1. What is summarization?
2. What can summaries look like?
3. Evaluation
4. Today’s focus: Single Document Summarization (text & multimodal)
Content

1. What is summarization?
2. What can summaries look like?
3. Evaluation
4. Today’s focus: Single Document Summarization (text & multimodal)

Disclaimer! The field is way too big to cover everything in one lecture, so we will just touch upon a selection of different work!
Content

1. What is summarization?
2. What can summaries look like?
3. Evaluation
4. Today’s focus: Single Document Summarization (text & multimodal)
What is summarization?
What is summarization?

Task: Find the important bits in all the data and return these
What is summarization?

Task: Find the important bits in all the data and return these
Why would you be interested in this?
Why would you be interested in this?

**Academically**
Can we learn a smaller representation that still captures our input?

**Practical standpoint**
Countless examples where amount of information available is too much to manually digest.
1. What is summarization?
2. What can summaries look like?
3. Evaluation
4. Today’s focus: Single Document Summarization (text & multimodal)
What can summaries look like?

Extractive summarization

Abstractive summarization
Extractive Summarization
Extractive Summarization
Extractive Summarization
Extractive Summarization
Extractive Summarization
Abstractive Summarization
Focus of this lecture

Single Document Summarization

- Text
- Multimodal
Content

1. What is summarization?
2. What can summaries look like?
3. Evaluation
4. Today’s focus: Single Document Summarization (text & multimodal)
Evaluation - what makes a summary ‘good’?
ROUGE


**ROUGE: Recall-Oriented Understudy for Gisting Evaluation**

ROUGE is a metric to evaluate textual summaries.
**ROUGE**


Predicted summary

ROUGE computes the quality of a summary, by comparing the number of overlapping n-grams in the predicted summary and the reference summary.

Reference summary

ROUGE is a metric to compute the quality of a textual summary, by comparing the number of overlapping n-grams in the predicted summary and the reference summary.
ROUGE


Predicted summary

ROUGE computes the quality of a summary, by comparing the number of overlapping n-grams in the predicted summary and the reference summary.

Reference summary

ROUGE is a metric to compute the quality of a textual summary, by comparing the number of overlapping n-grams in the predicted summary and the reference summary.
ROUGE

Predicted summary
ROUGE computes the quality of a summary, by comparing the number of overlapping n-grams in the predicted summary and the reference summary.

# Unigrams: 22
# Bigrams: 21

Reference summary
ROUGE is a metric to compute the quality of a textual summary, by comparing the number of overlapping n-grams in the predicted summary and the reference summary.

# Unigrams: 27
# Bigrams: 26
ROUGE


Predicted summary

ROUGE computes the quality of a summary, by comparing the number of overlapping n-grams in the predicted summary and the reference summary.

# Overlapping Unigrams: 21
# Overlapping Bigrams: 18

Reference summary

ROUGE is a metric to compute the quality of a textual summary, by comparing the number of overlapping n-grams in the predicted summary and the reference summary.
ROUGE


\[
ROUGE - N_{\text{Recall}} = \frac{\text{# overlapping } n\text{-grams}}{\text{# } n\text{-grams in reference summary}}
\]

\[
ROUGE - N_{\text{Precision}} = \frac{\text{# overlapping } n\text{-grams}}{\text{# } n\text{-grams in predicted summary}}
\]

\[
ROUGE - N_F = \frac{(1 + \beta^2) \times ROUGE - N_{\text{Recall}} \times ROUGE - N_{\text{Precision}}}{ROUGE - N_{\text{Recall}} + \beta^2 \times ROUGE - N_{\text{Precision}}}
\]
**ROUGE**


\[
ROUGE - 1_{\text{Recall}} = \frac{\text{# overlapping unigrams}}{\text{# unigrams in reference summary}} = \frac{21}{27} = 0.78
\]

\[
ROUGE - 1_{\text{Precision}} = \frac{\text{# overlapping unigrams}}{\text{# unigrams in predicted summary}} = \frac{21}{22} = 0.95
\]

\[
ROUGE - 1_F = \frac{(1 + \beta^2) \times ROUGE - 1_{\text{Recall}} \times ROUGE - 1_{\text{Precision}}}{ROUGE - 1_{\text{Recall}} + \beta^2 \times ROUGE - 1_{\text{Precision}}} = 2 \times \frac{0.78 \times 0.95}{0.78 + 0.95} = 0.86
\]

We use $\beta = 1$
ROUGE


$ROUGE - 1_{Recall} = \frac{\# \text{overlapping unigrams}}{\# \text{unigrams in reference summary}} = \frac{21}{27} = 0.78$

$ROUGE - 1_{Precision} = \frac{\# \text{overlapping unigrams}}{\# \text{unigrams in predicted summary}} = \frac{21}{22} = 0.95$

$ROUGE - 1_F = \frac{(1 + \beta^2) \times ROUGE - 1_{Recall} \times ROUGE - 1_{Precision}}{ROUGE - 1_{Recall} + \beta^2 \times ROUGE - 1_{Precision}} = 2 \times \frac{0.78 \times 0.95}{0.78 + 0.95} = 0.86$

We use $\beta = 1$

Same computation for Rouge-2, but use the bigram counts
ROUGE

ROUGE-L
Do the same, but for longest common subsequence.
(Use the union of LCS for multiple sentences.)

Multiple references
Use argmax of all ROUGE-N scores.
Correlation with Human Judgement

Okay?

People are not totally convinced and often perform a human evaluation as well.
Content

1. What is summarization?
2. What can summaries look like?
3. Evaluation
4. Today’s focus: Single Document Summarization (text & multimodal)
Single Document Summarization - Text

Extractive summarization

- Historical methods
- Binary labeling strategy
- Reinforcement learning approach
- ...

Abstractive summarization

- Encoder
- Attention
- Context vector
- Decoder
- $x_1, x_2, x_3, x_T$
- $y_{t-1}, y_t$
Single Document Summarization - Text

Extractive summarization

• Historical methods
• Binary labeling strategy
• Reinforcement learning approach
• ...

Disclaimer! The field is way too big to cover everything in one lecture, so we will just touch upon a selection of different work!

Abstractive summarization

Encoder

Attention

Context vector

Decoder

$y_{t-1} \quad y_t$

$x_1 \quad x_2 \quad x_3 \quad x_T$
Historical methods - **LexRank** as example

Make a graph from the document. Each sentence is a **node**.

Compute similarity between sentences. These scores are the **edges**.

Use PageRank to rank these sentences.

Select the highest ranking sentences as the summary.

Unsupervised!
Extractive Single Document Summarization - Text (2)


Binary labeling strategy

Classify each sentence as either ‘belongs to summary’ or as ‘does not belong to summary’.
Extractive Single Document Summarization - Text (3)


Reinforcement Learning Strategy
Each document is a context, combinations of sentences in the document are actions.

Learn a policy to perform the best action - get the best summary.

This method optimises directly for ROUGE!
Abstractive Single Document Summarization - Text (1)


Main idea / contributions

• Introduce encoder / decoder structure for summarisation.
• Add tags to the encoder.

Main idea / contributions

- Introduce encoder / decoder structure for summarisation.
- Add tags to the encoder.
- Introduce pointing probability, for unknown words.
Abstractive Single Document Summarization - Text (1)


Main idea / contributions

- Introduce encoder / decoder structure for summarisation.
- Add tags to the encoder.
- Introduce pointing probability, for unknown words.
- Introduce hierarchical attention mechanism.
Abstractive Single Document Summarization - Text (2)


Main idea / contributions

• Make use of pointer network, that can copy words from the input at any time - less unknown words.
• Introduce coverage mechanism that keeps track of previously paid attention.
Challenges

Questionable grammaticality.

Not very abstractive - Next method addresses this.
Main idea / contributions

• Use convolutions, to capture the hierarchical structure of the text.

• Add topic information, to force the summary into a certain direction.

• They introduce a new dataset!
Single Document Summarization - Multimodal
Multimodal Summarization with Multimodal Output


Figure 1: The illustration of our proposed task – Multimodal Summarization with Multimodal Output (MSMO). The image can help better understand the text in the red font.
Multimodal Summarization with Multimodal Output


Figure 2: The framework of our model.
Summary

- Automatic summarization has the goal to automatically find the important bits in all the data and return these.
Summary

• Automatic summarization has the goal to automatically find the important bits in all the data and return these.

• We can summarise different types of data. In this lecture we have focussed on text and we have briefly touched upon multimodal summarization.
Summary

• Automatic summarization has the goal to automatically find the important bits in all the data and return these.
• We can summarise different types of data. In this lecture we have focussed on text and we have briefly touched upon multimodal summarization.
• Summarization techniques can be divided into extractive methods and abstractive methods.
Summary

• Automatic summarization has the goal to automatically find the important bits in all the data and return these.
• We can summarise different types of data. In this lecture we have focussed on text and we have briefly touched upon multimodal summarization.
• Summarization techniques can be divided into extractive methods and abstractive methods.
• ROUGE and Human Evaluation are used to evaluate the produced summaries.
Questions?

Slides are available at maartjeth.github.io/#talks