



Unsupervised Extractive Text Summarization with Distance-Augmented Sentence Graphs

Jingzhou Liu
Carnegie Mellon University
liujingzhou@cs.cmu.edu

Dominic J. D. Hughes
UC Berkeley
dominic@theory.stanford.edu

Yiming Yang
Carnegie Mellon University
yiming@cs.cmu.edu

ABSTRACT

Supervised summarization has made significant improvements in recent years by leveraging cutting-edge deep learning technologies. However, the true success of supervised methods relies on the availability of large quantity of human-generated summaries of documents, which is highly costly and difficult to obtain in general. This paper proposes an unsupervised approach to extractive text summarization, which uses an automatically constructed sentence graph from each document to select salient sentences for summarization based on both the similarities and relative distances in the neighborhood of each sentences. We further generalize our approach from single-document summarization to a multi-document setting, by aggregating document-level graphs via proximity-based cross-document edges. In our experiments on benchmark datasets, the proposed approach achieved competitive or better results than previous state-of-the-art unsupervised extractive summarization methods in both single-document and multi-document settings, and the performance is competitive to strong supervised baselines.

CCS CONCEPTS

• Information systems → Summarization; • Computing methodologies → Natural language processing.

KEYWORDS

unsupervised summarization; extractive summarization; graph-based method

ACM Reference Format:

Jingzhou Liu, Dominic J. D. Hughes, and Yiming Yang. 2021. Unsupervised Extractive Text Summarization with Distance-Augmented Sentence Graphs. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21), July 11–15, 2021, Virtual Event, Canada*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3404835.3463111>

1 INTRODUCTION

Text summarization is the task of condensing a given document or a set of documents into a shorter piece of textual summary (a.k.a. single-document or multi-document summarization), which preserves the main contents of the input. Existing approaches can be

characterized into two kinds, i.e., extractive vs. abstractive. The extractive methods compose each summary by extracting a subset of passages (sentences or phrases) from the input text. The abstractive methods produce each summary based on an underlying generative model, where the output may include the words or phrases beyond the input text. Generally speaking, extractive summaries are piecewise fluent and accurate as they are directly selected from the input text, while abstractive summaries could be globally more coherent and/or versatile. Both kinds of approaches have been intensively studied [2, 13, 15, 17, 21]. In this paper, we focus on improving the state of the art in unsupervised extractive text summarization.

Representative examples of early work in extractive summarization include the Maximum-Marginal Relevance (MMR) principle [1] for selecting sentences based on both their relevance (to the central theme of the document) and the diversity of the selected sentences, and the PageRank scores [14] of sentences (as nodes) in sentence-similarity graphs [4, 11]. Those methods do not require any human-generated summaries for training and hence are unsupervised by nature. Recent methods in text summarization focus more on supervised neural networks, which adapt sequence-to-sequence translation, reinforcement learning and large-scale pretraining techniques to text summarization settings. Significant performance improvements in both extractive and abstractive summarization have been reported [6, 8, 13, 15, 17]. However, a fundamental limitation of those supervised methods is that their success heavily depends on the availability of large training corpora with human-generated high-quality summaries [5, 7, 12, 16], which are extremely costly to produce and difficult to obtain.

To avoid the expensive cost of data annotation, unsupervised neural learning for text summarization has received an increasing amount of attentions recently [9, 18, 19, 22]. One representative work is a new graph-based method named PACSUM proposed by [20]. Similar to previous graph-based summarization methods like PageRank for sentence selection [4, 11], PACSUM also uses a graph to represent sentences as the nodes and to encode the semantic similarity among sentences via weighted edges. The main differences in PACSUM are that 1) it uses *directed* links to encode both the semantic-similarity information and the partial-order information (which sentence is mentioned earlier than the other) for sentence pairs, which is more expressive than the previous models, and 2) it leverages large-scale neural pretraining (BERT [3]) for context sensitive embedding of sentences, which is a benefit from recent developments in deep learning. As a result, PACSUM is representative for the state-of-the-art performance in unsupervised extractive summarization, and achieves competitive results as supervised methods on evaluation benchmarks (CNNDM [7] and NYT [5]).



This work is licensed under a Creative Commons Attribution International 4.0 License.

SIGIR '21, July 11–15, 2021, Virtual Event, Canada.

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8037-9/21/07.

<https://doi.org/10.1145/3404835.3463111>

Inspired by the success of PACSUM, we seek further enhancements in unsupervised extractive summarization as the focus of this paper. Firstly, for the task of single-document summarization we propose to improve PACSUM by introducing our Distance-Augmented Sentence Graphs (DASG) formalism, which is more expressive for graph-based learning of sentence importance with respect to summarization. Secondly, we propose a generalization of the DASG formalism from the single-document setting to a multi-document setting, which is a less explored area of research, especially for graph-based text summarization methods.

We argue that the graph formulation in PACSUM may not be sufficient for fully leveraging the positional information of sentences within a document, as it only encode the pairwise ordering (which sentence is mentioned before the other sentence) but not the distance information (how far the two sentences are apart from each other). As an intuitive example, if two sentences are similar and adjacent in the input document, then the second one is often commenting on, or supplementing the first one, and hence may not necessarily contribute a strong evidence for the global importance of the first sentence with respect to summarization. On the other hand, if the two similar or identical sentences are apart from each other, then the repetitive mentions may be a good indicator for the importance (or centrality) of those sentences. Through this example, we see that encoding only the pairwise ordering of sentences is not sufficient for fully leveraging the positional information about sentences; instead, encoding the relative distances of sentence pairs is a better alternative. This leads to our key idea behind the proposed DASG approach for edge weighting in graph construction for single-document extractive summarization (see Section 2 for detailed description). The proposed model obtained the best results among unsupervised methods on two benchmark summarization datasets, and the performance is comparable to strong supervised baselines.

As the second contribution in this paper, we generalize the DASG formalism to the multi-document settings, as a significant step forward. Compared to the intensive studies on single-document summarization [4, 11], multi-document models are much less explored, especially with respect to unsupervised neural graph-based methods. We propose to accomplish multi-document summarization by first generating the DASGs for individual documents, and then merging the single-document DASGs into a multi-document DASG via proximity-based cross-document edges. The resulting model achieves competitive results with state-of-the-art unsupervised multi-document extractive summarization methods.

2 DISTANCE-AUGMENTED SINGLE-DOCUMENT SUMMARIZATION

2.1 Sentence Graph and Centrality

In unsupervised graph-based extractive summarization, the document is represented as a graph, where each node represents a sentence in the input document. Let D denote the input document, consisting of n sentences, $D = (s_1, s_2, \dots, s_n)$, and edge from node i to node j has weight e_{ij} . Centrality of a node measures its importance within a graph. In graph-based summarization methods, centrality is used to select the most salient sentence to construct summaries through ranking.

Traditional methods construct the sentence graph as an undirected graph $e_{ij} = e_{ji}$ (such as TextRank [11] and LexRank [4]), and centrality scores are computed via running PageRank. PACSUM, on the other hand, constructs a directed graph, and the centrality score of a node is computed by simply aggregating its incoming and outgoing edge weights:

$$\text{Centrality}(s_i) = \sum_{i \neq j} e_{ij}. \quad (1)$$

Edge weights are defined as:

$$e_{ij} = \begin{cases} \lambda_1 \times \text{sim}(s_i, s_j) & i < j \\ \lambda_2 \times \text{sim}(s_j, s_i) & i > j \end{cases}, \quad (2)$$

where λ_1 and λ_2 are system hyper-parameters. This asymmetric graph formulation assumes that the connection of two nodes has unequal contributions to their centrality scores based on the relative positions of the two sentences in the document.

2.2 Distance-Augmented Sentence Graph (DASG)

The sentence graphs in existing works (TextRank, LexRank, PACSUM) only consider sentence similarities and the order of a pair of sentences, but do not fully characterize the complex structures in the original documents. Intuitively, for a pair of similar sentences s_i and s_j , if s_i and s_j are next to each other in a paragraph, s_j is probably commenting or continuing s_i , and so this pair does not provide us enough information to adjust the importance of s_i or s_j ; but if they are far away from each other (e.g. one appears at the beginning and the other appears at the end), it is likely that these two sentences contain important information that needs to be reiterated. Therefore, besides sentence similarities and relative orders, the distance between a pair of sentences in the document is also important for more fine-grained modeling.

To exploit the idea of augmenting directed sentence graphs with distances, we propose the following edge weighting scheme:

$$e_{ij} = \delta(i, j) \times \text{sim}(s_i, s_j), \quad (3)$$

where $\delta(\cdot, \cdot)$ is a function that maps a pair of sentence indices to a coefficient decided by relative distance. Suppose the input document consists of n sentences. Theoretically each choice of distance $(1, \dots, n-1)$ should have a unique coefficient associated with it, but for efficiency and better generalization, we evenly divide these distances into k groups, where each group contains $m = \lceil n/k \rceil$ distances, and the distances within the same group has the same coefficient. We design the $\delta(\cdot, \cdot)$ function as the following step function:

$$\delta(i, j) = \begin{cases} \lambda_1^+ \lfloor \frac{i-j}{m} \rfloor + 1 & i < j \\ \lambda_1^- \lfloor \frac{i-j}{m} \rfloor + 1 & i > j \end{cases}, \quad (4)$$

where $\lambda_1^+, \dots, \lambda_k^+$ and $\lambda_1^-, \dots, \lambda_k^-$ are fixed hyper-parameters. Usually we set k to be 3, and more analysis can be found in Section 4.3. We use the aggregating approach to computing centrality scores:

$$\text{centrality}(s_i) = \sum_{i \neq j} \delta(i, j) \times \text{sim}(s_i, s_j) \quad (5)$$

Here sentences are encoded by a BERT model, and the similarity function $\text{sim}(\cdot, \cdot)$ is implemented as the inner product of the two sentence representations.

3 ADAPTION TO MULTI-DOCUMENT SUMMARIZATION

In this section we introduce a general framework for adapting our graph-based method to multi-document settings. Suppose we have t input documents $\mathcal{D} = \{D_1, \dots, D_t\}$, and the i -th document consists of n_i sentences, $D_i = (s_1^i, s_2^i, \dots, s_{n_i}^i)$. First we construct a graph for each input document using the single-document method, where the i -th document's graph is $(\{s_j^i\}, \{e_{**}^i\})$, resulting in t disconnected directed sub-graphs, and next we will discuss two options of adding cross-document links to connect all these sub-graphs.

3.1 Undirected Cross-Document Edges

For two sentences in two different documents, s_j^i and $s_{j'}^{i'}$, we want to add an edge connecting them to measure their cross-document centrality. Following the same idea in single-document summarization, the edge weights should be based on sentence similarities and relative importance, but unlike sentence pairs within a document, there is no information to determine whether one sentence is more important than another sentence in a different document, since the input documents usually do not have an order. A simple approach is to simply add undirected cross-document edges:

$$e_{jj'}^{ii'} = e_{j'j}^{i'i} = \text{sim}(s_j^i, s_{j'}^{i'}). \quad (6)$$

3.2 Proximity-Based Cross-Document Edges

Similar to the situation in single-document summarization, the symmetry of undirected edges weakens the model's power of selecting salient sentences, and ideally we desire *directed* edges between a pair of sentences so that we can distinguish the more important one. Consider a pair of sentences from two documents, s_j^i and $s_{j'}^{i'}$. A sentence in a document and its proximate sentences (the ones immediately before and after it) often talk about relevant content, including comments, supplementary information, etc., and so we can compute the similarity between s_j^i and proximity of $s_{j'}^{i'}$ (i.e., $(s_{j'-1}^{i'}, s_{j'}^{i'}, s_{j'+1}^{i'})$), and similarly, the similarity between $s_{j'}^{i'}$ and proximity of s_j^i . Intuitively, if the similarity of one direction is higher than the other (suppose s_j^i and proximity of $s_{j'}^{i'}$ has higher similarity), it is a reasonable indicator that s_j^i carries more information and contains more comprehensive content than $s_{j'}^{i'}$ does. Based on this idea, we propose proximity-based cross-document edge weight scheme:

$$\begin{aligned} e_{jj'}^{ii'} &= \text{sim}(s_{j'}^{i'}, [s_{j-1}^i, s_j^i, s_{j+1}^i]) \\ e_{j'j}^{i'i} &= \text{sim}(s_j^i, [s_{j'-1}^{i'}, s_{j'}^{i'}, s_{j'+1}^{i'}]) \end{aligned} \quad (7)$$

To extract representation of a sentence's proximity, we simply concatenate this sentence with its proximate sentences and feed them to a BERT encoder.

3.3 Multi-Document Centrality

Aggregating intra- and cross-document edges, we now can compute multi-document centrality score of a sentence:

$$\begin{aligned} &\text{centrality}(s_j^i) \\ &= \sum_{j' \neq j} \delta(j, j') \times \text{sim}(s_j^i, s_{j'}^{i'}) + \gamma \times \sum_{i' \neq i} \sum_{j'} \text{sim}(s_j^i, [s_{j'-1}^{i'}, s_{j'}^{i'}, s_{j'+1}^{i'}]), \end{aligned} \quad (8)$$

where γ is a hyper-parameter that balances intra- and cross-document edge weights, and if a sentence is the first or the last one in a document, we simply discard out-of-bound indices. Sentences across all documents are ranked by their centrality scores and top ranked sentences are preserved as the summary.

4 EXPERIMENTS

4.1 Datasets

We conduct experiments on three benchmark summarization datasets in news domain, two for single-document summarization (CNN/Daily Mail [7] and NYT [16]), and one for multi-document summarization (Multi-News [5]). For all datasets, we only use their validation and test set in our experiments.

4.2 Implementation Details

We use the publicly available BERT-base pre-trained model¹, and the BERT model is fine-tuned on the input documents and then fixed throughout the summarization experiments. In all experiments, the number of groups in the distance-augmented graph (k in Section 2.2) is set to 3. All hyper-parameters ($\gamma, \lambda_1^+, \dots, \lambda_k^+, \lambda_1^-, \dots, \lambda_k^-$) are tuned on 2000 validation samples with ground truth summaries on all three datasets, $\gamma \in \{0.2, 0.4, 0.6, 0.8\}$, $\lambda_k^* \in \{-2, -1.5, \dots, 1.5, 2\}$. Due to the large search space, we do not use full grid-search, but heuristics to choose hyper-parameters. During decoding, our model extracts 4, 3 and 10 sentences as output summaries for CNNDM, NYT and Multi-News, respectively. Trigram blocking [15] is used on all three datasets.

4.3 Single-Document Experiments

Main results of single-document summarization are shown in Table 1. Row 1-3 are strong abstractive or extractive supervised methods [10, 13, 17]. Row 4-9 are previous unsupervised extractive methods, with PACSUM (BERT) (row 9) being the state-of-the-art method. Row 11 is our method using degree aggregation for centrality computing (equation 5). For comparison, we also include a method that runs PageRank on our distance-augmented graph (row 10).

Comparing row 10 and 11, we can see that based on the same graph, PageRank performs much worse than degree aggregation. This is probably due to the fact that PageRank cannot directly address edges with negative weights. Comparing row 11 with row 4-9, our proposed single-document summarization surpasses all unsupervised comparing methods. Specifically, our method (row 11) achieves better results than PACSUM (BERT) on both NYT and CNNDM datasets, showing the effectiveness of DASG which allows the model to better capture sentence relations.

¹<https://github.com/huggingface/transformers>

ID	Method	NYT			CNNDM		
		R-1	R-2	R-L	R-1	R-2	R-L
0	ORACLE	61.9	41.7	58.3	54.7	30.4	50.8
1	REFRESH	41.3	22.0	37.8	41.3	18.4	37.5
2	POINTER-GENERATOR	42.7	22.1	38.0	39.5	17.3	36.4
3	BERTSUMEXT (LARGE)	-	-	-	43.9	20.3	39.9
4	LEAD-3	35.5	17.2	32.0	40.5	17.7	36.7
5	DEGREE (TF-IDF)	33.2	13.1	29.0	33.0	11.7	29.5
6	TEXTRANK (TF-IDF)	33.2	13.1	29.0	33.2	11.8	29.6
7	TEXTRANK (BERT)	29.7	9.0	25.3	30.8	9.6	27.4
8	PACSUM (TF-IDF)	40.4	20.6	36.4	39.2	16.3	35.3
9	PACSUM (BERT)	41.4	<u>21.7</u>	37.5	40.7	17.8	36.9
10	OURS (PAGERANK)	37.7	17.8	32.9	36.2	14.6	33.5
11	OURS (AGGREGATION)	42.2	21.8	38.2	41.6	18.5	37.8

Table 1: Single-document summarization results. ROUGE F1 are reported. Row 0-2, 4-9 are taken from [20], and row 3 is taken from [10]. Bold indicates the best result among unsupervised methods. Underline means not statistically significantly worse than the best result in the same column among unsupervised methods.

Method	CNNDM		
	R-1	R-2	R-L
PACSUM (BERT)	40.7	17.8	36.9
OURS ($k=1$)	40.5	17.9	36.5
OURS ($k=2$)	41.3	18.3	37.2
OURS ($k=3$)	41.6	18.5	37.8
OURS ($k=4$)	41.7	18.7	37.8

Table 2: Single-document results with different k s.

To further investigate the influence of k (number of groups in Section 2.2), we conduct experiments on CNNDM with different choice of k . Results are shown in Table 2. When $k = 1$, all distances have the same coefficient when computing edge weights, and so for a pair of sentences, only their relative order matters. Therefore, when $k = 1$ our model is equivalent to PACSUM. In Table 2, PACSUM is slightly better than OURS ($k=1$), most likely due to slightly different BERT encoder finetuning procedures. As k increases, we can see that the ROUGE scores also increase, indicating that when k gets larger, the model is able to perform more fine-grained modeling. When k increases to 4, performance improvements seem to be saturated, probably because the model overfits to the small validation set. Another reason could be sub-optimal hyper-parameters due to the large search space when $k = 4$.

4.4 Multi-Document Experiments

ID	Method	Multi-News		
		R-1	R-2	R-L
0	LEAD-3	39.41	11.77	14.51
1	POINTER-GENERATOR	42.80	14.19	16.75
2	FLATTRANSFORMER	44.32	15.11	20.50
3	HI-MAP	43.47	14.89	17.41
4	GRAPHSUM	46.07	17.42	23.21
5	LEXRANK	38.27	12.70	13.20
6	TEXTRANK (TF-IDF)	38.44	13.10	13.50
7	SUMMPIP	42.32	13.28	16.20
8	OURS (UNDIRECTED)	41.31	<u>13.01</u>	14.66
9	OURS (PROXIMITY)	42.60	<u>13.22</u>	<u>16.15</u>

Table 3: Multi-document summarization results. Row 0-7 are taken from [5] or original paper. Bold indicates the best result among unsupervised methods. Underline means not statistically significantly worse than the best result in the same column among unsupervised methods.

Results of multi-document summarization experiments are shown in Table 3. Row 1-4 are strong supervised methods [3, 9, 17], and row 5-7 are unsupervised methods. SUMMPIP [19] is the state-of-the-art unsupervised method. It constructs sentence discourse graphs for each document and then perform spectral clustering, where each cluster generates one sentence in the final summary.

OURS (UNDIRECTED) means cross-document edges in the model are undirected, and OURS (PROXIMITY) means that proximity-based cross-document edges are used in the sentence graph. First, comparing row 8 with row 9, we can see that using directed proximity-based edges is much better than using undirected edges, verifying our assumption that proximity-based edges help model distinguish important sentences from different documents. Comparing row 9 with unsupervised methods (row 5-7), our method clearly surpasses LEXRANK and TEXTRANK baselines, and is comparable with state-of-the-art method SUMMPIP. SUMMPIP consists of a pipeline of complex operations, while our method provides a simple framework that adapts to any single document graph-based method. Comparing row 9 with row 1-4, our method is comparable with some supervised baselines (row 1,2,3), but is still far behind the state-of-the-art supervised method (row 4).

5 CONCLUSIONS

In this paper we have proposed a graph-based single-document unsupervised extractive method that constructs a distance-augmented sentence graph (DASG) from a document that enables the model to perform more fine-grained modeling of sentences and better characterize the original document structures. We further adapted the single-document model to multi-document settings, by connecting every input document’s sentence graph through proximity-based cross-document edges. Experimental results on three summarization benchmark datasets showed the effectiveness of our proposed methods. Both of our proposed single- and multi-document unsupervised methods either achieves or is comparable to state-of-the-art unsupervised methods, and both methods are competitive with strong supervised baselines.

Acknowledgement We thank all the reviewers for their helpful comments. This work is supported in part by National Science Foundation (NSF) under grant IIS-1546329.

REFERENCES

- [1] Jaime Carbonell and Jade Goldstein. 1998. 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*. 335–336.
- [2] Yen-Chun Chen and Mohit Bansal. 2018. Fast Abstractive Summarization with Reinforce-Selected Sentence Rewriting. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 675–686.
- [3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [4] Günes Erkan and Dragomir R Radev. 2004. Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research* 22 (2004), 457–479.
- [5] Alexander R Fabbri, Irene Li, Tianwei She, Suyi Li, and Dragomir R Radev. 2019. Multi-news: A large-scale multi-document summarization dataset and abstractive hierarchical model. *arXiv preprint arXiv:1906.01749* (2019).
- [6] Sebastian Gehrmann, Yuntian Deng, and Alexander M Rush. 2018. Bottom-up abstractive summarization. *arXiv preprint arXiv:1808.10792* (2018).
- [7] Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching machines to read and comprehend. In *Advances in neural information processing systems*. 1693–1701.
- [8] Chris Kedzie, Kathleen McKeown, and Hal Daume III. 2018. Content selection in deep learning models of summarization. *arXiv preprint arXiv:1810.12343* (2018).
- [9] Wei Li, Xinyan Xiao, Jiachen Liu, Hua Wu, Haifeng Wang, and Junping Du. 2020. Leveraging Graph to Improve Abstractive Multi-Document Summarization. *arXiv preprint arXiv:2005.10043* (2020).
- [10] Yang Liu and Mirella Lapata. 2019. Text summarization with pretrained encoders. *arXiv preprint arXiv:1908.08345* (2019).
- [11] Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*. 404–411.
- [12] Shashi Narayan, Shay B Cohen, and Mirella Lapata. 2018. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. *arXiv preprint arXiv:1808.08745* (2018).
- [13] Shashi Narayan, Shay B Cohen, and Mirella Lapata. 2018. Ranking sentences for extractive summarization with reinforcement learning. *arXiv preprint arXiv:1802.08636* (2018).
- [14] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. *The PageRank citation ranking: Bringing order to the web*. Technical Report. Stanford InfoLab.
- [15] Romain Paulus, Caiming Xiong, and Richard Socher. 2017. A deep reinforced model for abstractive summarization. *arXiv preprint arXiv:1705.04304* (2017).
- [16] Evan Sandhaus. 2008. The new york times annotated corpus. *Linguistic Data Consortium, Philadelphia* 6, 12 (2008), e26752.
- [17] Abigail See, Peter J Liu, and Christopher D Manning. 2017. Get to the point: Summarization with pointer-generator networks. *arXiv preprint arXiv:1704.04368* (2017).
- [18] Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter J Liu. 2019. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. *arXiv preprint arXiv:1912.08777* (2019).
- [19] Jiming Zhao, Ming Liu, Longxiang Gao, Yuan Jin, Lan Du, He Zhao, He Zhang, and Gholamreza Haffari. 2020. SummPip: Unsupervised Multi-Document Summarization with Sentence Graph Compression. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1949–1952.
- [20] Hao Zheng and Mirella Lapata. 2019. Sentence centrality revisited for unsupervised summarization. *arXiv preprint arXiv:1906.03508* (2019).
- [21] Qingyu Zhou, Nan Yang, Furu Wei, Shaohan Huang, Ming Zhou, and Tiejun Zhao. 2018. Neural Document Summarization by Jointly Learning to Score and Select Sentences. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 654–663.
- [22] Chenguang Zhu, Ziyi Yang, Robert Gmyr, Michael Zeng, and Xuedong Huang. 2019. Make Lead Bias in Your Favor: A Simple and Effective Method for News Summarization. *arXiv preprint arXiv:1912.11602* (2019).