Automatic Text Summarization

Abeer Ahmad (38)
Mohammed Deifallah (59)

Outline

- Recap
- Problem Updates
- Model Architecture
- Training
- Testing
- What’s Next?

Recap

- We address the text summarization problem, the process of creating a short and coherent version of a longer document.
- We follow the single-document generic abstractive approach.
- Almost all abstractive models are variations of the Seq2Seq (Encoder-Decoder) network.

Evaluation Metrics

- We use ROUGE to evaluate our model.

- Recall-Oriented Understudy for Gisting Evaluation.
- ROUGE-1 refers to the overlap of 1-gram (each word) between the system and reference summaries.
- ROUGE-2 refers to the overlap of bigrams between the system and reference summaries.
- ROUGE-L: Longest Common Subsequence (LCS) based statistics.

BLEU VS. ROUGE

- Precision-oriented
- Proposed in 2002
- Formula:

$$\frac{number\_of\_overlapping\_words}{total\_words\_in\_system\_summary}$$

- Brevity penalty

- Recall-oriented
- Inspired from **BLEU** in 2004
- Formula:

$$\frac{number\_of\_overlapping\_words}{total\_words\_in\_reference\_summary}$$

- Verbosity penalty

State-of-the-Art Models

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

Outline

- Recap
- Problem Updates
- Model Architecture
- Training
- Testing
- What’s Next?

Problem Updates

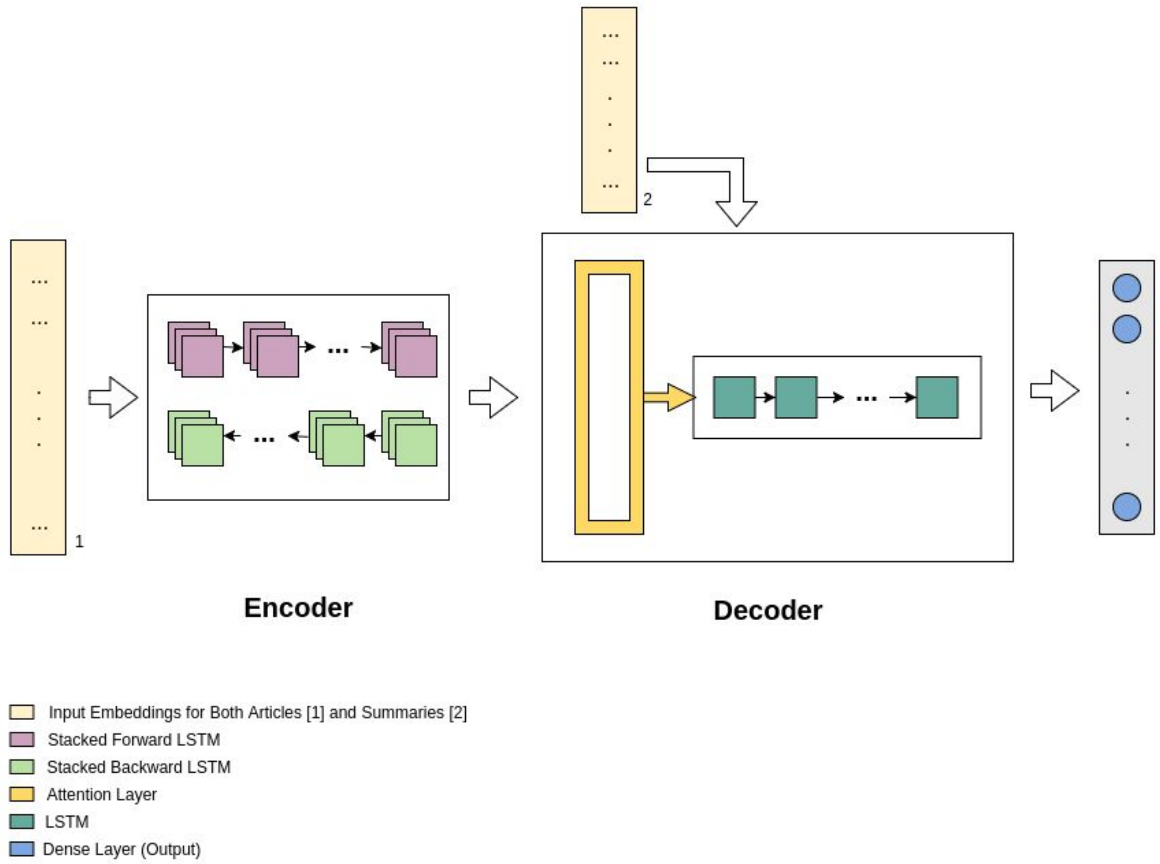
- We drop data preprocessing, except for sentence tokenization.
- We use an easier-to-handle dataset, Gigaword, instead of the CNN dataset for our baseline model.
- The Gigaword has 3,803,957 training samples and 189,651 test samples, with single-sentence ground truth summaries.
  - Article: `south korea '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`

Outline

- Recap
- Problem Updates
- Model Architecture
- Training
- Testing
- What's Next?

Model Architecture

- **Encoder**
  - Stacked Bidirectional LSTM with dropout (keep_prop defaults to 0.8)
- **Decoder**
  - LSTM network with attention



Workflow

1. Feed a batch of training samples, where each sample is tokenized and each token is represented by its GloVe embedding.
2. Encode each sample through the encoder network, generating an intermediate output.
3. Feed the encoder output as input to the decoder network, along with the ground truth summary embeddings.
4. Use the attention layer to mark a subset of encoder intermediate output vectors that significantly add meaning to the decoder generated summary.
5. Use a dense layer of the same size as our vocab to generate the most probable word at each timestep.

Outline

- Recap
- Problem Updates
- Model Architecture
- Training
- Testing
- What's Next?

Training

- We train our model with the following configuration:
  - 5-layer stacked LSTM
  - 300-dimensional GloVe embedding vectors
  - 150-unit LSTM cells for the encoder
  - 300-unit LSTM cells for the decoder
- We use Adam optimizer, with sparse cross entropy as a loss function.

Outline

- Recap
- Problem Updates
- Model Architecture
- Training
- Testing
- What’s Next?

Outline

- Recap
- Problem Updates
- Model Architecture
- Training
- Testing
- What’s Next?

What’s Next?

- Model Improvement for better results on the test set
- ROUGE evaluation
- Generalizing to the CNN dataset, if possible