AUTOMATIC TEXT SUMMARIZATION

Sütő Evelyne [1]

[1] Babeş-Bolyai University, Computer Science, Cluj-Napoca

INTRODUCTION

- **Purpose:** Present a the different approaches and problems of summarization
- Extractive summaries
- Abstractive summaries
- Why summary generation is important?
- How can we evaluate summarization systems
- Conclusions

EXTRACTIVE SUMMARIZATION

In extractive summarization we are mainly focused on calculating the importance of each sentence in our document, while still considering the grammatical correctness of the overall summary while choosing the most important sentences to represent the overall meaning of the text.

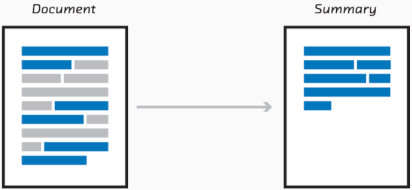

Figure 1: Summary in a nutshell Image source: `https://blog.fastforwardlabs.com/2016/04/21/summarization-as-a-gateway-to-computable-language.html`

WORD FREQUENCY AND STATISTICAL METHODS

WORD FREQUENCY AND STATISTICAL METHODS

- Frequency based [Luhn, 1958]

  Extraction based on:

  - Word frequency
  - Relative position of words in a sentence

- Cue, Key, Title and Location [Edmundson, 1969]

  Where:

  - **Cue** word (derived from a dictionary) is considered to introduce important information
  - **Key** words are used to consist of topic based words
  - **Title** assumption words in title must descriptive of subject
  - **Location** relies on certain characteristics of document structures, e.g topic sentences appear very early in sections

$$w_1 * C + w_2 * K + w_3 * T + w_4 * L$$

WORD FREQUENCY AND STATISTICAL METHODS

A more sophisticated measure for frequencies could be the tf-idf measure. The tf*idf weights of words are good indicators of importance, and they are easy and fast to compute.
[Nenkova and McKeown, 2012]

$$tf * idf = c(t) * log \frac{D}{d(t)}$$

Where c(t) is the term frequency in the corpus, D is the number of documents and d(t) is number of document t occurred in.

Can be easily used as input for a Naive Bayes classifier.

TEXT CONNECTIVITY

Lexical chains [Barzilay and Elhadad, 1999]

- Intuition, for using lexical chains rather than frequency
- Challenge in this area is to find a scoring system to select the most significant chain
- After selecting the strongest chains, for each chain choose the sentence that contains the first appearance of a representative member of the chain in the text

$$Score(Chain) = Length(Chain) * Homogeneity$$

TEXT CONNECTIVITY

G-Flow system [Christensen et al., 2013]

- Graph based system
- A graph that approximates the discourse relations across sentences based on discourse cues, deverbal nouns, co-reference.
- Nodes are sentences, and each edge represents a pairwise ordering constraint necessary for a coherent summary.
- Using this graph we can estimate how coherent our extracted text is
- The edges model: deverbal nouns, co-reference, lexical chains, cue words

GRAPHED BASED METHODS

GRAPH BASED METHODS

TextRank [Mihalcea and Tarau, 2004]

A graph-based ranking algorithm is a way of deciding on the importance ofa vertex within a graph, by taking into account global information recursively computed fromthe entire graph, rather than relying only on local vertex-specific information.

- Importance is computed recursively from the entire graph
- Graph-based ranking model is based voting or recommendation
- Each link equals a vote, the more vote an edge has the bigger its importance

GRAPH BASED METHODS

$$S(V_i) = (1-d) + d * \sum_{j \in In(V_i)} \frac{1}{|Out(V_j)|} S(V_j)$$

Where:

- In(V) = set of vertices pointing to V
- Out(V) = set of vertices to which V points to
- d = damping factor, probability of jumping from a given vertex to another random vertex, usually set to 0.85

The values of the vertices are assigned randomly at first and the algorithm runs until convergence is achieved.

ABSTRATIVE SUMMARIZATION

ABSTRACTIVE SUMMARIZATION

Abstractive summaries try to present a shorter version of the text input, based on deeper understanding of the conecpts being represented and it may contain newly added sentences,phrases as well to improve the generated abstract

Figure 2: Image source:
`https://www.summarizebot.com/`

ENCODER-DECODER

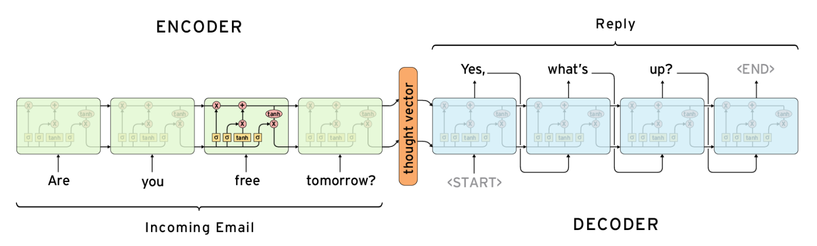

Figure 3: Image source:
`https://medium.com/botsupply/generative-model-chatbots/`

ENCODER-DECODER

Encoder decoder summary generation [Nallapati et al., 2016]

- Challenge: sequence lengths for the input and output is very different

- Summarization is not simple mapping because we need concentrate the main ideas of the text with some loss

ENCODER-DECODER

- **Model:** bi-directional RNN encoder with attention mechanism and a uni-directional RNN decoder

- Large vocabulary trick (LVT), by restricting the decoder's vocabulary

- Capture additional tags: parts-of-speech tags, named-entity tags, and TF and IDF statistics

- Look-up based embedding matrices for the vocabulary of each tag-type

- Solution for rare words using a switching generator-pointer

ENCODER-DECODER

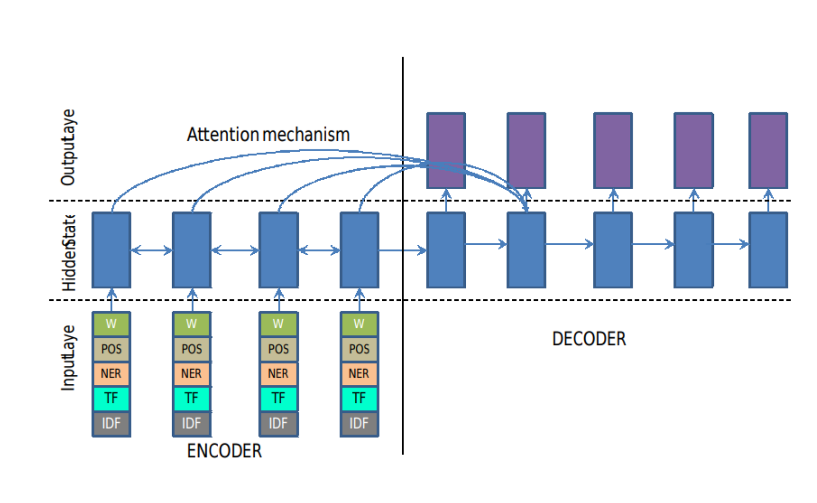

Figure 4: Image source: [Nallapati et al., 2016]

EVALUATION METRICS

Currently the ROUGE (Recall-Oriented Understudy for Gisty Evaluation) [Lin, 2004] metrics are the standard

ROUGE-N is an n-gram recall between a candidate summary and a set ofreference summaries

$$ROUGE - N = \frac{\sum_{S \in \{\text{Reference summary}\}} \sum_{n-gram \in S} Count_{match}(n - gram)}{\sum_{S \in \{\text{Reference summary}\}} \sum_{n-gram \in S} Count(n - gram)}$$

EVALUATION METRICS

$$R_{LCS} = \frac{LCS(X, Y)}{m}$$

$$P_{LCS} = \frac{LCS(X, Y)}{n}$$

$$ROUGE - L = \frac{(1 + b^2)R_{LCS}P_{LCS}}{R_{LCS} + b^2P_{LCS}}$$

Given two sequences X and Y, the longest common subsequence (LCS) of X and Y (X of length m and Y of length n) is a common subsequence with maximum length

CONCLUSION

- Summary generation problem is a very complex one.
- Endless approaches have been researched such as graph based methods, statistical computational methods, text connectivity, deep learning.
- People are looking for new approaches continously.
- For example researchers are beginning to use different reinforcement learning methods for sentence selection and even multi agent systems.

Barzilay, R. and Elhadad, M. (1999).
Using lexical chains for text summarization.
Advances in automatic text summarization, pages 111–121.

Christensen, J., Soderland, S., Etzioni, O., et al. (2013).
Towards coherent multi-document summarization.
In *Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: Human language technologies*, pages 1163–1173.

Edmundson, H. P. (1969).
New methods in automatic extracting.
Journal of the ACM (JACM), 16(2):264–285.

Lin, C.-Y. (2004).
Rouge: A package for automatic evaluation of summaries.
Text Summarization Branches Out.

Luhn, H. P. (1958).
The automatic creation of literature abstracts.
IBM Journal of research and development, 2(2):159–165.

Mihalcea, R. and Tarau, P. (2004).
Textrank: Bringing order into text.
In *Proceedings of the 2004 conference on empirical methods in natural language processing*.

Nallapati, R., Zhou, B., Gulcehre, C., Xiang, B., et al. (2016).
Abstractive text summarization using sequence-to-sequence rnns and beyond.
arXiv preprint arXiv:1602.06023.

Nenkova, A. and McKeown, K. (2012).
A survey of text summarization techniques.
In *Mining text data*, pages 43–76. Springer.