Tractable & Coherent Multi-Document Summarization: Discrete Optimization of Multiple Neural Modeling Streams via Integer Linear Programming

Litton J Kurisinkel, Nancy F. Chen
Institute for Infocomm Research, A*STAR, Singapore
litton_kurisinkel, nfychen@i2r.a-star.edu.sg

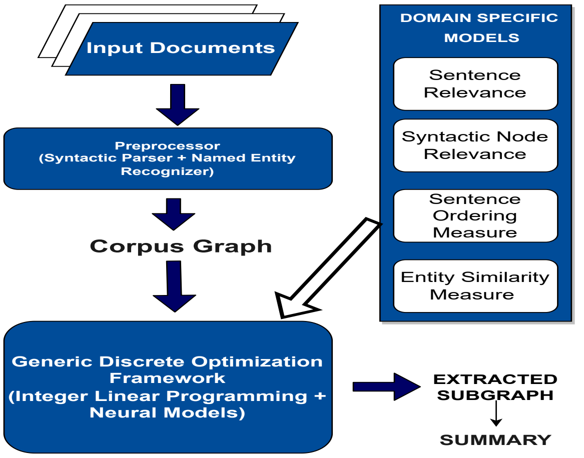

Figure 1: System Architecture(Ref. Figure 2 for a more detailed view)

Abstract

One key challenge in multi-document summarization is the generated summary is often less coherent compared to single document summarization due to the larger heterogeneity of the input source content. In this work, we propose a generic framework to jointly consider coherence and informativeness in multi-document summarization and offers provisions to replace individual components based on the domain of source text. In particular, the framework characterizes coherence through verb transitions and entity mentions and takes advantage of syntactic parse trees and neural modeling for intra-sentential noise pruning. The framework cast the entire problem as an integer linear programming optimization problem with neural and non-neural models as linear components. We evaluate our method in the news and legal domains. The proposed approach consistently performs better than competitive baselines for both objective metrics and human evaluation.

1 Introduction

Multi-Document summarization (MDS) approaches generate the summary of a corpus of documents consisting of a set of related topics. Extractive summarization techniques extract a subset of sentences which topically represent the input corpus in a stipulated summary space (Lin and Bilmes, 2011). On the other hand, abstractive summarization techniques constructs a semantic representation of the source text and constructs the summary in its own learnt writing style (Tan et al., 2017). Despite the attempts for abstractive summarization techniques using neural methods, extractive summarization techniques reserve its space for formulating ready to use summarization approaches. However, a set of (selected) sentences put together without considering the ordering and coherence of the content may not make much sense to the summary reader (Guinaudeau and Strube, 2013; Barzilay and Lapata, 2008). Hence, the to make such text generation applications more accessible to users, it is essential to improve qualitative dimensions such as *coherence*. Studies on human written summaries shows that generic information is more relevant for summary content while more specific information is considered to be irrelevant (Louis and Nenkova, 2011). In MDS, the unit of extraction is sentences. Long sentences could often contain irrelevant information that is not essential to the summary and thus regarded as *noisy* information (Knight and Marcu, 2000). Hence MDS systems should ideally be equipped for pruning intra-sentential noise.

Most previous work for extractive MDS gave less importance in improving summaries in qualitative dimensions such as coherence and provided an incoherent reading to the summary reader(Takamura and Okumura, 2009). A subset of previous work aimed at removing intra-sentential noise by sentence compression (Berg-Kirkpatrick et al., 2011). Such works were successful in removing intra- sentential noise, however the problem of incoherent reading can be more severe as the parts of the sentences pruned away can be important

Proceedings of EMNLP 2022 Industry Track, pages 247–253
December 9–11, 2020. ©2022 Association for Computational Linguistics

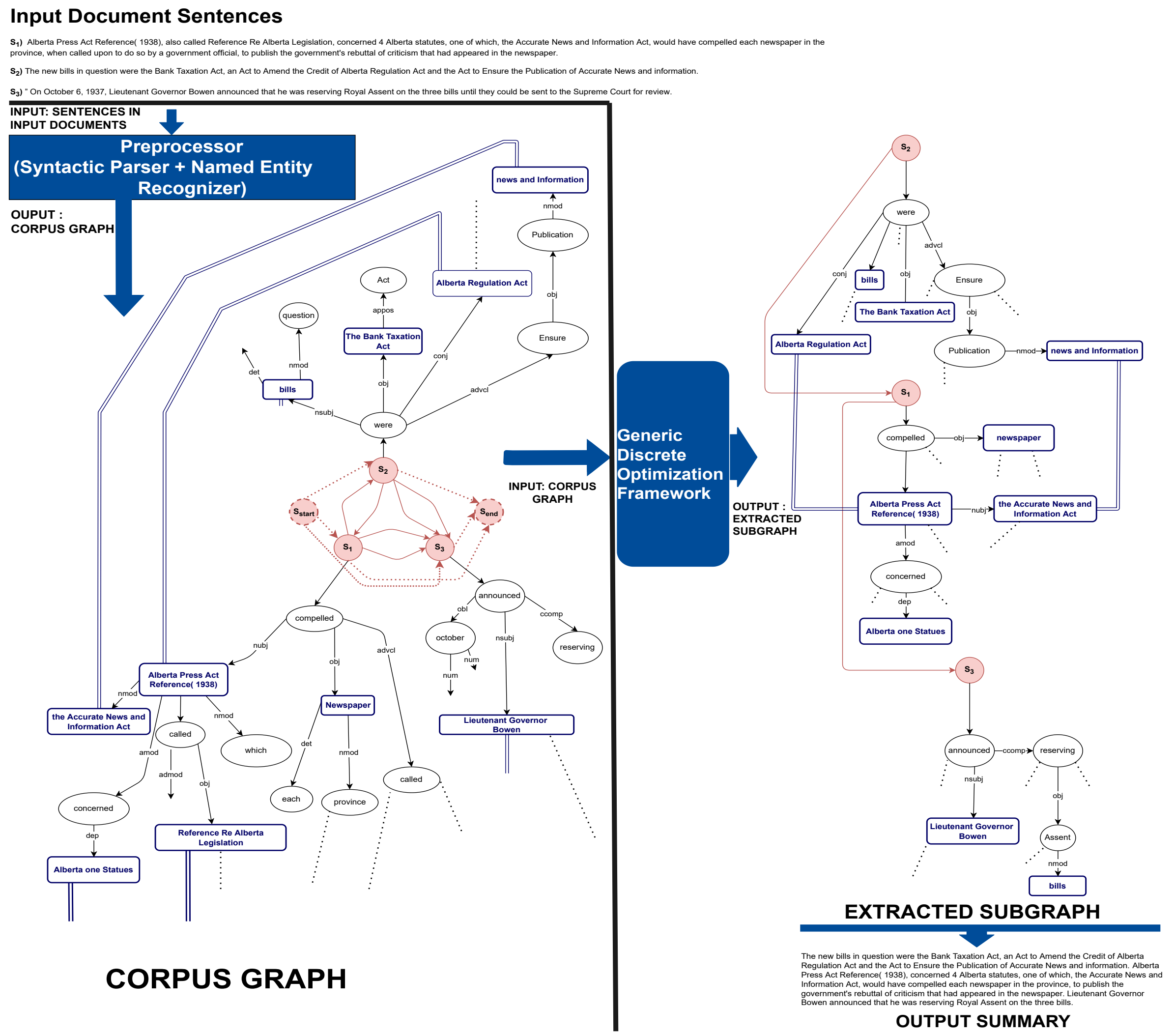

Figure 2: Flowchart for Coherent and Noise-Free Multi-Document Summarization with Input and Intermediate and Final Outputs: Figure depicts the summarization of an input corpus containing three sentences. Left side of the figure depicts the corpus graph constructed while right portion depicts the extracted sub-graph. Sentence nodes and sentence pair edges are depicted in red, syntactic tree nodes and edges are depicted in black and entity nodes, and entity pair edges are depicted in blue.

for topical continuity and for coherence. Very few works formulated methods to improve topical coherence between sentences in the summary by constructing corpus level discourse graph or by computing entity transition probabilities (Christensen et al., 2013; Wang et al., 2016). However such works had no mechanism for pruning intra- sentential noise. Through the current work we present a hybrid approach using Integer Linear Programming (ILP) and neural models which jointly does relevant and informative content selection, sentence compression for *noise pruning* and content ordering for *coherence*. We investigate for a framework as shown in Figure 1 which gives good cross-domain performance with minimal replacement of components.

2 Related Work

Text summarization can be achieved using extractive (Takamura and Okumura, 2009; Lin and Bilmes, 2011; Wang et al., 2008) and abstractive methods (Bing et al., 2015; Li, 2015). Extractive summarization has the advantage of output fluency due to direct use of human-written texts. However, because of the higher level of granularity exhibited by sentences, these approaches cannot ensure a noise free and coherent summary. A subset of previous extractive summarization approaches utilized parsed sentence structures to execute noise pruning while extracting content for summary (Morita et al., 2013; Berg-Kirkpatrick et al., 2011). But these techniques can merely prune noise, and cannot ensure a coherent reading for the summary reader. The attempt to achieve *coherence* in muti-document summarization was attempted by some of the extractive summarization system. (Christensen et al., 2013; Wang et al., 2016). However the attempts to achieve coherence in an MDS scenario often cope with intra- sentencial noise for coherence.

3 Method

Approach: We formulate an approach which ensures the construction of multi-document extractive summaries encompassing relevant and coherent content as depicted using Figures 1 and 2. The approach involves two steps, represented by blue rectangular boxes in the figure.

- Preprocessing: Identify named entities in the documents and parse the sentences in the input set of documents using a syntactic dependency parser. Construct a 'Corpus Graph' which provides means for tracking coherent and relevant content and for noise removal.
- Summary Extraction: Extract a sequence of noise- pruned sequence of syntactic subtrees which hold coherent, relevant and grammatically accurate information. The sequence of subtrees can be directly linearized to a coherent sequence of sentences.

The following subsections explain each one of these steps in detail.

3.1 Prepossessing and Corpus Graph Construction

At this stage, named entities in the input corpus sentences are identified and chunked [1]. Subsequently, the sentences in the input corpus is parsed into syntactic dependency trees [2]. The set of syntactic dependency trees are transformed into a corpus graph by adding extra nodes and edges (Figure 2). The graphs contains Sentence nodes (S_i) and Dummy start (S_{start}) and end nodes (S_{end}), Sentence Pair Edges (E_{ij}), Syntactic Trees Nodes N_{ij}, Named Entity Node N_{ij}, Entity Pair Edges ($EE_{ij,mn}$) represented by red colored circles, red colored arrows, ellipses drawn in black lines, rectangles drawn in blue colored lines and blue colored double lines respectively.

[1] https://stanfordnlp.github.io/CoreNLP/ner.html

[2] https://nlp.stanford.edu/software/lex-parser.shtml

3.2 Summary Extraction

Using Corpus Graph, we tranform summary extraction into a sub- graph extraction problem. The extracted subgraph SG should be of the form of the graph depicted at the right portion of the Figure 2. The extract should represent a sequence of sentence nodes with syntactic subtrees containing the most salient information attached to them. The syntactic substrees are formed by removing noisy portions of the original syntactic trees. The sequence should maximize coherence quantified by the selected set of sentence pair and entity pair edges. The total size of the text content held by the selected sequence of subtrees should be within the specified summary size. We formulate the sub- graph extraction from corpus graph as an integer linear programming(ILP) problem. Our ILP formulation is given below.

Maximize,

$$\begin{aligned} \mathrm{F(X)} = {} & \lambda_1 \sum_{S_i \epsilon SN} SSal(S_i) * sx_i + \\ & \lambda_2 \sum_{n_{ij} \epsilon TN} NSal(n_{ij}) * nx_{ij} + \\ & \lambda_3 \sum_{e_{ij,ik} \epsilon TE} ESal(e_{ij,ik}) * ey_{ij,ik} + \\ & \lambda_4 \sum_{e_{ij,ik} \epsilon EE} eSim(ne_{ij}, ne_{ik}) * sy_{ij,ik} + \\ & \lambda_5 \sum_{E_{ij} \epsilon SE} Prob(S_i, S_j) * Ey_{ij} - \\ & \lambda_6 \sum_{E_{ij} \epsilon SE} SSim(S_i, S_j) * Ey_{ij} \end{aligned} \quad (1)$$

Subject to Constraints,

$$\forall S_i \epsilon S, S_j \epsilon S$$

$$2 * (Ey_{ij} + Ey_{ji}) - (Sx_i + Sx_i) <= 1 (c_1)$$

$$\forall S_i, \sum_j Ey_{ij} = 1, \sum_j Ey_{ji} = 1 (c_2)$$

$$\sum_{S_i \epsilon S} Ey_{start,i} = 1, \sum_{S_i \epsilon S} Ey_{i,end} = 1 (c_3)$$

$$\sum_j Ey_{ij} - \sum_j Ey_{ji} = 0(c_4)$$

$$\sum_{S_i \epsilon S} Sx_i + \sum_{E_{ij} \epsilon SE} Ey_{ij} = 0(c_5)$$

$$\forall N_{ij} \epsilon TN, Sx_i - nx_{ij} >= 0(c_6)$$

$$\forall S_j \epsilon S, \sum_{N_{ij} \epsilon TN} nx_{ij} - Sx_i >= 0(c_7)$$

$$\forall e_{ij,ik} \epsilon TE$$

$$nx_{ij} + nx_{ij} - 2ey_{ij,ik} <= 1(c_8)$$

$$nx_{ij} + nx_{ij} - 2ey_{ij,ik} >= 0(c_9)$$

if DepReln($e_{ij,ik}$)$\epsilon GramRelns$

$$nx_{ij} - nx_{ik} = 0(c_{10})$$

$$\forall S_i \epsilon S,$$

$$sx_i - nx_{iroot} = 0(c_{11})$$

$$sx_i - (\sum_{N_{ij} \epsilon TN} nx_{ij} - \sum_{N_{ij} \epsilon TE} ey_{ij,ik}) = 0(c_{12})$$

$$\forall n_{ij} \epsilon TN, \sum_{e_{ij,ik} \epsilon TE} ey_{ij,ik} - nx_{ij}(c_{13})$$

$$\forall ee_{ij,mn} \epsilon EE,$$

$$2 * sy_{ij,mn} - (nx_{ij} + nx_{mn}) >= 0(c_{14})$$

$$2 * sy_{ij,mn} - (nx_{ij} + nx_{mn}) <= 1(c_{15})$$

$$(Ey_{im} + Ey_{mi}) - sy_{ij,mn} >= 0(c_{16})$$

$$\sum_{n_{ij} \epsilon TN} size(n_{ij}) * nx_{ij} <= SumSize(c_{17})$$

$$\sum_{S_i \epsilon S} sx_i <= N(c_{18})$$

Where,

X = $(.sx_i., .Ey_{ij}., .nx_{ij}., .ey_{ij,ik}., .sy_{ij,ik}.)$ represents a binary indicator vector corresponding to a candidate subgraph SG to be extracted. SN is the set of sentence nodes, SE is the set of sentence pair edges, TN is the set of syntactic tree nodes, TE is the sentence tree edges and EE is the set of entity pair edges. Indicator variables in SG represents different components of the graph as follows.

3.2.1 Linear Components of F:

$SSal$, $NSal$ and $ESal$ computes salience of sentence, node and edge respectively. $SSim$ computes the similarity between sentences while $ESim$ computes similarity between entities. $Prob$ returns the transition probablity of main verbs of parameter sentences, pre- computed using a large domain specific corpus. $SSim$ is used to penalize the redundant content in candidate summaries. $ESim$ and $Prob$ contributes for encouraging coherence of the summary to be extracted.

3.2.2 ILP Constraints:

The constraints c_1 to c_5 ensures the consistency between sentences nodes and sentence pair edges in SG. Also ensures that extracted subgraph contains a sequence of sentence nodes. c_6 to c_9 ensures the consistency of selection between sentence nodes, tree nodes and tree edges. $DepReln$ return the dependency relation corresponding to the parameter tree edge and $GramRelns$ contains the dependency relations required ensure grammaticality. The constraint c_{10} ensures that nodes which are essential for a syntactic subtree to hold grammatically accurate information won't be pruned away. nx_{iroot} indicates the root node of the syntactic tree corresponding to S_i and the constraint c_{11} ensures that subtrees extracted are rooted at the original root node. Constraints c_{12} to c_{16} ensures that entity pair edges are active only when corresponding named entity nodes are selected and when corresponding sentences are neighbours in the sequence of sentence nodes contained in SG.

4 Experiments and Results

4.1 Data

We evaluate our method using the test sets of DUC 2004 [3] and corpus MDS testset released by (Zopf et al., 2016) for law & Politics domain. We resort to standard ROUGE metric (Lin, 2004) for measuring content selection and rely on human evaluation for measuring coherence and linguistic quality. We tune our hyper parameters using the DUC 2003 dataset[4].

[3] http://duc.nist.gov/data.html
[4] http://duc.nist.gov/data.html

	D-2004(News)				Law			
System	R-1	R-2	R-L	R-W	R-1	R-2	R-L	R-W
Lin and Bilmes (2011)	39.3	10.7	38.7	15.7	41.4	9.7	40.75	21.90
Berg-Kirkpatrick et al. (2011)	36.3	8.3	37.3	13.7	40.14	9.12	39.91	21.70
Bing et al. (2015)	34.0	7.3	33.0	11.6	41.3	8.09	41.3	21.12
Christensen et al. (2013)	37.3	8.2	37.0	13.9	36.25	7.18	37.0	18.16
Wang et al. (2016)	39.0	9.3	37.3	13.7	39.30	8.75	38.25	19.30
Current System + G-Flow	38.3	9.8	38.0	13.6	40.7	9.33	40.91	21.37
Current System + BertSum	37.7	9.3	37.7	12.9	-	-	-	-

Table 1: Comparison with state of the art. In the Table R represents Rouge

	Coh			Inf			Gram		
Peer Systems(PS)	PS	OS	AMB	PS	OS	AMB	PS	OS	AMB
Lin and Bilmes (2011)	21	79	0	60	30	10	63	29	8
Berg-Kirkpatrick et al. (2011)	19	72	9	32	45	23	31	37	32
Bing et al. (2015)	20	67	13	34	60	6	37	40	23
Christensen et al. (2013)	43	54	3	18	60	12	70	17	13
Wang et al. (2016)	40	52	8	30	61	9	70	27	3
Kappa	73			77			72		

Table 2: Human Evaluation: In the Table, PS is Peer System, OS is Our System, Amb is Ambiguous, Coh is Coherence, Inf is Informativeness and Gram is Grammaticality

4.2 Settings

4.2.1 Salience Functions

- $SSal$: To compute sentence salience we use the same linear regression function proposed by Christensen et al. (2013). For news domain, we also leverage BertSum (Liu and Lapata, 2019) for computing sentence relevance.
- $NSal$: To compute syntactic tree node salience we use the neural model proposed by Kurisinkel et al. (2019) which leverage syntactic context information to compute the salience of a node.
- $ESal$:We set the weight of syntactic tree edge e as the freeuency of D_{bigram} which is the bigram constituted by the words incident on e.

4.2.2 Similarity Functions: $SSim$ & $ESim$

We relay on overlapping words for entity similarity and the similarity is computed using Jaccards Index(Hamers et al., 1989) entity word sets. For sentence similarity we rely on the method suggested by (Pawar and Mago, 2018). They compute the semantic similarity between sentences based on word similarity, sentence similarity and word order similarity.

4.2.3 Verb Transition Probability: $Prob$

We learn the fully connected neural network to learn verb transition probabilities. To learn the probabilities, we collect corpus of 16000 and 4500 documents in news and legal domains respectively. We extract main verbs from each sentence in the document using Standford Parser[5].

5 Evaluation

5.1 Evaluation of Content Coverage

We evaluated content coverage of the summary using objective metrics such as ROUGE. As show in the Table 1, results are reported in terms of ROUGE-1, ROUGE-2 and ROUGE -L. (Lin and Bilmes, 2011) consistently performed well in terms of ROUGE score. They incorporate a monotone sub-modular scoring function which is designed for quantitatively improving content coverage. Submodular maximization functions cannot be incorporated in an ILP setting with provision for improving coherence. Our approach yielded results that is comparable with (Lin and Bilmes, 2011) while out performing most of the other systems considered for evaluation. Other systems which incorporated coherence (Wang et al., 2016; Christensen et al., 2013) did not perform well in the evaluation for content coverage. We observe that these systems compromised on relevant content without any means for removing intra- sentencial noise. However, our approach for coherent summarization incorporates means for intra- sentential noise pruning performed well in terms of ROUGE evaluation.

[5]https://nlp.stanford.edu/software/lex-parser.shtml

System Summary
The new bills in question were the Bank Taxation Act, and the Act to Ensure the Publication of Accurate News and information. Alberta Press Act Reference-LRB- 1938-RRB-, concerned 4 Alberta statutes, one of which, the Accurate News and Information Act, would have compelled each newspaper in the province, to publish the government's rebuttal of criticism that had appeared in the newspaper. Coverage by the Edmonton Journal earned the newspaper a special Pulitzer Prize" for its editorial leadership in defense of the freedom of the press.
Original Sentences
The new bills in question were the Bank Taxation Act, an Act to Amend the Credit of Alberta Regulation Act and the Act to Ensure the Publication of Accurate News and information. Alberta Press Act Reference-LRB- 1938-RRB-, also called Reference Re Alberta Legislation, concerned 4 Alberta statutes, one of which, the Accurate News and Information Act, would have compelled each newspaper in the province, when called upon to do so by a government official, to publish the government's rebuttal of criticism that had appeared in the newspaper. Coverage by the Edmonton Journal was particularly strong and eventually earned the newspaper a special Pulitzer Prize" for its editorial leadership in defense of the freedom of the press.
Reference Summary
The Accurate News and Information Act was a statute passed by the Legislative Assembly of Alberta, Canada, in 1937, at the instigation of William Aberhart's Social Credit government. It would have required newspapers to print "clarifications" of stories that a committee of Social Credit legislators deemed inaccurate, and to reveal their sources on demand. The act was a result of the stormy relationship between Aberhart and the press, which dated to before the 1935 election, in which the Social Credit League was elected to government.

Table 3: Tree Combination vs Phrase Combination

5.2 Human Evaluation

We conducted human evaluation for evaluating summaries in other qualitative dimensions such as coherence, grammaticality and informativeness. Evaluators are four post graduate students in linguistics. During each evaluation process one among the peer systems compete with our system. 20 summaries generated by each one of the systems competing systems are chosen for evaluation. Summaries are shown to the evaluators in random order to avoid any kind of bias. For evaluating coherence and grammaticality, for each summary pair competing summaries, evaluators are asked to choose the best one in terms of the aspect under evaluation. For informativeness, they are asked to read the reference summary and asked to choose the most informative one. Results are shown in the Table 2. Our approach performed consistently better than other systems in the evaluation for coherence. We used verb transition probability in combination with entity similarity for modelling coherence. We observe this as the reason for our better performance in comparison with (Christensen et al., 2013) and (Wang et al., 2016). Our method performed comparably with other systems in the evaluation for informativeness. Obviously the methods which don't modify the original sentences performed better than our method in the evaluation for grammaticality. However we performed better than (Berg-Kirkpatrick et al., 2011). This shows that explicit use of neural model for computing node relevance using syntactic context and grammatical rules based on syntactic relations helped in maintaining grammaticality.

6 Discussions

A system summary generated for an input corpus in Law domain is shown in the Table in the next page. The table also contains the sequence of original sentences in the corpus with intra- sentencial noisy information and the corresponding reference summary. Clearly our method were successful in removing intra- sentencial noise and organizing summary for a coherent ordering. The neighboring sentences contained similar entities and the order of main verbs $((were, compelled, was))$ is the most likely one as per the transition probabilities computed using neural model for verb transition. The summary is also infomative as per the human written abstractive reference summary

Acknowledgements

This research was partially supported by the National Re- search Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme. We would also like to thank anonymous reviewers for several very insightful feedback on how to improve the paper.

References

Regina Barzilay and Mirella Lapata. 2008. Modeling local coherence: An entity-based approach. *Computational Linguistics*, 34(1):1–34.

Taylor Berg-Kirkpatrick, Dan Gillick, and Dan Klein. 2011. Jointly learning to extract and compress. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 481–490. Association for Computational Linguistics.

Lidong Bing, Piji Li, Yi Liao, Wai Lam, Weiwei Guo, and Rebecca J Passonneau. 2015. Abstractive multi-document summarization via phrase selection and merging. *arXiv preprint arXiv:1506.01597*.

Janara Christensen, Stephen Soderland, Oren Etzioni, 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.

Camille Guinaudeau and Michael Strube. 2013. Graph-based local coherence modeling. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 93–103.

Lieve Hamers et al. 1989. Similarity measures in scientometric research: The jaccard index versus salton's cosine formula. *Information Processing and Management*, 25(3):315–18.

Kevin Knight and Daniel Marcu. 2000. Statistics-based summarization-step one: Sentence compression. *AAAI/IAAI*, 2000:703–710.

Litton J Kurisinkel, Yue Zhang, and Vasudeva Varma. 2019. Domain adaptive neural sentence compression by tree cutting. In *European Conference on Information Retrieval*, pages 475–488. Springer.

Wei Li. 2015. Abstractive multi-document summarization with semantic information extraction. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1908–1913.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

Hui Lin and Jeff Bilmes. 2011. A class of submodular functions for document summarization. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 510–520. Association for Computational Linguistics.

Yang Liu and Mirella Lapata. 2019. Text summarization with pretrained encoders. *arXiv preprint arXiv:1908.08345*.

Annie Louis and Ani Nenkova. 2011. Text specificity and impact on quality of news summaries. In *Proceedings of the Workshop on Monolingual Text-To-Text Generation*, pages 34–42.

Hajime Morita, Ryohei Sasano, Hiroya Takamura, and Manabu Okumura. 2013. Subtree extractive summarization via submodular maximization. In *ACL (1)*, pages 1023–1032. Citeseer.

Atish Pawar and Vijay Mago. 2018. Calculating the similarity between words and sentences using a lexical database and corpus statistics. *arXiv preprint arXiv:1802.05667*.

Hiroya Takamura and Manabu Okumura. 2009. Text summarization model based on maximum coverage problem and its variant. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 781–789. Association for Computational Linguistics.

Jiwei Tan, Xiaojun Wan, and Jianguo Xiao. 2017. Abstractive document summarization with a graph-based attentional neural model. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1171–1181.

Dingding Wang, Tao Li, Shenghuo Zhu, and Chris Ding. 2008. Multi-document summarization via sentence-level semantic analysis and symmetric matrix factorization. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 307–314. ACM.

Xun Wang, Masaaki Nishino, Tsutomu Hirao, Katsuhito Sudoh, and Masaaki Nagata. 2016. Exploring text links for coherent multi-document summarization. pages 213–223.

Markus Zopf, Maxime Peyrard, and Judith Eckle-Kohler. 2016. The next step for multi-document summarization: A heterogeneous multi-genre corpus built with a novel construction approach. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1535–1545.