From Standard Summarization to New Tasks and Beyond: Summarization with Manifold Information

Shen Gao[1] , **Xiuying Chen**[1] , **Zhaochun Ren**[3] , **Dongyan Zhao**[1] and **Rui Yan**[1,2*]

[1]Wangxuan Institute of Computer Technology, Peking University, Beijing, China
[2]Beijing Academy of Artificial Intelligence
[3]School of Computer Science and Technology, Shandong University, Qingdao, China
{shengao,xy-chen,zhaody,ruiyan}@pku.edu.cn zhaochun.ren@sdu.edu.cn

Abstract

Text summarization is the research area aiming at creating a short and condensed version of the original document, which conveys the main idea of the document in a few words. This research topic has started to attract the attention of a large community of researchers, and it is nowadays counted as one of the most promising research areas. In general, text summarization algorithms aim at using a plain text document as input and then output a summary. However, in real-world applications, most of the data is not in a plain text format. Instead, there is much manifold information to be summarized, such as the summary for a web page based on a query in the search engine, extreme long document (*e.g.,* academic paper), dialog history and so on. In this paper, we focus on the survey of these new summarization tasks and approaches in the real-world application.

1 Introduction

The rapid growth of World Wide Web means that document floods spread throughout the Internet. Readers get drown in the sea of documents, wondering where to access. Text summarization system aims at generating a condensed version of a document and conveying the main idea to the reader. Users can save a lot of time by reading the summary instead of the whole document to capture the main idea. Hence, many websites and applications deploy automatic summarization systems, and researchers in Natural Language Processing (NLP) field also focus on the text summarization task.

Generally speaking, there are two types of text summarization. One is designed for the most common scenario that summarizes a plain text with hundreds of words, and the most popular usage is the news summarization. The other one, on the contrary, is designed for generating summary with manifold information in which input may be a structured document or document with additional knowledge. Different from the plain text summarization task, these *new summarization* tasks aim to produce a better and appropriate summary by incorporating manifold information in many real-world applications
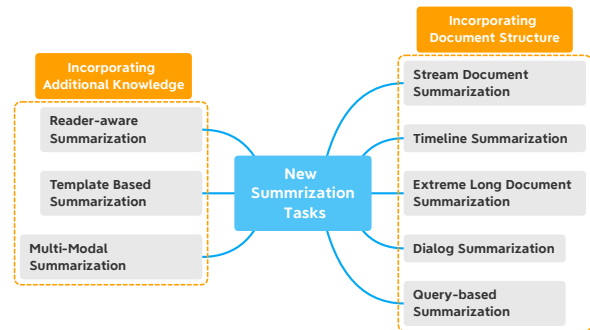
Figure 1: New summarization tasks introduced in this paper.

and scenarios. For example, for a search engine, it is better to summarize the web page according to the user's query instead of just summarizing the web page ignoring the query. Another example is dialog summarization, with the development of online chatting, people always chat with people on the web for business or chitchat (*e.g.,* on Slack or Whatsapp). Especially in the business scenario, it is helpful to give the user a summary of what has been talked in the past days before they starting a new dialog session, or give a brief introduction to the people who just join the group chat about what has been discussed in this group.

In contrast to the prosperity of survey on plain text summarization task, there are no systematic introductions to approaches about how to build an efficient summarization system which can leverage the manifold information (such as structure information of document and additional knowledge) in the research community. Thus, in this survey paper, we present a literature review for *new summarization tasks* and their corresponding methods, the tasks are listed in Figure 1.

Unlike the conventional plain text summarization methods which are built all with hand-crafted rules or feature engineering, recently many researchers begin to develop some data-driven approaches to build summarization systems. Since these approaches can leverage the publicly available large scale dataset and the rapid progress of deep learning approaches, instead of using time-consuming hand-crafted feature engineering. Therefore, we believe it is useful and valuable to summarize recent progress on new summarization tasks. This survey paper is partially based on our continu-

DUC	ROUGE-1	ROUGE-2	ROUGE-L
Extractive methods			
Lead-3	40.42	17.62	36.67
[Nallapati *et al.*, 2017]	39.60	16.20	35.30
[Narayan and et al., 2018]	40.00	18.20	36.60
[Wu and Hu, 2018]	41.25	18.87	37.75
[Zhang *et al.*, 2019]	42.37	19.95	38.83
[Liu and Lapata, 2019]	43.25	20.24	39.63
Abstractive methods			
[See *et al.*, 2017]	39.53	17.28	36.38
[Paulus and et al., 2018]	39.87	15.82	36.90
[Hsu and et al., 2018]	40.68	17.97	37.13
[Celikyilmaz and et al., 2018]	41.69	19.47	37.92
[Liu and Lapata, 2019]	42.13	19.60	39.18

Table 1: Leaderboard of document summarization task on CNN/-DailyMail dataset.

Figure 2: Techniques used in neural-based summarization methods.

ous efforts on building summarization models for new summarization tasks. We will introduce the problem formulation, data collection and the proposed methods for these tasks.[1]

2 Preliminary: Standard Summarization

In this section, we will introduce some generally used summarization frameworks based on conventional and neural-based learning methods. These frameworks are used as the basis of the methods for new summarization tasks.

2.1 Conventional Methods

Early conventional approaches to extractive summarization include: Centroid-based methods [Radev and et al., 2004], tree and graph based methods [Kikuchi and et al., 2014], Submodular methods [Morita and et al., 2013] and ILP-based methods [Gillick and et al., 2009]. Nevertheless, extractive approaches only extract some phrases or sentences from the original document as the summary, and it can not produce the condensed and fluent summary [Yates and et al., 2007]. On the contrary, the conventional abstractive summarization methods usually extract some words from document, and then reorder and perform linguistically-motivated transformations to the words [Dorr and et al., 2003]. However, these paraphrase-based generation method are easy to produce influent sentences.

2.2 Neural Methods

In contrast to the conventional learning methods, neural-based approaches provide an end-to-end method to summarization task. In this section, we introduce some widely used techniques in these methods, as shown in Figure 2, and we split the these methods into two categories: extractive and abstractive. Most of these works conduct the experiments on a benchmark dataset CNN/DailyMail, and we list the performance in terms of ROUGE score [Lin, 2004] at Table 1.

First, we will introduce the extractive methods which extract sentences from the document as the summary. Since the extractive methods use the sentence as the basic unit, the first
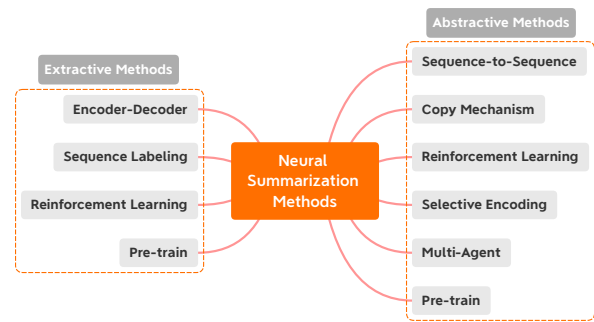
step is to obtain a sentence representation. The most common way [Cheng and Lapata, 2016] is to employ a recurrent neural network (RNN) or convolutional neural network (CNN) to encode the words in a sentence, and then obtain a vector representation. After obtaining the sentence representation, Cheng and Lapata [2016] first propose a framework with *encoder-decoder* using an RNN to tackle the extractive summarization task, which uses the encoder to obtain the vector representation of sentences and use the decoder with an attention mechanism to extract sentence. Since the *encoder-decoder* summarization framework [Narayan and et al., 2018] needs two RNN that computes slowly, many researchers start to use the *sequence labeling* framework [Nallapati *et al.*, 2017] for this task which use an RNN to read the sentences only once. To deeply understand the document, researchers incorporate the memory network [Chen *et al.*, 2018] into summarization framework which gives the reasoning ability to the model. Since previous works use the cross-entropy as the loss function to train the model which has a gap with the testing stage that use the ROUGE [Lin, 2004] score, [Wu and Hu, 2018] propose to use the *reinforcement learning* method to directly optimize the ROUGE score. In recent two years, the *pre-training* techniques growth rapidly in NLP field. Researchers employ the language model pre-training model (like Elmo, BERT) into summarization task [Liu and Lapata, 2019] to achieve better performance. From Table 1, we can find that [Liu and Lapata, 2019] achieve the state-of-the-art performance on the CNN/DailyMail dataset.

The *sequence-to-sequence* based text generation methods [Sutskever *et al.*, 2014; Li *et al.*, 2019; Gao *et al.*, 2019c] make generating fluent and concise summary possible, and Nallapati and et al. [2016] firstly apply the text generation method to the abstractive summarization task. Next, many extensions on the text generation framework are proposed to achieve better performance in generating summary. To avoid the out-of-vocabulary (OOV) problem, *copy mechanism* [Gu and et al., 2016; See *et al.*, 2017] is proposed which directly copy the OOV words (like the name of person, place or institution) from the source document into the generated summary. Similarly, researchers also use the *reinforcement learning* method in abstractive summarization [Paulus and et al., 2018; Wang and et al., 2018; Liu and et al., 2018a] for the same reason as the extractive methods. To help the summary generation module focus on the salient parts of source document, *selective encod-*

---

[1]Due to strict format limits, a longer version will be on the arXiv.

ing [Zhou and et al., 2017] are proposed to encode the important semantic parts and ignore the trivial parts. Encoding the source document is a crucial step in summarization, researchers [Celikyilmaz and et al., 2018] propose to divide the hard document encoding task into several sub-tasks and solve it by *multiple collaborating encoder agents*. After encoding the document, an attention mechanism is used to fuse all the document representations produced by all the agents, and then generate the summary. *Pre-training* language model helps the model to capture the semantic of text, that motivates the researchers to use the it as the encoder of summarization framework. Researchers employ the pre-training technique into document reading module [Liu and Lapata, 2019] to capture the main idea of the document, and achieve the state-of-the-art performance in abstractive summarization task. These methods use a plain text as input, and they can not utilize the document structure or other easily acquired knowledge to improve the summarization performance.

3 Summarization by Incorporating Document Structure

In this section, we will introduce some summarization tasks in which input is structural text instead of a plain text. These special structures will help the summarization model to capture the document main idea.

3.1 Stream Document Summarization

Stream summarization task was first introduced in TAC 2008 [Dang and Owczarzak, 2008], which targets at summarizing new documents in a continuously growing *text stream* such as news events and twitters. When the new document arrives in a sequence, the stream summary needs to be updated along with, considering previous information meantime.

Initial works include [Boudin and et al., 2008], which proposes a scalable sentence scoring method for query-oriented update summarization. In this method, candidate sentences are selected according to a combined criterion of query relevance and dissimilarity with previously read sentences. Delort and Alfonseca [2012] present an unsupervised probabilistic approach to model novelty in a document collection and apply it to the generation of update summaries. Ge and et al. [2016] propose a graph ranking based method, Burst Information Networks, as a novel representation of a text stream. In this method, the graph node is a burst word (including entities) with the time span of one of its burst periods, and an edge between two nodes indicates how strongly they are related. [Mnasri and et al., 2017] is the state-of-the-art work on the DUC and TAC dataset, which examines how integrating a semantic sentence similarity into an update summarization system can improve its results.

Current state-of-the-art methods for this task are all based on human-engineered and extractive methods. Nowadays, there are many stream data provided on the internet, such as tweets focus on the same news topic and news of a big event (like a presidential election, a natural disaster). Consequently, the abstractive-based summarization methods will be a hot research area and will be explored in the near future.

3.2 Timeline Summarization

Classic news summarization plays an important role with the exponential document growth on the Web. Many approaches are proposed to generate summaries but seldom simultaneously consider evolutionary characteristics of news plus to traditional summary elements. Timeline summarization is an important research task which can help users to have a quick understanding of the overall evolution of any given topic. It thus attracts much attention from research communities in recent years. To solve this task, one should first identify which sub-events are salient and then generate a summary. The big difference between timeline summarization and stream summarization task is whether the model can see all the sub-event.

Timeline summarization task is firstly proposed by Allan and et al. [2001], in this paper, they propose a method that extracts a single sentence from each event within a news topic. Later, a series of works [Yan *et al.*, 2011b; Yan *et al.*, 2011a; Yan *et al.*, 2012] further investigate the timeline summarization task, and all of them are based on conventional learning method to extract sentences from the timeline data. For instance, [Yan *et al.*, 2011b] formulate the timeline summarization task as a balanced optimization problem via iterative substitution. The objective function used in this method is measured by four properties: relevance, coverage, coherence, and diversity. In recent years, as an important case of timeline information, social media data is used by many timeline summarization research works. For example, Ren *et al.* [2013] considered the task of time-aware tweets summarization, based on a user's history and collaborative social influences from "social circles".

The previous works are all based on extractive methods, which are not as flexible as abstractive approaches. Chen *et al.* [2019] firstly propose a key-value memory network-based architecture to store the events described in the timeline. In this key-value memory network, they use the event time representation as the key, and split the value into two slots: global value and local value. The local value only captures event information from current event and the global value stores the global characteristics of events in different time position. Finally, an RNN-based decoder is employed to generate the summary abstractively. To train their model, they release the first large-scale timeline summarization dataset which contains 179,423 document-summary pairs collected from a cyclopedia website.

3.3 Extreme Long Document Summarization

Different from the previous summarization task, in some scenarios, the input document can be very long, such as an academic paper or a patent document which is longer than the news article. Thus, summarizing such an extreme long document is still a challenging problem when using existing summarization methods. The biggest challenge of this task is to extract the salient information and central idea from a large amount of information.

First, we introduce some benchmark datasets used the extreme long document summarization. In the era of using conventional machine learning methods, in this task, all of the researchers use the small-scale scientific papers as the

dataset [Teufel and Moens, 2002], and the number document summary data pairs is less than 100. In recent years, most of the researchers employ the neural-based methods which require a large amount of data to train the model. Thus, many large-scale long document datasets have been proposed and the data comes from Wikipedia [Liu and et al., 2018b], scientific papers [Cohan and et al., 2018], patent documents [Sharma *et al.*, 2019], etc. Liu and et al. [2018b] propose to use a Wikipedia web page with all the reference articles and the results fetched from Google as the long text input, and there are 2,332,000 document summary data pairs in this dataset. [Cohan and et al., 2018] propose a large-scale scientific paper summarization dataset which is collected from arXiv and PubMed, and it contains 348,000 document and summary pairs. The average document length is 4938 and 3016 words in arXiv and PubMed respectively, which is 6 times longer than the news dataset CNN/Daily-Mail. [Sharma *et al.*, 2019] propose a larger long document summarization dataset BigPatent, which is 10 times larger than the PubMed and arXiv. It contains 1,341,362 US patent document and summarization pairs and the average document length is 3,572 words. This dataset is more suitable for abstractive summarization task since the summary has more novel n-grams than other long document summarization datasets.

In the initial works of this task, researchers [Teufel and Moens, 2002] use a supervised classifier to select content from a scientific paper based on some human-engineered features. Then, researchers have extended these works by applying more sophisticated classifiers to identify more fine-grain categories. To determine whether a sentence should be included in the summary, [Collins *et al.*, 2017] directly use the section each sentence appears in as a categorical feature with values like Highlight, Abstract, Introduction, etc.

As for the neural network based methods, Liu and et al. [2018b] firstly use an extractive summarization method to coarsely identify salient information and then employ a neural abstractive model to generate the summary. [Cohan and et al., 2018] propose a hierarchical model that uses two-level RNN to encode the words and sections respectively, and then they use an attention decoder to forms a context vector from both word and sentence level information. They also conduct experiments on arXiv and PubMed datasets, and their model outperforms the baseline methods on these datasets. [Xiao and Carenini, 2019] propose an extractive method for this task using both the global context of the whole document and the local context within the current topic, and this method achieves state-of-the-art performance on the previous two datasets.

3.4 Dialog Summarization

In recent years, online chatting becomes more and more popular [Qiu *et al.*, 2019]. When the chatting history becomes very long, it is time-consuming for people to review all the context before starting a new dialog. Thus, some researchers focus on the task of summarizing the dialog history. Different from summarizing a document, the salient information is scattered in the whole dialog history.

[Ganesh and Dingliwal, 2019] first propose this task and

they propose a pipeline method that consists of a sequence labeling module to identify the salient utterance and a Seq2Seq module with attention and copy mechanism. Since their dataset is in a small-scale, they use a news summarization dataset CNN/DailyMail to train the abstractive summarization module and evaluate on a small scale dialog summarization dataset with only 45 sessions. To leverage the neural-based text generation method, [Gliwa and et al., 2019] propose the first large scale dataset SAMSum for this task. Different from previous papers working on chit-chats, [Liu and et al., 2019] propose a framework to generate a summary for online customer service, which can help the staff to know what was happen without going through long and sometimes twisted utterances.

As another branch of dialog summarization task, the meeting summarization task is to generate a summary of meeting transcriptions. [Shang and et al., 2018] propose an unsupervised abstractive meeting summarization using a graph-based model and budgeted submodular maximization. In recent years, people usually hold a meeting using video calls instead of just using audio. Consequently, additional visual information can be used in meeting summarization, such as the participant's head pose and eye gaze. [Li and et al., 2019] propose a multi-modal encoding framework that incorporates this visual information and employs a topic segmentation method to identify the topic transition in a dialog flow. Finally, they employ the Pointer-Generator [See *et al.*, 2017] network to fusion the encoded information and generate the summary.

3.5 Query-based Summarization

In the typical web search scenario, the search engine provides a list of web pages associated with their summaries. Different from the traditional document summarization, in this scenario, the summary should summarize the query focused aspect of the web page instead of the main idea. Inspire by this application, many researchers start to focus on the query-based summarization task, whose goal is to generate a summary that highlights those points that are relevant in the context of a given query.

In this task, most of the methods are based on conventional machine learning methods. [Li and Li, 2014] propose a semi-supervised graph-based model and incorporate the LDA topic model into summarization. [Feigenblat and et al., 2017] propose an unsupervised multi-document query-based summarization method using a cross-entropy method which is a generic Monte-Carlo framework for solving hard combinatorial optimization problems. Different from previous sentence extraction methods, [Wang and et al., 2013] employ a sentence compression method which uses three compression strategy: rule-based, sequence-based and tree-based to produce the summary.

To avoid generating repeated phrases and increasing the diversity of summary, [Nema and et al., 2017] firstly propose a neural-based Seq2Seq framework which ensures that context vectors in attention mechanism are orthogonal to each other. Specifically, to alleviate the problem of repeating phrases in the summary, we treat successive context vectors as a sequence and use a modified LSTM cell to compute the new state at each decoding time step. In decoding steps, the atten-

tion mechanism is also used to focus on different portions of the query at different time steps.

4 Summarization by Incorporating Additional Knowledge

In some summarization applications, there are many different types of additional knowledge that can be used to help the model enhance the performance. The model can leverage this additional knowledge to capture the main idea of the document or generate more fluent summaries.

4.1 Reader-aware Summarization

In most of the news websites, they provide an area for the readers to post their comments on the news article. In most cases, the reader comments concentrate on the main idea of the news article. Thus, these comments can be used to help the summarization model to capture the main idea of the news, and then the model can generate a better summary with this help. In this section, we will introduce two kinds of methods which are based on conventional learning methods and neural networks respectively.

In the beginning, [Hu and et al., 2008] firstly propose to understand the input document with the feedback of readers using a graph-based method, where they identify three relations (topic, quotation, and mention) by which comments can be linked to one another. [Li and et al., 2015] employ a sparse coding based framework for this task which jointly considers news documents and reader comments via an unsupervised data reconstruction strategy.

Next, we turn to the methods using neural networks. [Li and et al., 2017b] propose a sentence salience estimation framework based on a neural generative model called Variational Auto-Encoder (VAE). In contrast to the previous methods which use sentence extraction methods on a small-scale dataset, [Gao et al., 2019b] first propose a large-scale dataset and a neural generative method RASG on this task. This dataset contains 863,826 data samples, and each data sample has several a document, a summary and several reader comments (the average comments number of a document is 9.11). The proposed RASG method is a generative-adversarial based learning method which conducts the interaction between the reader comments and news article to capture the reader attention distribution on the article, and then use the reader focused article information to guide the summary generation process.

4.2 Template Based Summarization

To generate a fluent summary, template based summarization method first retrieves a summary template and then edits it into the new summary of the current document. Existing methods can be classified into two categories: hard-editing and soft-editing. More specifically, hard-editing methods force the system to generate the summary which is in the same language pattern as the template. Conversely, soft-editing methods can use partial words in the template and generate more flexible summaries.

[Wang and Cardie, 2013] introduced a template-based focused abstractive meeting summarization system. Their system first applies a MultipleSequence Alignment algorithm to generate templates and extracts all summary-worthy relation instances from the document. Then, the templates are filled with these relation instances. [Oya and et al., 2014] propose a hard-editing method that employs a multi-sentence fusion algorithm in order to generate summary templates.

Since the hard-editing methods are not very flexible, soft-editing methods become popular in recent years due to the development of the neural text generation method. [Cao et al., 2018] employ existing summaries as soft templates to generate a new summary. In this method, they use an information retrieval system to retrieve summaries of a similar document and then use an attention-based generator to fuse the information from the template and current document. [Wang et al., 2019] leverages template discovered from training data to softly select key information from each source article to guide its summarization process. However, this method ignores the dependency between the template document and the input document. Following these works, [Gao et al., 2019a] propose to analyze the dependency and use this dependency to help the model identify which facts in the input document are the salient facts that should be mentioned in the summary. Furthermore, they use the relationship between template document and template summary to extract the summary template that can be reused in generating a new summary. In [Gao et al., 2019a], they also propose a large-scale dataset (contains 2,003,390 document and summary pairs) in which summaries are all in patternized language, and their method achieves state-of-the-art performance on this dataset.

4.3 Multi-Modal Summarization

With the increase of multi-media data on the web, many researchers focus on the multi-modal summarization task in recent years. Compared to the traditional text summarization task setting, in the multi-modal summarization task, the visual information is incorporated along with the input document into the text summarizing process to improve the quality of the generated summary.

In the beginning, we first introduce some datasets of image-based multi-modal summarization. [Li et al., 2018; Chen and Zhuge, 2018] propose two large-scale multi-modal summarization datasets, and each data sample in these datasets contains a source sentence, an image collected from the webpage and a summary. Different from the previous two datasets which output is only text, [Zhu et al., 2018] propose the first large-scale multi-modal input and multi-modal output summarization dataset which input is a document with several images and the ground truth is a text summary with the most relevant image selected from the input images. [Li and et al., 2017a; Palaskar and et al., 2019] propose two video-based multi-modal summarization dataset which contains document, video and summary pairs, and the number of data sample are 1000 and 79114 respectively.

Next, we will introduce some existing methods of multi-modal summarization task. For the image-based multi-modal summarization task, [Li et al., 2018; Chen and Zhuge, 2018] propose to use a Seq2Seq based abstractive model which has image and document encoders to obtain the representations of multi-modal input and an RNN based decoder with multi-modality attention to generate the summary. Since there are

some abstract concepts in the source document which can not find a counterpart in the image, and not all the visual information is useful for generating summary. To avoid introducing noise into summarization, [Li *et al.*, 2018] propose to use an image attention filter and an image context filter. As for the video-based summarization, [Palaskar and et al., 2019] employ a ResNeXt-101 3D convolutional neural network to model the video frames and then fuse this video information into the Seq2Seq using a hierarchical attention mechanism.

5 Conclusion

We have witnessed a rapid surge of summarization studies recently, especially the research in the many new summarization tasks. Summarization systems are catching on fire: the state-of-the-art performance has been pushed higher and higher. In the real-world applications, most of them are not in a traditional summarization setting, and they usually leverage manifold information. Since large-scale data become more easily available in current era, and it requires much time for people to obtain the overall information for an event. We may stand at the entrance of future success in more advanced summarization systems. It is our hope that this survey provides an overview of the challenges and the recent progress as well as some future directions in these new summarization tasks which leverage the manifold information.

Acknowledgments

References

[Allan and et al., 2001] James Allan and et al. Temporal summaries of new topics. In *SIGIR*, pages 10–18. ACM, 2001.

[Boudin and et al., 2008] Florian Boudin and et al. A scalable mmr approach to sentence scoring for multi-document update summarization. In *COLING*, 2008.

[Cao *et al.*, 2018] Ziqiang Cao, Wenjie Li, Sujian Li, and Furu Wei. Retrieve, rerank and rewrite: Soft template based neural summarization. In *ACL*, 2018.

[Celikyilmaz and et al., 2018] Asli Celikyilmaz and et al. Deep communicating agents for abstractive summarization. In *NAACL*, 2018.

[Chen and Zhuge, 2018] Jingqiang Chen and Hai Zhuge. Abstractive text-image summarization using multi-modal attentional hierarchical rnn. In *EMNLP*, 2018.

[Chen *et al.*, 2018] Xiuying Chen, Shen Gao, Chongyang Tao, Yan Song, Dongyan Zhao, and Rui Yan. Iterative document representation learning towards summarization with polishing. In *EMNLP*, 2018.

[Chen *et al.*, 2019] Xiuying Chen, Zhangming Chan, Shen Gao, Meng-Hsuan Yu, Dongyan Zhao, and Rui Yan. Learning towards abstractive timeline summarization. In *IJCAI*, 2019.

[Cheng and Lapata, 2016] Jianpeng Cheng and Mirella Lapata. Neural summarization by extracting sentences and words. In *ACL*, pages 484–494, 2016.

[Cohan and et al., 2018] Arman Cohan and et al. A discourse-aware attention model for abstractive summarization of long documents. In *NAACL-HLT*, 2018.

[Collins *et al.*, 2017] Ed Collins, Isabelle Augenstein, and Sebastian Riedel. A supervised approach to extractive summarisation of scientific papers. In *CoNLL*, pages 195–205, August 2017.

[Dang and Owczarzak, 2008] Hoa Trang Dang and Karolina Owczarzak. Overview of the tac 2008 update summarization task. In *TAC*, 2008.

[Delort and Alfonseca, 2012] Jean-Yves Delort and Enrique Alfonseca. Dualsum: a topic-model based approach for update summarization. In *EACL*, 2012.

[Dorr and et al., 2003] Bonnie J. Dorr and et al. Hedge trimmer: A parse-and-trim approach to headline generation. In *HLT-NAACL*, 2003.

[Feigenblat and et al., 2017] Guy Feigenblat and et al. Unsupervised query-focused multi-document summarization using the cross entropy method. In *SIGIR*, 2017.

[Ganesh and Dingliwal, 2019] Prakhar Ganesh and Saket Dingliwal. Abstractive summarization of spoken and written conversation. *ArXiv*, abs/1902.01615, 2019.

[Gao *et al.*, 2019a] Shen Gao, Xiuying Chen, Piji Li, Zhangming Chan, Dongyan Zhao, and Rui Yan. How to write summaries with patterns? learning towards abstractive summarization through prototype editing. In *EMNLP*, 2019.

[Gao *et al.*, 2019b] Shen Gao, Xiuying Chen, Piji Li, Zhaochun Ren, Lidong Bing, Dongyan Zhao, and Rui Yan. Abstractive text summarization by incorporating reader comments. In *AAAI*, 2019.

[Gao *et al.*, 2019c] Shen Gao, Zhaochun Ren, Yihong Zhao, Dongyan Zhao, Dawei Yin, and Rui Yan. Product-aware answer generation in e-commerce question-answering. In *WSDM*, 2019.

[Ge and et al., 2016] Tao Ge and et al. News stream summarization using burst information networks. In *EMNLP*, 2016.

[Gillick and et al., 2009] Dan Gillick and et al. A scalable global model for summarization. In *ILP 2009*, 2009.

[Gliwa and et al., 2019] Bogdan Gliwa and et al. Samsum corpus: A human-annotated dialogue dataset for abstractive summarization. In *EMNLP*, 2019.

[Gu and et al., 2016] Jiatao Gu and et al. Incorporating copying mechanism in sequence-to-sequence learning. *Acl*, page 11, 2016.

[Hsu and et al., 2018] Wan-Ting Hsu and et al. A unified model for extractive and abstractive summarization using inconsistency loss. In *ACL*, 2018.

[Hu and et al., 2008] Meishan Hu and et al. Comments-oriented document summarization: understanding documents with readers' feedback. In *SIGIR*, 2008.

[Kikuchi and et al., 2014] Yuta Kikuchi and et al. Single document summarization based on nested tree structure. In *ACL*, 2014.

[Li and et al., 2015] Piji Li and et al. Reader-aware multi-document summarization via sparse coding. In *IJCAI*, 2015.

[Li and et al., 2017a] Haoran Li and et al. Multi-modal summarization for asynchronous collection of text, image, audio and video. In *EMNLP*, 2017.

[Li and et al., 2017b] Piji Li and et al. Reader-aware multi-document summarization: An enhanced model and the first dataset. In *NFiS@EMNLP*, 2017.

[Li and et al., 2019] Manling Li and et al. Keep meeting summaries on topic: Abstractive multi-modal meeting summarization. In *ACL*, 2019.

[Li and Li, 2014] Yanran Li and Sujian Li. Query-focused multi-document summarization: Combining a topic model with graph-based semi-supervised learning. In *COLING*, pages 1197–1207, August 2014.

[Li *et al.*, 2018] Haoran Li, Junnan Zhu, Tianshang Liu, Jiajun Zhang, and Chengqing Zong. Multi-modal sentence summarization with modality attention and image filtering. In *IJCAI*, 2018.

[Li *et al.*, 2019] Juntao Li, Lidong Bing, Lisong Qiu, Dongmin Chen, Dongyan Zhao, and Rui Yan. Learning to write stories with thematic consistency and wording novelty. In *AAAI*, 2019.

[Lin, 2004] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, July 2004.

[Liu and et al., 2018a] Linqing Liu and et al. Generative adversarial network for abstractive text summarization. In *AAAI*, 2018.

[Liu and et al., 2018b] Peter J. Liu and et al. Generating wikipedia by summarizing long sequences. In *ICLR*, 2018.

[Liu and et al., 2019] Chunyi Liu and et al. Automatic dialogue summary generation for customer service. In *KDD*, 2019.

[Liu and Lapata, 2019] Yang Liu and Mirella Lapata. Text summarization with pretrained encoders. In *EMNLP*, 2019.

[Mnasri and et al., 2017] Maali Mnasri and et al. Taking into account inter-sentence similarity for update summarization. In *IJC-NLP*, 2017.

[Morita and et al., 2013] Hajime Morita and et al. Subtree extractive summarization via submodular maximization. In *ACL*, 2013.

[Nallapati and et al., 2016] Ramesh Nallapati and et al. Abstractive text summarization using sequence-to-sequence rnns and beyond. In *CoNLL*, 2016.

[Nallapati *et al.*, 2017] Ramesh Nallapati, Feifei Zhai, and Bowen Zhou. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. In *AAAI*, 2017.

[Narayan and et al., 2018] Shashi Narayan and et al. Ranking sentences for extractive summarization with reinforcement learning. In *NAACL*, 2018.

[Nema and et al., 2017] Preksha Nema and et al. Diversity driven attention model for query-based abstractive summarization. In *ACL*, 2017.

[Oya and et al., 2014] Tatsuro Oya and et al. A template-based abstractive meeting summarization: Leveraging summary and source text relationships. In *INLG*, 2014.

[Palaskar and et al., 2019] Shruti Palaskar and et al. Multimodal abstractive summarization for how2 videos. In *ACL*, 2019.

[Paulus and et al., 2018] Romain Paulus and et al. A deep reinforced model for abstractive summarization. In *ICLR*, 2018.

[Qiu *et al.*, 2019] Lisong Qiu, Juntao Li, Wei Bi, Dongyan Zhao, and Rui Yan. Are training samples correlated? learning to generate dialogue responses with multiple references. In *ACL*, 2019.

[Radev and et al., 2004] Dragomir R. Radev and et al. Centroid-based summarization of multiple documents. volume 40, pages 919–938, 2004.

[Ren *et al.*, 2013] Zhaochun Ren, Shangsong Liang, Edgar Meij, and Maarten de Rijke. Personalized time-aware tweets summarization. In *SIGIR*, pages 513–522. ACM, 2013.

[See *et al.*, 2017] Abigail See, Peter J. Liu, and Christopher D. Manning. Get to the point: Summarization with pointer-generator networks. In *ACL*, 2017.

[Shang and et al., 2018] Guokan Shang and et al. Unsupervised abstractive meeting summarization with multi-sentence compression and budgeted submodular maximization. In *ACL*, 2018.

[Sharma *et al.*, 2019] Eva Sharma, Chen Li, and Lu Wang. Big-patent: A large-scale dataset for abstractive and coherent summarization. In *ACL*, 2019.

[Sutskever *et al.*, 2014] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. Sequence to sequence learning with neural networks. In *NIPS*, 2014.

[Teufel and Moens, 2002] Simone Teufel and Marc Moens. Summarizing scientific articles: Experiments with relevance and rhetorical status. *Computational Linguistics*, 28:409–445, 2002.

[Wang and Cardie, 2013] Lu Wang and Claire Cardie. Domain-independent abstract generation for focused meeting summarization. In *ACL*, pages 1395–1405, August 2013.

[Wang and et al., 2013] Lu Wang and et al. A sentence compression based framework to query-focused multi-document summarization. In *ACL*, pages 1384–1394, August 2013.

[Wang and et al., 2018] Li Wang and et al. A reinforced topic-aware convolutional sequence-to-sequence model for abstractive text summarization. In *IJCAI*, 2018.

[Wang *et al.*, 2019] Kai Wang, Xiaojun Quan, and Rui Wang. Biset: Bi-directional selective encoding with template for abstractive summarization. In *ACL*, pages 2153–2162, July 2019.

[Wu and Hu, 2018] Yuxiang Wu and Baotian Hu. Learning to extract coherent summary via deep reinforcement learning. In *AAAI*, 2018.

[Xiao and Carenini, 2019] Wen Xiao and Giuseppe Carenini. Extractive summarization of long documents by combining global and local context. In *EMNLP*, 2019.

[Yan *et al.*, 2011a] Rui Yan, Liang Kong, Congrui Huang, Xiaojun Wan, Xiaoming Li, and Yan Zhang. Timeline generation through evolutionary trans-temporal summarization. In *EMNLP*, pages 433–443. ACM, 2011.

[Yan *et al.*, 2011b] Rui Yan, Xiaojun Wan, Jahna Otterbacher, Liang Kong, Xiaoming Li, and Yan Zhang. Evolutionary timeline summarization: a balanced optimization framework via iterative substitution. In *SIGIR*, pages 745–754. ACM, 2011.

[Yan *et al.*, 2012] Rui Yan, Xiaojun Wan, Mirella Lapata, Wayne Xin Zhao, Pu-Jen Cheng, and Xiaoming Li. Visualizing timelines: Evolutionary summarization via iterative reinforcement between text and image streams. In *CIKM*, pages 275–284. ACM, 2012.

[Yates and et al., 2007] Alexander Yates and et al. Textrunner: Open information extraction on the web. In *HLT-NAACL*, 2007.

[Zhang *et al.*, 2019] Xingxing Zhang, Furu Wei, and Ming Zhou. Hibert: Document level pre-training of hierarchical bidirectional transformers for document summarization. In *ACL*, 2019.

[Zhou and et al., 2017] Qingyu Zhou and et al. Selective encoding for abstractive sentence summarization. In *ACL*, 2017.

[Zhu *et al.*, 2018] Junnan Zhu, Haoran Li, Tianshang Liu, Yu Zhou, Jiajun Zhang, and Chengqing Zong. Msmo: Multimodal summarization with multimodal output. In *EMNLP*, 2018.