Interactive Document Summarization

Raoufdine Said, Adrien Guille[0000−0002−1274−6040]

Université de Lyon, Lyon 2, ERIC UR 3083, France
raoufdine.said@univ-lyon2.fr
adrien.guille@univ-lyon2.fr

Abstract. With the advent of modern chatbots, automatic summarization is becoming common practice to quicken access to information. However the summaries they generate can be biased, unhelpful or untruthful. Hence, in sensitive scenarios, extractive summarization remains a more reliable approach. In this paper we present an original extractive method combining a GNN-based encoder and a RNN-based decoder, coupled with a user-friendly interface that allows for interactive summarization.

Keywords: Interactive Summarization · Extractive Summarization · Graph Neural Network · Recurrent Neural Network · User Interface.

1 Introduction

One of the most common uses for modern chatbots is automatic summarization [1]. While the abstractive summaries generated by these tools can be relevant, the language models they're based on might be biased, *e.g.* politically biased [9] or gender biased [5], they can sometimes be untruthful because large language models are prone to hallucinate [6], or they can simply be unhelpful, as chatbots might fail to follow instructions correctly [8]. This limits their reliability and can make them ill-suited to summarize sensitive documents, like scientific articles, encyclopedic articles or press articles, among others, where information shouldn't be altered. A viable alternative under this scenario is extractive summarization, which consists in composing the summary from actual sentences picked from the original document. In this paper, we first present an original method for extractive summarization, by combining two existing methods. The implementation of the model is available here: https://github.com/baragouine/radsum. Next, we present an interface that allows users to interactively summarize documents, the code of which is available here: https://github.com/baragouine/radsum_app/. For a video demonstration, see https://youtu.be/vBenEaCIwkI.

2 Summarization Method

We combine two existing methods, namely (i) HeterSUMGraph [11], which we slightly modify and only use for the encoding part of our architecture, and (ii) SummaRuNNer [7], which we use for the decoding part only. Fig. 1 illustrates the overall methodology. The rational behind combining these methods is to have a more expressive encoder while having an interpretable decoder.

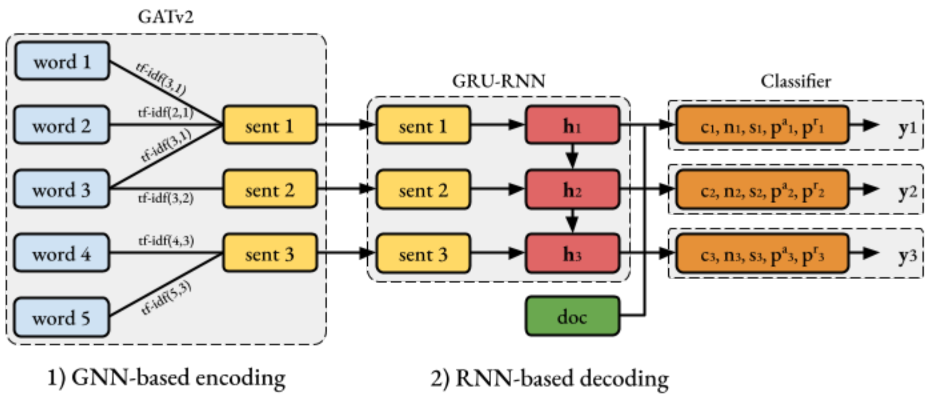


Fig. 1. Overall architecture of the implemented summarization method.

Document Encoding. We convert the input document to a graph as in HeterSUMGraph [11]. It is a bipartite graph with two kinds of vertices: word vertices and sentence vertices. More precisely there is one vertex per unique word in the document and one vertex per sentence in the document, connected according to the composition of the sentences. Edges are weighted with *tf-idf* scores measured at the sentence level, meaning *tf* is calculated as the number of times the word occurs in the sentence, while *idf* is defined as the inverse of the degree of the word vertex. Word embeddings are propagated with a 2-layer graph neural network to obtain sentence representations. Whereas HeterSUMGraph uses GAT layers [10], we implement more expressive GATv2 layers [2], with an attention mechanism that accounts for discretized tf-idf weights via edge embeddings.

Summary Decoding. We pass the sentence representations to a recurrent neural network based on the GRU cell [3] to further contextualize them and proceed similarly to SummaRuNNer in classifying sentences sequentially, following their order in the document. The probability to keep the i^{th} sentence is calculated in terms of 5 scores:

- **content score**: a linear function of the representation of this sentence, $\mathbf{W}_c\mathbf{h}_i$;
- **salience score**: a bilinear function of the representation of this sentence and the representation of the whole document, $\mathbf{h}_i^\top \mathbf{W}_s \mathbf{d}$;
- **novelty score**: a bilinear function of the representation of sentence i and the representation of the document up to sentence $i-1$, $\mathbf{h}_i^\top \mathbf{W}_r \tanh(\mathbf{s}_i)$;
- **absolute position score**: a linear function of the embedding of the absolute position of the sentence, $\mathbf{W}_{ap}\mathbf{p}_i^a$;
- **relative position score**: a linear function of the embedding of the relative position of the sentence, $\mathbf{W}_{rp}\mathbf{p}_i^r$.

The probability to keep the i^{th} sentence in the summary is calculated as follows:

$$\mathrm{p}(1|\mathbf{h}_i, \mathbf{s}_i, \mathbf{d}) = \sigma\left(\mathbf{W}_c\mathbf{h}_i + \mathbf{h}_i^\top\mathbf{W}_s\mathbf{d} + \mathbf{h}_i^\top\mathbf{W}_r\tanh(\mathbf{s}_i) + \mathbf{W}_{ap}\mathbf{p}_i^a + \mathbf{W}_{rp}\mathbf{p}_i^r\right), \tag{1}$$

where $\mathbf{d}$ is the representation of the whole document obtained by averaging all the hidden states $\{\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n\}$, and $\mathbf{s}_i$ is the representation of the document up to the previous sentence, a weighted average of $\{\mathbf{h}_1, \ldots, \mathbf{h}_{i-1}\}$:

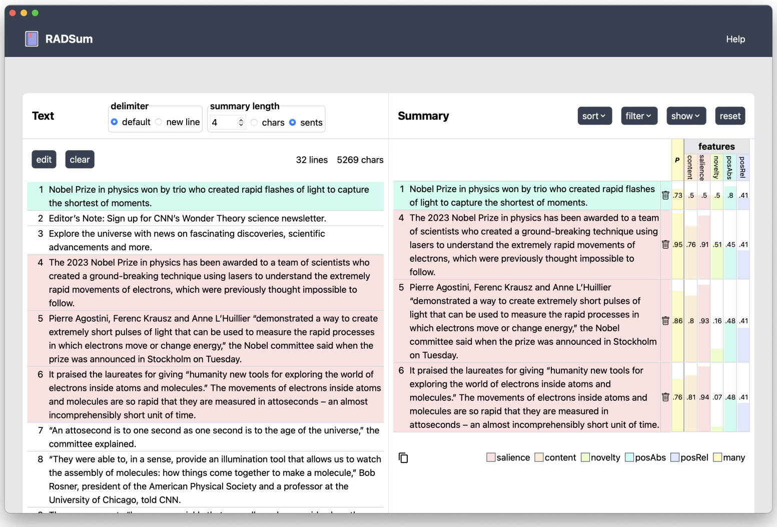
$$\mathbf{s}_i = \sum_{j=1}^{i-1}\mathbf{h}_j\mathrm{p}(1|\mathbf{h}_j, \mathbf{s}_j, \mathbf{d}). \tag{2}$$

3 User Interface

Table 1. Performance on the NYT corpus. The gain over SummaRuNNer is given after the + sign.

	ROUGE-1	**ROUGE-L**
SummaRuNNer	45.3	34.65
HeterSUMGraph	46.76 +2.0%	35.21 +1.6%
RadSum	46.91 +2.4%	35.35 +2.0%

For the purpose of the demonstration, we train our method, RadSum, on the New York Times annotated corpus [4]. Its performance in terms of ROUGE-1 and ROUGE-L scores is reported in Table 1, along with the scores achieved by SummaRuNNer and HeterSUMGraph. Fig. 2 shows the default interface of the proposed application. The left side of the windows allows to input the document to summarize and adjust general settings. The right side of the window shows the summary extracted from that document. For each sentence, the contribution of the 5 scores to the overall probability is depicted with a bar chart. Each sentence is colored according to the dominant score (scaled between 0 and 1 with the sigmoid function). When inputting a press article about the Nobel Prize in Physics awarded in October 2023, it produces a 4-sentence long summary that focuses on the laureates and the implications of their discovery by picking the 1st sentence, particularly for its position in the document, and sentences 4-6 mostly because of their content and salience. Fig. 3 shows how one could leverage the app to tailor the summary according to its needs. Another 4-sentence long summary focusing on the discovery itself is obtained by selecting sentences 19, 28, 29 and 31 solely based on the salience score (using the filter button). It also shows how one can manually filter the sentences: here, sentence 31 has been removed from the summary, which resulted in the automatic addition of sentence 30 (the next most salient sentence) to keep the summary at a length of 4 sentences.



Text

delimiter: [✓ default] new line

summary length: 4 chars [✓ sents]

32 lines 5269 chars

1 Nobel Prize in physics won by trio who created rapid flashes of light to capture the shortest of moments.

2 Editor's Note: Sign up for CNN's Wonder Theory science newsletter.

3 Explore the universe with news on fascinating discoveries, scientific advancements and more.

4 The 2023 Nobel Prize in physics has been awarded to a team of scientists who created a ground-breaking technique using lasers to understand the extremely rapid movements of electrons, which were previously thought impossible to follow.

5 Pierre Agostini, Ferenc Krausz and Anne L'Huillier "demonstrated a way to create extremely short pulses of light that can be used to measure the rapid processes in which electrons move or change energy," the Nobel committee said when the prize was announced in Stockholm on Tuesday.

6 It praised the laureates for giving "humanity new tools for exploring the world of electrons inside atoms and molecules." The movements of electrons inside atoms and molecules are so rapid that they are measured in attoseconds – an almost incomprehensibly short unit of time.

7 "An attosecond is to one second as one second is to the age of the universe," the committee explained.

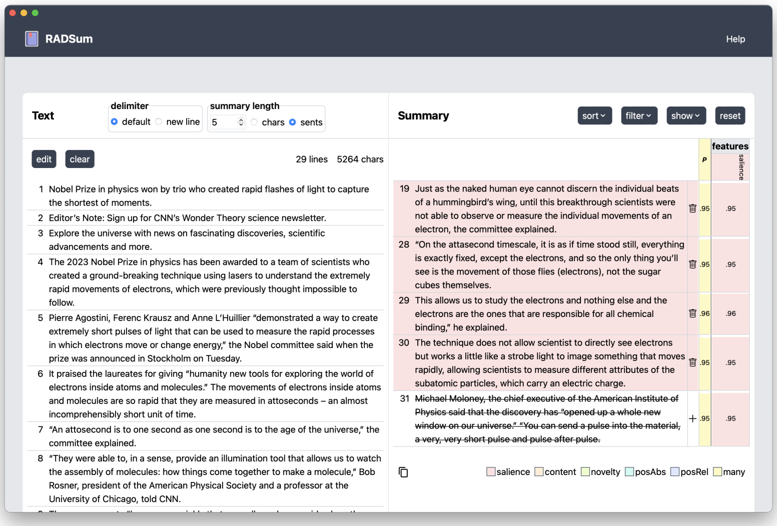
8 "They were able to, in a sense, provide an illumination tool that allows us to watch the assembly of molecules: how things come together to make a molecule," Bob Rosner, president of the American Physical Society and a professor at the University of Chicago, told CNN.

Summary

		P	features				
			content	salience	novelty	posAbs	posRel
1	Nobel Prize in physics won by trio who created rapid flashes of light to capture the shortest of moments.	.73	.5	.5	.5	.8	.41
4	The 2023 Nobel Prize in physics has been awarded to a team of scientists who created a ground-breaking technique using lasers to understand the extremely rapid movements of electrons, which were previously thought impossible to follow.	.95	.76	.91	.51	.45	.41
5	Pierre Agostini, Ferenc Krausz and Anne L'Huillier "demonstrated a way to create extremely short pulses of light that can be used to measure the rapid processes in which electrons move or change energy," the Nobel committee said when the prize was announced in Stockholm on Tuesday.	.86	.8	.93	.16	.48	.41
6	It praised the laureates for giving "humanity new tools for exploring the world of electrons inside atoms and molecules." The movements of electrons inside atoms and molecules are so rapid that they are measured in attoseconds – an almost incomprehensibly short unit of time.	.76	.81	.94	.07	.48	.41

🟥 salience 🟧 content 🟩 novelty 🟦 posAbs 🟪 posRel 🟨 many

Fig. 2. Default interface. The *P* column (in yellow) corresponds to the probability to keep the sentence.



Text

delimiter: [✓ default] new line

summary length: 5 chars [✓ sents]

29 lines 5264 chars

1 Nobel Prize in physics won by trio who created rapid flashes of light to capture the shortest of moments.

2 Editor's Note: Sign up for CNN's Wonder Theory science newsletter.

3 Explore the universe with news on fascinating discoveries, scientific advancements and more.

4 The 2023 Nobel Prize in physics has been awarded to a team of scientists who created a ground-breaking technique using lasers to understand the extremely rapid movements of electrons, which were previously thought impossible to follow.

5 Pierre Agostini, Ferenc Krausz and Anne L'Huillier "demonstrated a way to create extremely short pulses of light that can be used to measure the rapid processes in which electrons move or change energy," the Nobel committee said when the prize was announced in Stockholm on Tuesday.

6 It praised the laureates for giving "humanity new tools for exploring the world of electrons inside atoms and molecules." The movements of electrons inside atoms and molecules are so rapid that they are measured in attoseconds – an almost incomprehensibly short unit of time.

7 "An attosecond is to one second as one second is to the age of the universe," the committee explained.

8 "They were able to, in a sense, provide an illumination tool that allows us to watch the assembly of molecules: how things come together to make a molecule," Bob Rosner, president of the American Physical Society and a professor at the University of Chicago, told CNN.

Summary

		P	features
			salience
19	Just as the naked human eye cannot discern the individual beats of a hummingbird's wing, until this breakthrough scientists were not able to observe or measure the individual movements of an electron, the committee explained.	.95	.95
28	"On the attasecond timescale, it is as if time stood still, everything is exactly fixed, except the electrons, and so the only thing you'll see is the movement of those flies (electrons), not the sugar cubes themselves.	.95	.95
29	This allows us to study the electrons and nothing else and the electrons are the ones that are responsible for all chemical binding," he explained.	.96	.96
30	The technique does not allow scientist to directly see electrons but works a little like a strobe light to image something that moves rapidly, allowing scientists to measure different attributes of the subatomic particles, which carry an electric charge.	.95	.95
31	~~Michael Moloney, the chief executive of the American Institute of Physics said that the discovery has "opened up a whole new window on our universe." "You can send a pulse into the material, a very, very short pulse and pulse after pulse.~~	.95	.95

🟥 salience 🟧 content 🟩 novelty 🟦 posAbs 🟪 posRel 🟨 many

Fig. 3. Example of interaction: the user has selected salience as the only score and has removed sentence 31 from the summary.

References

1. Azaria, A., Azoulay, R., Reches, S.: Chatgpt is a remarkable tool – for experts (2023)
2. Brody, S., Alon, U., Yahav, E.: How attentive are graph attention networks? ICLR (2022)
3. Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. EMNLP (2014)
4. Durrett, G., Berg-Kirkpatrick, T., Klein, D.: Learning-based single-document summarization with compression and anaphoricity constraints. ACL (2016)
5. Jentzsch, S., Turan, C.: Gender bias in BERT - measuring and analysing biases through sentiment rating in a realistic downstream classification task. GeBNLP workshop @ ACL (2022)
6. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y.J., Madotto, A., Fung, P.: Survey of hallucination in natural language generation. ACM Computing Surveys **55**(12) (2023)
7. Nallapati, R., Zhai, F., Zhou, B.: Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. AAAI (2017)
8. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C.L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., Lowe, R.: Training language models to follow instructions with human feedback. ArXiV Technical Report (2022)
9. Rozado, D.: The political biases of chatgpt. Social Sciences **12**(3) (2023)
10. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph Attention Networks. ICLR (2018)
11. Wang, D., Liu, P., Zheng, Y., Qiu, X., Huang, X.: Heterogeneous graph neural networks for extractive document summarization. ACL (2021)