# A Scoring Model Assisted by Frequency for Multi-Document Summarization

Yue Yu[1,3](✉), Mutong Wu[1], Weifeng Su[1,2](✉), and Yiu-ming Cheung[3]

[1] Computer Science and Technology Programme,
Division of Science and Technology, Hefei, China
[2] Guangdong Key Lab of AI and Multi-Modal Data Processing, BNU-HKBU
United International College, Guangdong, China
o930201601@mail.uic.edu.cn, {mutongwu,wfsu}@uic.edu.cn
[3] Department of Computer Science, Hong Kong Baptist University,
Hong Kong, China
{csyueyu,ymc}@comp.hkbu.edu.hk

**Abstract.** While position information plays a significant role in sentence scoring of single document summarization, the repetition of content among different documents greatly impacts the salience scores of sentences in multi-document summarization. Introducing frequencies information can help identify important sentences which are generally ignored when only considering position information before. Therefore, in this paper, we propose a scoring model, SAFA (**S**elf-**A**ttention with **F**requency Gr**a**ph) which combines position information with frequency to identify the salience of sentences. The SAFA model constructs a frequency graph at the multi-document level based on the repetition of content of sentences, and assigns initial score values to each sentence based on the graph. The model then uses the position-aware gold scores to train a self-attention mechanism, obtaining the sentence significance at its single document level. The score of each sentence is updated by combing position and frequency information together. We train and test the SAFA model on the large-scale multi-document dataset Multi-News. The extensive experimental results show that the model incorporating frequency information in sentence scoring outperforms the other state-of-the-art extractive models.

**Keywords:** Multiple document summarization · Position information · Frequency · Graph

## 1 Introduction

Document summarization usually has two directions: Single Document Summarization (SDS) and Multi-Document Summarization (MDS). While many previous SDS researches have achieved competitive results [3,9,24,25,30], there is still a room for MDS development. Though some of the previous SDS researches could

This work is supported by the BNU-HKBU United International College research grant.


I. Farkaš et al. (Eds.): ICANN 2021, LNCS 12895, pp. 309–320, 2021.
https://doi.org/10.1007/978-3-030-86383-8_25

be generalized on MDS field, MDS models still need to be developed according to their own characteristics. Some studies [8,18,26,27,33] show that position information, one of the main factors for identifying salient sentences, could have substantially positive effects on single-document summarization. As the study [29] shows, however, position information is less effective than the frequency of sentences for improving the scores of evaluations (e.g. ROUGE [17]). Repetition among multiple documents has rarely been considered as an indicator of salience information before, because it is not applicable for SDS. For the MDS dataset, documents in a document pair mainly discuss the same event, thus having many repetitive contents. A sentence should be considered significant if its content has been repeated over many documents. On the other hand, not all documents have a similar narrative structure. It might not be sufficient to depend on the position information only for MDS tasks to solve this problem. We need to balance the frequency and position information when scoring sentences.

In this paper, we propose a model - **S**elf-**A**ttention with **F**requency Gr**a**ph (SAFA) to consider the position and frequency of a sentence at the same time when scoring the sentence. In SAFA, a global sentences frequency graph is first constructed for each document pair based on frequency information at the multi-document level. A node in the graph stands for a sentence. The weight of the edge between two nodes is the cosine similarity of the nodes sentences. The graph then groups sentences with high similarities together. Based on the similarities between sentences and the size of each group, a newly defined variable centrality directly represents "to what degree a sentence contains the repeated contents in the document pair". Centrality for each sentence is thus assigned as its initial *Score* to indicate its frequency information at multi-document level. At the same time, we utilize the self-attention mechanism with a Bi-LSTM neural network to extract the sentence information at the single-document level. The training process of the Bi-LSTM neural network uses position-aware pseudo values as the training target, thus incorporating the position information in the self-attention values. These values are then used to update the frequency graph and assign new scores, which combine both the centrality and the self-attention values to nodes. Finally, a sentence ranking and selecting process is implemented to generate the summary. Therefore, each sentence's final score considers both the position information from the single-document level and the frequency from the multi-document level.

We train and test our model on Multi-News [7], the first large-scale MDS news dataset released in 2019. The results show that SAFA outperforms the other extractive models on ROUGE scores. We have also conducted ablation studies on the two features to show that they are both indispensable parts of MDS. The human evaluations show that the summary from our proposed model contains more important information with strong readability.

The contributions of this paper are:

- We exploit the role of frequency information in multi-document summarization. Ablation experiments show that the model combing two features (position and frequency) performs better than considering only one feature.

– We propose a scoring model - SAFA, which incorporates the frequency of a sentence at the multi-document level with the positional information at the single document level.
– We conduct extensive experiments on the large-scale MDS dataset Multi-News, which show that the method combining these two features outperforms the other extractive models to the best of our knowledge. The results of the human evaluation show that the summaries generated from our model hold strong readability and contain more information.

## 2 Related Work

**Graph-Based Methods.** [6,8,11,16,22,28,31,34] have been widely applied to solve document summarization issues. Some use the graph-based ranking mechanism [6,22,28] for sentence extraction. For instance, LexRank [6] proposed a stochastic graph-based method to calculate the importance of relative units. TextRank [22] proposed a ranking system with two unsupervised methods to extract keyword or sentence. Others introduce graphs to explore the structures between sentences and documents. For example, GraphSum [16], a abstractive summarization model, combines multiple graphs into a neural network which has a similar structure as Transformer [32]. This inspires us to use a graph constructed from the frequency information to identify the repeated contents among multiple documents.

In summarization, a graph could be used to represents the input documents where the nodes are sentences and the edges are relations between the two connecting nodes [34]. This provides us a way to directly measure the repetitive contents among multiple documents by putting the similarity between two sentences to the edge to introduce frequency information to the neural network

**Position Information** has usually been a significant factor for finding the salience of a word or sentence in document summarization [5,12,21,26,27]. Many researches [8,18,26,27,33] approve of the positive effect of position information on extracting important content when dealing SDS issues. They state that sentences at the beginning position of a document have higher probabilities to being salient. For a MDS task, the attributes of single document is also an indispensable factor to obtain a structured summary.

**The Attention Mechanism** was first proposed in solving image classification problems [23]. In recent years, the incorporation of self-attention in neural networks has solved many NLP field problems such as abstractive summarization [20], text parsing [13,14], and speaker identification [1]. In this paper, we use a self-attention mechanism [19] with Bi-LSTM neural network to obtain the salience of a sentence within a single document based on the position information.

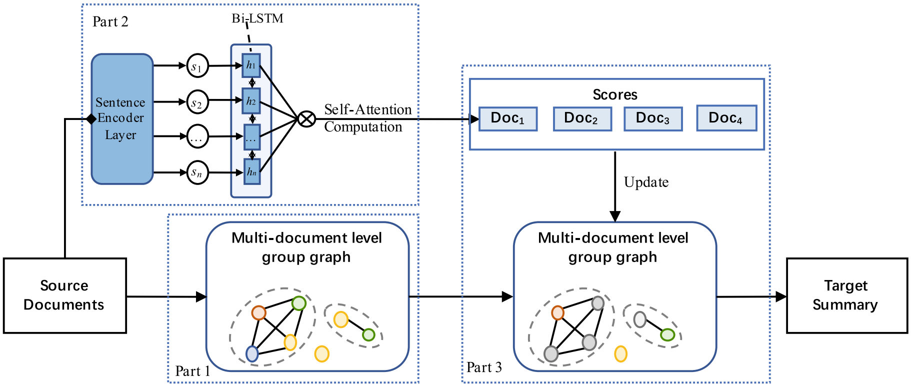


**Fig. 1.** The pipeline of the model. (Part 1) is a global frequency graph for each document pair. Each group is a cluster of similar sentences. (Part 2) is to implement self-attention for every single document. It contains a sentence encoder and a Bi-LSTM neural network for the self-attention computation. (Part 3) is the final scoring step which takes both the result of (Part 1) and (Part 2) into account to update the final score of each sentence.

## 3 SAFA Model

This section describes the specific details of the proposed SAFA model and explains how the model combines position and frequency information together. Figure 1 illustrates the pipeline of the SAFA model. It consists of three parts: (1) a global frequency graph at multi-document level generating sentence groups for each document pair to incorporate frequency information, (2) a Bi-LSTM neural network applying the position-aware self-attention mechanism to every single document, and (3) a sentence scoring algorithm to take both frequency (from part 1) and position information (from part 2) into account to update scores indicating salience of sentences.

### 3.1 Frequency Graph

This sub-section provides the details of how the frequency graph is generated for each document pair. The graph is used to represent the sentences and measure the similarity between them. The construction of the graph for each document pair takes all sentences together. A sentence is first randomly selected in each document pair. For a new sentence having any cosine similarity with the selected sentence greater than a threshold $N_t$, it is put into the group of the selected sentence with the greatest cosine similarity. Otherwise, it waits for another comparison or join no group at the end. The process terminates if each sentence belongs to one and only one group.

Sentences within the group $G_i$ are connected by edges and are considered as neighboring nodes to each other. We define the centrality $C_j^i$ for sentence $s_j$ in group $G_i$ containing the frequency information as the average similarity between node $j$ and its neighboring nodes. The centrality measures how similar the sentence is with other sentences in the same group.

$$C_j^i = \sum_{s_k \in G_i / s_j} CosSim(s_j, s_k) \tag{1}$$

where $G_i/s_j$ is the set of sentences in group $G_i$ except $s_j$, and $n$ is the group size. If the sentence does not belong to any group, its sentence centrality is 0. Each node will have its *Centrality* as its initial value for *Score*.

Therefore, a group is the collection of all similar sentences. That is, these sentences describe similar things within one document or among multiple documents. The sentence centrality for each node in this global graph contains information on the frequency of the sentence and is set as the initial score for the scoring model.

### 3.2 Self-attention for Single Document

This sub-section explains how the self attention which takes the position information of each sentence is computed within one document.

A Bi-LSTM is applied to extract the representation of each sentence within the single document it belongs to. For each document, the sentence embedding is calculated by BERT [4], and the sentence embeddings are the inputs of the Bi-LSTM neural network. The hidden states $\mathbf{H}$ are then used by the self-attention mechanism to get the vector of self-attention values $\mathbf{a}$ for all sentences in the single document. The self-attention vector is calculated as:

$$\mathbf{a} = \mathit{softmax}(\mathbf{v}\mathit{tanh}(\mathbf{W}_h \mathbf{H}^T)) \tag{2}$$

where $\mathbf{W}_h$ is a learnable weight and $\mathbf{v}$ is a learnable vector.

For each sentence $s_i$, we calculate a pseudo training target $t_i$ for it. The target acts as a "gold score" for training. The $t_i$ for each sentence $s_i$ in a single source document is the sum of the cosine similarities between $s_i$ and all sentences in the corresponding gold summary. A $softmax$ function is applied to make the sum of all $t_i$ in one document to 1:

$$t_i = softmax(\sum_{s_j \in G_i} Cos(s_i, s_j)) \tag{3}$$

The (a) graph in Fig. 2 shows the distribution of the average target values for sentences in the same position. The target values are monotonically decreasing, indicating that the training target is position-aware. In other words, since the training process of obtaining the self-attention value incorporates the position information, the self-attention value will also take position into account.

### 3.3 Sentence Scoring

This subsection shows how the final score of each sentence is computed and how the final extractive summaries are constructed.

A *Score* combining information of both frequency and position is updated for each sentence, indicating the salience of a sentence at both single document level and multi-document level. The $Score_i$ for $s_i$ is updated as:

$$Score_i = \mathbf{a}[i] + C_i * N_{G_j} \tag{4}$$

where $\mathbf{a}[i]$ is the self-attention value containing positional information and salience of $s_i$ at single document level; $Centrality_i$ contains frequency information at multi-document level. Since each group is allowed to output only one sentence in the selection process to reduce redundancy, we multiply the $Centrality_i$ with the size of group $N_{G_j}$ to indicate that large groups (i.e. sentences with high frequency) should have more weight.

Finally, the sentence in each group with the highest score is selected to form the summary. The summary thus considers the frequency and positional information of sentences, as well as the salience at both single document and multi-document level.

## 4 Experiments and Result

### 4.1 Dataset and Experimental Setup

The model is trained and tested on the large scale Multi-News MDS dataset[7]. Multi-News is a large-scale dataset for multi-document summarization tasks. There are 44,972 document pairs in the training set, 5622 pairs for both validation and test sets. Each document pair contains a gold summary and 2 to 10 documents describing the same event. The threshold $N_t$ of forming a group is 0.3. For the Bi-LSTM self-attention part, the size of the hidden state and the dimension of the self-attention weight matrix are 300.

### 4.2 Experimental Result

We compare our model with six representative models that perform well on Multi-News. The first three are extractive; The others are abstractive:

- *LexRank* [6] is a graph-based method for computing relative importance.
- *TextRank* [22] is a ranking model based on the global graph. It uses eigenvectors to compute the importance of each sentences.
- *MMR* [2] combines query-relevance with information-novelty, ranking the sentences based on relevance and redundancy.
- *Hi-MAP* [7] uses a pointer-generator and incorporates the MMR algorithm to generate the summary.
- PG-MMR [15] is a pointer-generator model, applying the MMR algorithm in the encoder to filter out unimportant sentences, and then put all important sentences in the decoder to form a summary.

**Table 1.** ROUGE F1 scores for models trained and tested on the Multi-News dataset. Models with "*" are abstractive.

| Method | R-1 | R-2 | R-SU |
|---|---|---|---|
| PG-MMR* | 40.55 | 12.36 | 15.87 |
| CopyTransformer* | 43.57 | 14.03 | 17.37 |
| GraphSum* | **45.70** | **17.12** | **19.06** |
| LexRank | 38.27 | 12.70 | 13.20 |
| TextRank | 38.44 | 13.10 | 13.50 |
| MMR | 38.77 | 11.98 | 12.91 |
| Hi-MAP | 43.47 | 14.89 | 17.41 |
| **SAFA** (our model) | **45.47** | **15.91** | **18.87** |

**Table 2.** Ablation study on the Multi-News dataset.

| Model | R-1 | R-2 | R-SU |
|---|---|---|---|
| *w/o* group | 37.93 | 11.77 | 13.55 |
| *w/o* position | 43.65 | 14.98 | 17.44 |
| **SAFA** | **45.47** | **15.91** | **18.87** |

- *CopyTransformer* [10] randomly chooses an attention head to copy the distribution.
- *GraphSum* [16] puts graphs into a neural network with structure similar to the Transformer.

We evaluate all models with the automatic ROUGE metric using version "ROUGE-1.5.5" on both the dataset and the Multi-News dataset. It has three ROUGE values for each model: the overlaps of unigrams (R-1), bigrams (R-2), and the skip-bigrams with a maximum distance of four words (R-SU).

Table 1 shows the results of all models, with the performances of the three abstractive models listed at top, the three extractive models at middle, and the proposed model at the bottom. We could see that overall our model performs competitively in terms of R-1 and R-SU. For extractive models, SAFA outperforms the other models for all ROUGE values. To evaluate the overall quality of the generated summary, we implement human evaluation.

### 4.3 Ablation Study

We implement three experiments to demonstrate that the information of both position and frequency are indispensable for generating high quality summary:

- *w/o* **group** chooses sentences based on the self-attention value only. Since the average of the self-attention values among sentences at the same position is proportional to the position, this experiments only takes the positional information into account;

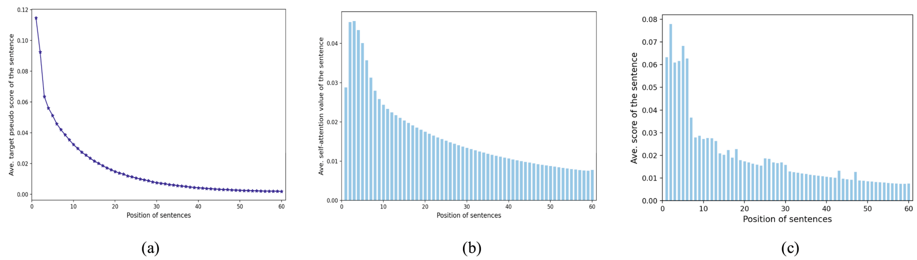


**Fig. 2.** (a) The trend of average pseudo score for different positions; (b) The average self-attention value of sentences at each position; The posture of the value is on decrease. Although first two sentences are smaller than the following sentences, the overall trend is position-aware; (c) The average score of sentences at each position. After the accumulation of the value of the group information, the posture of score become position-agnostic for many positions.

**Table 3.** The precision value of first three sentences for $w/o$ group and $w/o$ position. Because this experiment only considers the first three sentences, precision is a better metric than ROUGE.

| Model | R-1-p | R-2-p | R-l-p |
|---|---|---|---|
| $w/o$ group | 33.40 | 11.63 | 17.82 |
| $w/o$ position | **54.75** | **16.99** | **30.91** |

- $w/o$ **position** groups are ranked in descending order based on sizes. Then one sentence is chosen randomly from each group starting from the largest group until the summary reaches the minimum length. Groups with only one sentence will not be considered. Since the groups are formed based on frequency information only, this experiment exclude the positional information;
- **SAFA** choose sentences based on the newly defined score of our model, which balances the frequency and positional information.

The ROUGE scores of the three experiments are shown in Table 2. We could see that the best result is achieved when the model SAFA takes both the frequency and the positional information into account. In addition, (b) in Fig. 2 shows the average self attention values of the sentences at the same position. (c) in Fig. 2 shows the histograms of the average values of the newly defined scores for sentences at the same position. From (b) in Fig. 2, we could see that the self attention values are strictly decreasing except for the first two positions, which further demonstrates that the self-attention values contains positional information. (c) in Fig. 2, however, shows that the incorporation of the frequency information makes the scores not so "position-aware" that it provides chances for the significant sentences appearing late in the documents to be included in the summary.

**Table 4.** Human evaluation according to informativeness, fluency and non-redundancy.

| Method | Inf. | Flu. | Non-Redu. |
|---|---|---|---|
| GraphSum | 32 | 28 | 34 |
| SAFA | 41 | 30 | 29 |
| GOLD | 44 | 50 | 50 |

We also test the precision of outputting the first three sentences for experiments without position and without frequency group to see which factor works better in a specific range. The results are shown in Table 3. While the "*w/o* position" experiment selects the fist three sentences with the highest attention, the "*w/o* group" experiment selects one sentence from each of the groups with the three largest sizes. The higher precision demonstrates that within the range of the first three sentences of gold summaries, there are more sentences with high frequency than sentences appearing early. In other words, the frequency information plays a significant role thus should be incorporated.

### 4.4 Human Evaluation

Since the rouge scores could not assess the quality of a summary comprehensively, we conduct human evaluations from the levels of informativeness, non-redundancy, and fluency to test whether a summary is complete, precise, and readable. Specifically, informativeness measures whether the summary contains all important information and details; non-redundancy checks whether the summary is precise; and fluency measures whether the summary is written with correct grammar. 50 document pairs are randomly selected from the Multi-News testset. Two models: GraphSum and the proposed SAFA model are implemented to generate summaries. These generated summaries with the gold summaries are distributed to 10 native speakers with 5 pairs to each. The person is asked to add 1 to the level (i.e. informativeness, fluency, non-redundancy) if a summary performs well on this level. The person will not know which system the summary is from, or whether the summary is the gold one beforehand.

The human evaluation results of the three systems and the gold summaries are shown in Table 4. Since our SAFA model balances both the position information and frequency, the information contained in the sentences that appear frequently but late is also included. Therefore, our model ranks top for informativeness and fluency.

## 5 Conclusion

To better address the special characteristics of MDS, we propose a scoring model that takes both the frequencies and the position of sentences into account The introduction of the frequency graph with the self-attention mechanism to update the calculation of sentence scores enables the model to outperform the other extractive models on the Multi-News dataset and rank at the top for the informativeness and fluency in the human evaluation procedure.

## References

1. An, N.N., Thanh, N.Q., Liu, Y.: Deep CNNs with self-attention for speaker identification. IEEE Access **7**, 85327–85337 (2019). https://doi.org/10.1109/ACCESS.2019.2917470, https://ieeexplore.ieee.org/document/8721628/
2. Carbonell, J., Goldstein, J.: The use of MMR, diversity-based reranking for reordering documents and producing summaries. In: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval - SIGIR 1998, New York, New York, USA, pp. 335–336. ACM Press (1998). https://doi.org/10.1145/290941.291025. http://portal.acm.org/citation.cfm?doid=290941.291025
3. Cheng, J., Lapata, M.: Neural summarization by extracting sentences and words. In: 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers (2016). https://doi.org/10.18653/v1/p16-1046
4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: NAACL HLT 2019–2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference (2019)
5. Edmundson, H.P.: New methods in automatic extracting. J. ACM **16**(2), 264–285 (1969). https://doi.org/10.1145/321510.321519. https://dl.acm.org/doi/10.1145/321510.321519
6. Erkan, G., Radev, D.R.: LexRank: graph-based lexical centrality as salience in text summarization. J. Artif. Intell. Res. **22**, 457–479 (2004). https://doi.org/10.1613/jair.1523. http://arxiv.org/abs/1109.2128, https://jair.org/index.php/jair/article/view/10396
7. Fabbri, A.R., Li, I., She, T., Li, S., Radev, D.R.: Multi-news: a large-scale multi-document summarization dataset and abstractive hierarchical model, pp. 1074–1084 (2019). https://doi.org/10.18653/v1/p19-1102. http://arxiv.org/abs/1906.01749
8. Gao, Y., Zhao, W., Eger, S.: SUPERT: Towards New Frontiers in Unsupervised Evaluation Metrics for Multi-Document Summarization (2020). https://doi.org/10.18653/v1/2020.acl-main.124
9. Gehrmann, S., Deng, Y., Rush, A.: Bottom-Up Abstractive Summarization (2019). https://doi.org/10.18653/v1/d18-1443
10. Gehrmann, S., Deng, Y., Rush, A.M.: Bottom-up abstractive summarization. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018, August 2018. https://doi.org/10.18653/v1/d18-1443. http://arxiv.org/abs/1808.10792

11. Hariharan, S., Srinivasan, R.: Studies on graph based approaches for singleand multi document summarizations. Int. J. Comput. Theory Eng. 519–526 (2009). https://doi.org/10.7763/IJCTE.2009.V1.84. http://www.ijcte.org/show-26-177-1.html
12. Katragadda, R.: Sentence Position revisited: a robust light-weight Update Summarization 'baseline' Algorithm. Computational Linguistics (2009)
13. Kitaev, N., Cao, S., Klein, D.: Multilingual constituency parsing with self-attention and pre-training. In: ACL 2019–57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (2020). https://doi.org/10.18653/v1/p19-1340
14. Kitaev, N., Klein, D.: Constituency parsing with a self-attentive encoder. In: ACL 2018–56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers), May 2018. https://doi.org/10.18653/v1/p18-1249. http://arxiv.org/abs/1805.01052
15. Lebanoff, L., Song, K., Liu, F.: Adapting the neural encoder-decoder framework from single to multi-document summarization. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018, August 2018. https://doi.org/10.18653/v1/d18-1446. http://arxiv.org/abs/1808.06218
16. Li, W., Xiao, X., Liu, J., Wu, H., Wang, H., Du, J.: Leveraging Graph to Improve Abstractive Multi-Document Summarization, pp. 6232–6243 (2020). https://doi.org/10.18653/v1/2020.acl-main.555
17. Lin, C.Y.: Rouge: a package for automatic evaluation of summaries. In: Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004) (2004)
18. Lin, C.Y., Hovy, E.: Identifying topics by position. In: Proceedings of the Fifth Conference on Applied Natural Language Processing, Morristown, NJ, USA, pp. 283–290. Association for Computational Linguistics (1997). https://doi.org/10.3115/974557.974599. http://portal.acm.org/citation.cfm?doid=974557.974599
19. Lin, Z., et al.: A structured self-attentive sentence embedding. In: 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings, March 2017. http://arxiv.org/abs/1703.03130
20. Liu, P.J., et al.: Generating Wikipedia by summarizing long sequences. In: 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings (2018)
21. Lunh, H.P.: The automatic creation of literature abstracts. IBM J. Res. Dev. **2**, 159–165 (1958)
22. Mihalcea, R., Tarau, P.: TextRank: bringing order into texts. In: Proceedings of EMNLP (2004)
23. Mnih, V., Heess, N., Graves, A., Kavukcuoglu, K.: Recurrent models of visual attention. In: Advances in Neural Information Processing Systems, June 2014. http://arxiv.org/abs/1406.6247
24. Nallapati, R., Zhou, B., dos Santos, C., Gulçehre, Ç., Xiang, B.: Abstractive text summarization using sequence-to-sequence RNNs and beyond. In: CoNLL 2016–20th SIGNLL Conference on Computational Natural Language Learning, Proceedings (2016). https://doi.org/10.18653/v1/k16-1028
25. Narayan, S., Cohen, S.B., Lapata, M.: Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018 (2018). https://doi.org/10.5281/zenodo.2399762
26. Nenkova, A.: Automatic text summarization of newswire: lessons learned from the document understanding conference. In: Proceedings of the National Conference on Artificial Intelligence (2005)

27. Ouyang, Y., Li, W., Lu, Q., Zhang, R.: A study on position information in document summarization. In: COLING 2010–23rd International Conference on Computational Linguistics, Proceedings of the Conference (2010)
28. Radev, D.R., Jing, H., Styś, M., Tam, D.: Centroid-based summarization of multiple documents. Inf. Process. Manag. **40**(6), 919–938 (2004). https://doi.org/10.1016/j.ipm.2003.10.006. https://linkinghub.elsevier.com/retrieve/pii/S0306457303000955
29. Schilder, F., Kondadadi, R.: FastSum: fast and accurate query-based multi-document summarization. In: ACL 2008: HLT - 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference (2008)
30. See, A., Liu, P.J., Manning, C.D.: Get to the point: summarization with pointer-generator networks. ACL 2017–55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers), vol. 1, pp. 1073–1083 (2017). https://doi.org/10.18653/v1/P17-1099
31. Thakkar, K.S., Dharaskar, R.V., Chandak, M.B.: Graph-based algorithms for text summarization. In: 2010 3rd International Conference on Emerging Trends in Engineering and Technology, pp. 516–519. IEEE (2010). https://doi.org/10.1109/ICETET.2010.104. http://ieeexplore.ieee.org/document/5698380/
32. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30, June 2017. http://arxiv.org/abs/1706.03762
33. Yih, W.T., Goodman, J., Vanderwende, L., Suzuki, H.: Multi-document summarization by maximizing informative content-words. In: IJCAI International Joint Conference on Artificial Intelligence (2007)
34. Zheng, H., Lapata, M.: Sentence centrality revisited for unsupervised summarization. In: ACL 2019–57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, vol. 2, pp. 6236–6247 (2020). https://doi.org/10.18653/v1/p19-1628