

Extractive summarisation of biomedical research articles using TextRank, WordRank, and a hybrid approach

Kristina Levina

Course code: 732A81

LiU-ID: krile102

Abstract

This project aims at developing a tool to generate a summary automatically based on the main text of a biomedical research manuscript. To this end, extractive summarisation is employed. The motivation behind choosing extractive summarisation is, in essence, the necessity to preserve the key sentences from the main text. Extractive summarisation will retrieve sentences based on their importance without rephrasing them, thereby excluding misinterpretations. This is crucial for scientific texts. In this project, SumPubMed dataset was used. According to the literature, SumPubMed dataset was created specifically for testing summarisation tools of biomedical research articles. Two algorithms were chosen to implement this task: TextRank as a baseline and a pretrained transformer from Hugging Face. Their performance was assessed using X and Y metrics. The obtained results show that Y better suits for the considered task.

1 Introduction

With increasing volume of published articles in medical research, it becomes increasingly difficult for doctors, medical staff, and public health officials to stay updated. Sometimes, a

2 Theory

2.1 PageRank

2.2 ROUGE

2.3 BLEU

3 Data

4 Method

4.1 TextRank

4.2 WordRank

4.3 Hybrid

5 Results

Kaykobad Reza et al. (2020)

[scale=0.5]unigrams.pdf

Figure 1: An example graph

6 Discussion

7 Conclusion

8 Preamble

Table 1 shows the syntax supported by the style files. We encourage you to use the natbib styles. You can use the command `\citet` (cite in text) to get “author (year)” citations, like this citation to a paper by [Gusfield \(1997\)](#). You can use the command `\citep` (cite in parentheses) to get “(author, year)” citations ([Gusfield, 1997](#)). You can use the command `\citealp` (alternative cite without parentheses) to get “author, year” citations, which is useful for using citations within parentheses (e.g. [Gusfield, 1997](#)).

References

- James W. Cooley and John W. Tukey. 1965. [An algorithm for the machine calculation of complex Fourier series](#). *Mathematics of Computation*, 19(90):297–301.
- Dan Gusfield. 1997. *Algorithms on Strings, Trees and Sequences*. Cambridge University Press, Cambridge, UK.
- Md. Kaykobad Reza, Rifat Rubayatul Islam, Sadik Siddique, Md. Mostofa Akbar, and M. Sohel Rahman. 2020. [Automatic summarization of scientific articles from biomedical domain](#). In *Proceedings of International Joint Conference on Computational Intelligence*, pages 591–602, Singapore. Springer Singapore.

Output	natbib command	Old ACL-style command
(Cooley and Tukey, 1965)	\citep	\cite
Cooley and Tukey, 1965	\citealp	no equivalent
Cooley and Tukey (1965)	\citet	\newcite
(1965)	\citeyearpar	\shortcite
Cooley and Tukey's (1965)	\citeposs	no equivalent
(FFT; Cooley and Tukey, 1965)	\citep[FFT;][]	no equivalent

Table 1: Citation commands supported by the style file. The style is based on the natbib package and supports all natbib citation commands. It also supports commands defined in previous ACL style files for compatibility.