-1	ROUGE-2	ROUGE-L
Pretraining-Based Model (2019)	**41.71**	**19.49**	**38.79**
GAN Model (2018)	39.92	17.65	36.71
Pointer-Generator Model (2017)	39.53	17.28	36.38

- These results are all based on text summarization models trained over the CNN dataset.

Outline

- What Is Text Summarization?
- Text Summarization Techniques Classification
- Our Target
- Related Work
- Dataset
- Evaluation Metric
- Model Architecture
- Training
  - Hyperparameters
  - Loss Function
- Evaluation
- Best Result
- How to Extend?

Dataset

- CNN, part of DeepMind Q&A dataset was our first choice. This dataset contains different CNN articles, supplemented with multiple highlights.
- We then sticked to a preprocessed version of the Gigaword dataset, a larger (4x) but easier-to-handle dataset than the CNN; as it has its ground truth labels as single-sentence summaries.
    - **Article:** "`south korea 's nuclear envoy kim sook urged north korea monday to restart work to disable its nuclear plants and stop its `` typical '' brinkmanship in negotiations #`"
    - **Ground Truth Summary:** "`envoy urges north korea to restart nuclear disablement`"
- This way, it is more like ***"Headline Generation"*** problem.

Outline

Evaluation Metric

- Nearly all approaches taken to tackle similar problems use ROUGE as an evaluation metric.
- We used the same metric for our model evaluation:
  - ROUGE-1
  - ROUGE-2
  - ROUGE-L

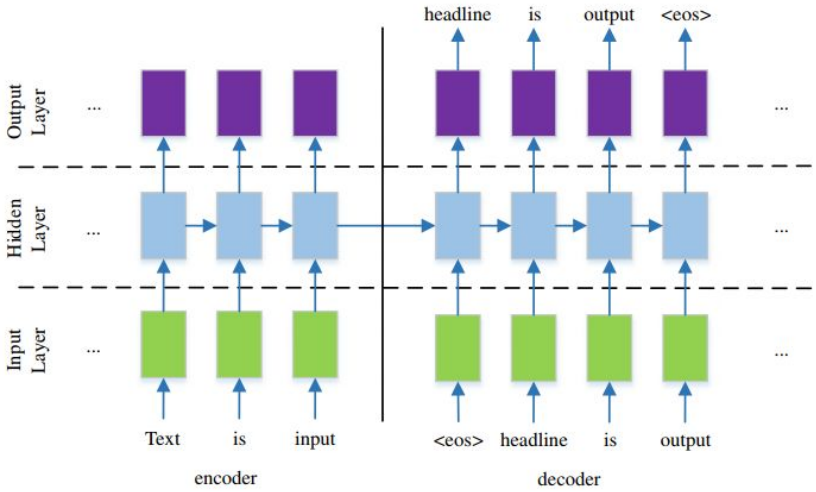
Outline

- What Is Text Summarization?
- Text Summarization Techniques Classification
- Our Target
- Related Work
- Dataset
- Evaluation Metric
- Model Architecture
- Training
  - Hyperparameters
  - Loss Function
- Evaluation
- Best Result
- How to Extend?

Model Architecture

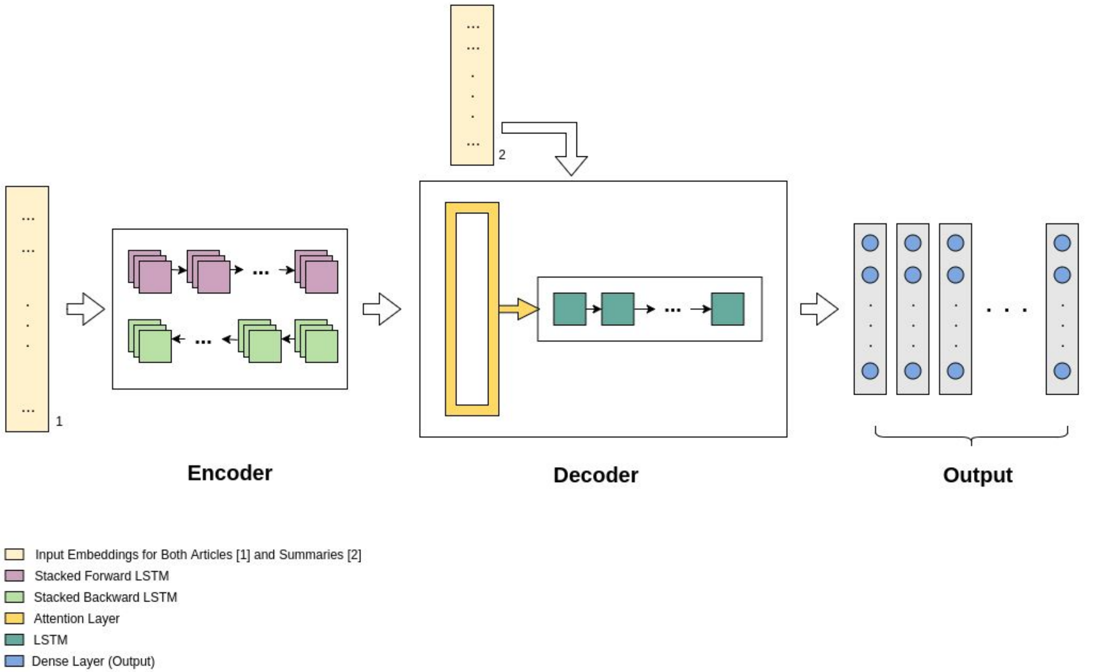
Encoder-Decoder Seq2Seq Network; where:

- Encoder: stacked bidirectional LSTM with dropout (keep_prop defaults to 0.8)
- Decoder: LSTM network with attention



[Source]

Model Architecture - A Closer Look



Outline

- What Is Text Summarization?
- Text Summarization Techniques Classification
- Our Target
- Related Work
- Dataset
- Evaluation Metric
- Model Architecture
- Training
  - Hyperparameters
  - Loss Function
- Evaluation
- Best Result
- How to Extend?

Training

Hyperparameters

- number of stacked LSTM layers (default 2)
- number of LSTM hidden units (default 150)
- emeddings
  - learned VS. pretrained
  - size (default 300)
- batch size (default 64)
- learning rate (default 0.01)
- dropout keep-probability (default 0.8)

Loss Function

- sparse softmax cross entropy

Outline

- What Is Text Summarization?
- Text Summarization Techniques Classification
- Our Target
- Related Work
- Dataset
- Evaluation Metric
- Model Architecture
- Training
    - Hyperparameters
    - Loss Function
- Evaluation
- Best Result
- How to Extend?

Evaluation

- To evaluate our model, we first trained it under various combinations of hyperparameter values that we anticipated would make significant difference in the obtained results.
- For each trained model, we used it to generate summaries for our test set articles, and evaluated the resulting summaries compared to the ground truth titles using ROUGE.
- Following are some of the trials we performed...

Evaluation

- pretrained GloVe embeddings
- 512-sample batch
- 3 epochs
- others -- default

- Average Training Time per Epoch:
    - 3 hours
- Results:
    - ROUGE-1: 31
    - ROUGE-2: 12.53
    - ROUGE-L: 27.27

Evaluation

- 200-feature self-learned embeddings
- 100-unit LSTM
- 64-sample batch
- 1 epoch
- others -- default

- Average Training Time per Epoch:
  - 5 hours

- Results:
  - ROUGE-1: 19.77
  - ROUGE-2: 6.02
  - ROUGE-L: 16.7

Evaluation

- pretrained GloVe embeddings
- 4-layer stacked LSTM
- 512-sample batch
- 3 epochs
- others -- default

- Average Training Time per Epoch:
  - 3 hours
- Results:
  - ROUGE-1: 19.37
  - ROUGE-2: 5.55
  - ROUGE-L: 17.02

Outline

- What Is Text Summarization?
- Text Summarization Techniques Classification
- Our Target
- Related Work
- Dataset
- Evaluation Metric
- Model Architecture
- Training
    - Hyperparameters
    - Loss Function
- Evaluation
- Best Result
- How to Extend?

Best Result

ROUGE

- ROUGE-1: 31
- ROUGE-2: 12.53
- ROUGE-L: 27.27

Model Configuration

- number of stacked LSTM layers (default 2)
- number of LSTM hidden units (default 150)
- emeddings
  - pretrained (GloVe)
  - size (default 300)
- batch size (512)
- learning rate (default 0.01)
- dropout keep-probability (default 0.8)

Outline

- What Is Text Summarization?
- Text Summarization Techniques Classification
- Our Target
- Related Work
- Dataset
- Evaluation Metric
- Model Architecture
- Training
    - Hyperparameters
    - Loss Function
- Evaluation
- Best Result
- How to Extend?

How to Extend?

- Apply ***"Beam Search"*** technique for better sequence generation.
- Support ***"Pointer-Generator"*** method to tackle factual errors, OOV (out-of-vocab) words and repeating.
- Generalize decoder network to work on CNN and other datasets with multi-sentence summaries.

Thank You