Query-based Text Summarization: Techniques and Classification

Varad Unhale
Department of Electronics and Telecommunication Engg.,
SCTR's Pune Institute of Computer Technology,
Pune, India
varadunhale.vu@gmail.com

Suyash Udchan
Department of Electronics and Telecommunication Engg.,
SCTR's Pune Institute of Computer Technology,
Pune, India
suyash.udchan@gmail.com

Mohak Shah
Department of Electronics and Telecommunication Engg.,
SCTR's Pune Institute of Computer Technology,
Pune, India
shahmohak2311@gmail.com

Bhakti Kadam
Department of Electronics and Telecommunication Engg.,
SCTR's Pune Institute of Computer Technology,
Pune, India
bdkadam@pict.edu

***Abstract*—Owing to the proliferation of internet and multimedia applications, there has been a huge increase in digital content in the form of text, images, audio and video worldwide. The various applications of text summarization has attracted the computer vision researchers to generate the optimal text summaries. Several methodologies have been reported in literature to generate text summaries. However, the key challenge is incorporating user's preference as text summarization is a subjective task. User generated queries act as a guiding beacon in the summarization process. Instead of producing generic text summaries, query-based text summarization techniques offer user preferred responses. This survey begins by clarifying the fundamental concepts and objectives of text summarization, emphasizing on the role of user queries in the summarization process. It proceeds to categorize existing query-based summarization techniques into distinct paradigms, including extractive, abstractive, and hybrid approaches, highlighting the advantages and limitations of each. This paper presents the evolution of query-based text summarization with developed techniques, available datasets, evaluation metrics, and performance comparison. The paper dives deeper into the paradigms of query-based text summarization and also provides the directions for prospective future research in the domain.**

***Index Terms*—Query-based Text Summarization, Extractive Summarization, Abstractive Summarization, Deep Learning, Evaluation Metrics**

I. Introduction

Text summarization is an essential task in a world where information is abundant but time and attention are limited. The ability to distill the most crucial information from lengthy documents or vast datasets enables users to quickly access the information that matters most to them. However, generic summarization methods may fall short of addressing the specific information needs of users. This is where query-based text summarization comes into play. Query-based text summarization, also known as topic-based, user-focused, or query-focused summarization, integrates user-provided query information to craft the text summaries. In contrast to generic text summarization, which aims to provide broad document summaries, query-based summarization extracts or generates text that covers query-related key points and condenses the documents. In essence, text summarization can be seen as a higher-level category comprising three forms: generic summarization, extractive query-based summarization, and abstractive query-based summarization. It often results in a summary of content related to the user's question.

There have been surveys, like the one conducted by Afantenos et al. [17], that addressed text summarization in the context of medical documents. Nevertheless, these surveys limited their focus to medical-related queries. In 2010, Damova et al. [35] conducted a survey on query-based text summarization, but the field has evolved significantly in the following decade, with the introduction of technologies such as Word2Vec, end-to-end sequence generation models, and various attention mechanisms. These advances have contributed to the progress in query-based text summarization techniques. The challenge lies in developing an efficient and effective query-based text summarization method that can distill essential information from extensive textual datasets in response to user-generated queries, considering the diversity of summarization paradigms and their unique advantages and limitations. Therefore, this survey aims to provide a comprehensive overview of the specific techniques in query-based text summarization and bridge the gap in the literature, considering the recent developments in this field.

This paper presents a survey of generic and query-based text summarization techniques including existing techniques, performance comparison, limitations and challenges in the same.

II. Problem Statement

Given an input text, the aim of query-based text summarization model is to generate a meaningful text summary focusing on user inputted query. Figure 1 shows the systematic block diagram of query-based text summarization. Text preprocessing, Text extraction, Query pre-processing, training of the deep learning model and generating a text summary are the major steps in query-focused text summarization. From the perspective of text summarization, query-based text summarization can be categorized into two main approaches: query-based extractive and query-based abstractive text summarization. This type of summarization takes into account user-provided queries, which can be in the form of individual words or complete sentences. Additionally, research in this area can be divided into Single-Document Summarization (SDS) and Multi-Document Summarization (MDS). Figure 2 shows the broad categorization of Text Summarization techniques.

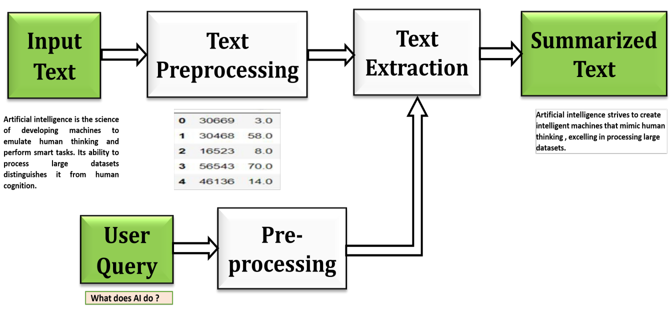

Fig. 1. Block Diagram of Query-based Text Summarization

The survey's structure overview begins by examining a key division: extractive techniques and abstractive techniques in text summarization. These two approaches differ fundamentally, with extractive techniques focusing on selecting prominent sections from a document and directly extracting them. Consequently, query-based extractive summarization primarily evaluates the relevance of the content concerning the query. In contrast, abstractive text summarization aims to identify the essential information in the text and generate new text that encapsulates the identified information, which is not present in the original document. While most research in text summarization revolves around either extractive or abstractive techniques, there are also efforts to combine these two approaches using hybrid methods.

This survey also explores the key findings in these areas following the discussions on extractive and abstractive summarization. In both the query-based extractive and abstractive text summarization sections, we primarily present the relevant research within two fundamental categories of machine learning techniques: unsupervised learning and supervised learning. By synthesizing various papers within each category, the common characteristics are extracted to summarize the core concepts of each category and offer insights for future research.

III. Existing Techniques

One of the most important aspects of developing an effective deep learning or machine learning algorithm is to have a validated evaluation that ensures the generalization of the developed algorithm or method. Specifically, evaluation is composed of two parts: evaluation datasets and evaluation metrics. Different machine learning tasks require different evaluation datasets and metrics to be chosen for the correct generalization purpose. So, before introducing summarization techniques, one important question to be answered for query-based text summarization is: how should the summaries predicted by the model be evaluated? This section briefly discusses the existing approaches reported in literature for generic and query-based text summarization. Table I presents a comparison of existing techniques in terms of methodology used, datasets experimented, achieved results and limitations.

In the paper authored by Aakash Sinha et al. [10], feed-forward neural networks are utilized in conjunction with Word2Vec and Fasttext for word and sentence vectorization, respectively, all while employing the softmax activation function. Their work is evaluated using the DUC 2002 dataset and ROUGE metrics, demonstrating that optimal performance is achieved when processing 40 sentences at a time. The paper authored by Ángel Herandez-Castenda et al. [14] employs clustering, Genetic Algorithm, LDA, TF-IDF, and N-grams techniques. They have evaluated their method using the DUC02 and TAC11 datasets, achieving an impressive 86% accuracy rate for their generated summaries and suggesting its potential.

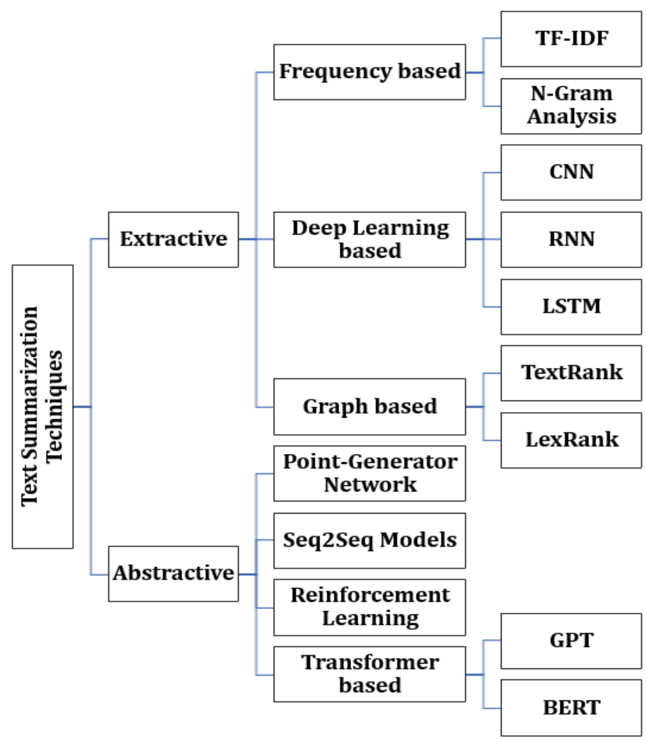

Fig. 2. Categorization of Text Summarization techniques

The research work presented in [1] utilizes cosine similarity, word-order similarity, semantic similarity, a hybrid similarity measure, and clustering algorithms for query-based summarization. Their research indicates that setting the 'page-len' parameter to 65, processing 65 sentences at a time, yields the best ROUGE scores.

TABLE I: Comparative study of Video Activity Classification techniques

Research Work	Methodology	Dataset	Results	Limitations
[10]	Feedforward neural networks, Word2Vec for word vectorization Fasttext for sentence, vectorization, Softmax activation function.	DUC 2002 Metrics: ROUGE-1 and ROUGE-2.	The best performance in terms of ROUGE scores was observed when using a 'page-len' parameter set to 40, indicating that processing 40 sentences at a time produced the best results.	Enhancing the model's capability to generate summaries exceeding the 'page-len', thus expanding its practicality across a wider range of summarization tasks.
[27]	BERT, a bidirectional Transformer model, boosts NLP tasks through pre-training and efficient fine-tuning, achieving top results.	BooksCorpus and English Wikipedia.	Unsupervised pre-training improves language understanding, benefits low-resource tasks, and extends to versatile use in deep bidirectional architectures.	Current language models are unidirectional, limiting context and pre-training options, impacting their performance and highlighting the need for more versatile models.
[1]	Cosine similarity, word-order similarity,semantic similarity,a hybrid similarity measure, and clustering algorithms (DBSCAN and Agglomerative clustering).	Various online websites of Amrita School of Engineering	These results suggest that setting the 'page-len' parameter to 65, which means processing 65 sentences at a time, led to the best performance in terms of ROUGE scores for the query- based text summarization method.	Summarization methods can excel or struggle in specific domains due to specialized terminology.
[4]	DGS- Summarizer, Q-Summarizer & QInc-Summarizer.	DUC02 dataset.	In particular, the Q-Summarizer method was the most effective in generating query-based summaries.	Further experiments on larger text corpora are needed to evaluate the performance of our method.
[2]	Tokenization, Normalization and Stop-Word Removal, Part-of-Speech Tagging Lemmatization Context Modeling using Latent Semantic Indexing Term-document matrix generation Retrieving query specific information Text Rank Algorithm.	TD-OFS, DUC-2005, publicly available news and email datasets.	The framework uses the LSA technique to build an intuitive semantic structure in addition to an enhanced PageRank algorithm, that aims at reducing redundant data without the use of any supervised learning technique.	To develop a semi supervised model that takes into account the sentences required by the user, in order to increase the semantic efficiency of the summary while including another level of feature set of the data.
[3]	Unsupervised Learning, Methods Approaches Based on Document Graphs	Publicly available news and email Dataset and transcripts.	The study outcomes revealed that when the 'page-len' parameter was adjusted to 25, signifying the concurrent processing of 25 sentences, it yielded the most promising performance in relation to ROUGE scores for the query- based text summarization approach.	An existing methods may face challenges when handling very short queries, lacking the necessary context for informative summarization.

[14]	The paper introduces a text summarization method utilizing Doc2Vec, LDA, and a genetic algorithm to cluster sentences based on semantic and lexical features.	DUC02 dataset.	Excellent summarization results, outperforming prior methods on DUC02, displaying competitiveness on TAC11.	The paper lacks explicit limitations but potential concerns include scalability and GA parameter tuning.
[5]	Term Sense, Extraction Specificity Power,Informativeness Power, Sentence Selection, Maximal Marginal Relevance (MMR), Evaluation Metrics , Normalization.	News corpus built by the Monitor and Research Center National Language Resource.	Optimal summarization results were achieved with a compression rate of 20 for documents with many closely related sub-topics.	Further experiments on larger text corpora are needed to evaluate the performance of our method.
[15]	Fuzzy C-Means clustering.	News articles from CNN, Improved F-Measure.	Improved F-Measure.	Additional tests on more extensive text datasets are required to assess how well our approach performs.
[14]	Comparison of two word sequence models (n-grams and MFS) using unsupervised learning (K-means) for automatic extractive text summarization. Different term weights (BOOL, TF, IDF, TFIDF) were also tested.	Standard DUC 2002.	1.3-grams with IDF weighting achieved the best Recall, MFS with BOOL weighting came in second for Recall. 3.1-grams with BOOL weighting were third in terms of Recall.	Limited to extractive summarization. Dependency on predefined parameters like the number of clusters in K-means, The choice of n-gram size for summarization is not clear.
[30]	A text summarization model is built using a sequence-to-sequence architecture with attention and a pointer mechanism. It takes a document and a query as input, utilizing bidirectional RNN encoders. The model employs attention to highlight important document parts and a generator mechanism to generate summaries.	The dataset used in the paper is created from CNN and Daily Mail news articles and contains document-query-answer triples.	The experiment results show that the model performs better when queries are incorporated, as indicated by ROUGE scores. The model benefits from the information provided by queries. Despite its performance, the model scores lower than a strong baseline, primarily due to the dataset's nature.	The model's performance relies on queries, posing a limitation when they are absent or less informative. Its lower performance compared to a strong baseline, influenced by dataset characteristics, indicates potential variability in effectiveness across different text corpora or domains.
[24]	MeanSum is a model for summarizing multiple documents without human-labeled data. It tests different architectures and pre-trained models to enhance summarization quality.	The document mentions using Yelp dataset.	The MeanSum model performs well in evaluations, surpassing extractive models in ROUGE scores and sentiment accuracy. It maintains word overlap, which correlates with ROUGE metrics.	The unsupervised abstractive model is designed for multi-document summarization and doesn't work well for single-document summarization due to the lack of redundancy cues.
[18]	Extractive text summarization using a statistical novel approach based on sentence ranking.	DUC 2002.	The proposed model improves the accuracy compared to traditional approaches, but specific quantitative results are not provided in the document.	The need for labeled data, anaphora and cataphora problems, and the ongoing.

[16]	The document introduces a neural network model for generating concise summaries from input sentences, employing a combination of neural language models and different encoders for context capture. The model is trained using negative log-likelihood on input-summary pairs, and beam search decoding is utilized for efficient summary generation.	DUC-2003 and DUC-2004, The training data is sourced from the Gigaword dataset, with training pairs formed by pairing article headlines with their first sentences, and several baseline methods are compared for benchmarking.	The study conducted experiments on summarization models using DUC-2004 and Gigaword datasets, highlighting the importance of combining article information and language models for better performance. While ABS and MOSES+ models outperformed TOPIARY, they fell short of human performance, indicating room for further improvement in abstractive summarization.	Pretrained language models like BERT can inherit biases from the internet data they are trained on, which can result in biased or objectionable outputs. Even fine-tuning the model on specific data may not fully eliminate these biases. Users should be cautious about potential biases when employing these models in production.
[17]	The paper discusses various methods and approaches to text summarization, including supervised, unsupervised, and hybrid techniques.	CNN/Daily Mail, Gigaword, NYT, DUC, 20NG, TIDSUMM, TT News, SummMac and metrics for text summarization.	The paper reviews various research works and their results, covering different algorithms and techniques for text summarization.	The paper mentions some common challenges in text summarization, such as the evaluation of summaries, the need for labeled data, anaphora and cataphora problems, and the ongoing research to find the perfect model for generating human-like summaries.
[13]	Word Vector Embedding approach for Extractive Text Summarization and Neural Network for Extractive Summarization using Supervised Learning method.	DUC2002 dataset.	The paper discusses various summarization techniques and their effectiveness. Results are mentioned in the context of different techniques and methods.	Not explicitly mentioned, but it highlights areas for potential improvements, such as dataset size and theme diversity, suggesting that results could be enhanced with more data and advanced techniques.
[12]	The document explores methods for summarizing multiple documents, including graph-based and cluster-based approaches. These methods involve techniques like TextRank, graph algorithms, and clustering to identify important information.	ROUGE-N, ROUGE-L, ROUGE-W, ROUGE-S, and ROUGE-SU.	It highlights the methodologies and approaches used in various summarization techniques.	The paper suggests challenges in multi-document summarization, such as the need for efficient algorithms, the potential loss of information, and the difficulty of handling domain-specific knowledge.
[35]	Extractive text summarization, feature extraction, neural network, NLTK toolkit, Porter stemmer algorithm, ffnet library.	Wikipedia articles as input, F1 score for evaluation.	The system's performance varied based on different features. The best results were obtained when considering citations (feature f9), achieving an F1 score of 0.223.	The paper doesn't explicitly state limitations, but it's crucial to recognize that summarization performance may vary with different datasets and domains. The effectiveness of the system in real-world scenarios is not addressed.

The authors Ahmed A. Mohamed et al. introduced multiple summarization methods in [15], with the Q-Summarizer method standing out as particularly effective for generating query-based summaries with high relevance to the query. The paper authored by Samridhi Murarka et al.[2] presents a hybrid approach for query-relevant text summarization. Their methodology involves tokenization, normalization, stop-word removal, part-of-speech tagging, lemmatization, and context modeling using Latent Semantic Indexing. They also incorporate a Text Rank Algorithm. Although their dataset and metrics used are not explicitly mentioned, they propose the integration of a powerful encoder and beam search decoding for generating improved summaries.

A neural attention model is designed by the authors in [17] o generate concise summaries from input sentences. Their approach combines a neural language model with various encoders, with a focus on minimizing the negative log-likelihood of input-summary pairs during training, utilizing stochastic gradient descent for parameter estimation. The evaluation of their model employs the DUC-2003 and DUC-2004 datasets, consisting of 500 news articles paired with human-generated reference summaries, with all system outputs truncated to 75 characters to ensure unbiased length evaluation.

In a separate study by Johan Hasselqvist et al. [18], they present a sequence-to-sequence model with attention and a pointer mechanism tailored for query-based abstractive summarization. This model takes both a document and a query as input, processing them using bidirectional Recurrent Neural Network (RNN) encoders. Employing attention, it strategically focuses on relevant segments of the document and employs a generator mechanism to craft summaries. Their dataset, derived from CNN and Daily Mail news articles, comprises document-query-answer triples, with human-written highlights transformed into Cloze-style questions. The experimental results showcase the model's improved performance when queries are integrated, as evidenced by ROUGE scores. However, despite its notable performance, the model still falls short of a strong baseline, primarily due to dataset characteristics. An analysis of the model's attention mechanism reveals its tendency to intensely focus on a select few words in the document, with a bias towards the text's beginning, although it retains the ability to select entities from further back in the document. Table II provides an information about evaluation metrics used in text summarization.

TABLE II
EVALUATION METRICS FOR TEXT SUMMARIZATION

Metric	Information	Papers Cited In
ROUGE	ROUGE is a widely used metric for evaluating the quality of text summaries by measuring the overlap between machine-generated summaries and reference summaries in terms of precision, recall, and F1 score.	[1], [2], [3], [5], [9], [10], [11], [12], [20], [21], [24], [25]
F1 Score	The F1 score is a metric that combines precision and recall into a single value, making it useful for summarization evaluation. It measures the balance between precision and recall.	[7], [16]
MMR-MD Parameter	MMR-MD parameter is used to control redundancy in multi-document summaries. It affects the trade-off between removing redundancy and maintaining relevance in summaries.	[13]

IV. CONCLUSION

This comprehensive survey provides as a definitive exploration of the intricacies surrounding query-based text summarization. Through the meticulous categorization of research papers, it has illuminated four distinct taxonomies, artfully differentiating between machine learning methods, namely supervised and unsupervised, and the summary types, whether extractive or abstractive. By doing so, this survey not only highlights the remarkable developments in query-based text summarization but also underscores the notable gaps and scarcities in certain taxonomies, offering a concise summary of the current state of research. Furthermore, it emphasizes the potential for cross-pollination of methodologies from the broader domain of generic text summarization into the specialized field of query-based summarization.

REFERENCES

[1] Gayathri Venu,Madhuri Chandu "Extractive Approach For Query Based Text Summarization" , IEEE,2020

[2] Samridhi Murarka, Akshat Singhal "Query-based Single Document Summarization using Hybrid Semantic and Graph-based Approach ", IEEE,2020

[3] Nazreena Rahman, Bhogeswar Borah "A survey on existing extractive techniques for query-based text summarization", IEEE , 2020

[4] Ahmed A. Mohamed; Sanguthevar Rajasekaran "Improving Query-Based Summarization Using Document Graphs", IEEE,2018

[5] Xinghuo Ye; Hai Wei "Query-Based Summarization for Search Lists" IEEE,2019

[6] Muhammad Yahya Saeed, Muhammad Awais , Ramzan Takib ,Muhammad Younas "Unstructured Text Documents Summarization With Multi-Stage Clustering" IEEE, 2020

[7] Nazreena Rahman, Bhogeswar Borah "2015 International Symposiwn on Advanced Computing and Communication (ISACC)"

[8] Asad Abdia,∗, Siti Mariyam Shamsuddina, Ramiz M. Aliguliyevb "QMOS: Query-based multi-documents opinion-oriented summarization"

[9] Yan Du And Hua Huo "News Text Summarization Based on Multi-Feature and Fuzzy Logic" IEEE 2020

[10] Aakash Sinha,Abhishek Yadav, Akshay Gahlot "Extractive Text Summarization using Neural Networks"IEEE 2018

[11] V. K. Gupta and T. J. Siddiqui "Multi-document summarization using sentence clustering" 2012 4th International Conference on Intelligent Human

[12] Y. K. Meena, A. Jai and D. Gopalani "Survey on graph and cluster based approaches in multi-document text summarization" , IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE-2014), IEEE, Jaipur, India, 2014

[13] A. A. Mohamed " Generating user-focused, content-based summaries for multi-documents using document graphs", 5th IEEE, 2015

[14] Ángel Herandez-Castenda , Rene Arnulfo Garcia-Hernandez, Yulia Ledenza, And Christian Eduardo Millan-Hernandez "Extractive Automatic Text Summarization Based on Lexical-Semantic Keywords"IEEE 2020

[15] Ahmed A. Mohamed, Sanguthevar Rajasekaran "Improving Query-Based Summarization Using Document Graphs" IEEE

[16] Xinghuo Ye, Hai Wei "Query-Based Summarization for Search Lists" IEEE 2018

[17] StergosAfantenos,VangelisKarkaletsis,andPanagiotisStamatopoulos. 2005. Summarization from medical documents: a survey. Artificial intelligence in medicine 33, 2 (2005), 157–177.

[18] Johan Hasselqvist,Niklas Helmertz,Mikael Kågebäck "Query based abstractive summarization using Neural Networks"IEEE 2015
[19] Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova "BERT: Pretraining of deep bidirectional transformers for language understanding" Springer 2019
[20] Stergos Afantenos, Vangelis Karkaletsis, and Panagiotis Stamatopoulos. 2005. Summarization from medical documents: a survey. Artificial intelligence in medicine 33, 2 (2005), 157–177.
[21] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014).
[22] Breck Baldwin and Thomas S Morton. 1998. Dynamic coreferencebased summarization. In Proceedings of the Third Conference on Empirical Methods for Natural Language Processing. 1–6.
[23] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014).
[24] Eric Chu and Peter Liu. 2019. MeanSum: a neural model for unsupervised multi-document abstractive summarization. In International Conference on Machine Learning. PMLR, 1223–1232.
[25] John Conroy, Judith D Schlesinger, and Dianne P O'leary. 2006. Topicfocused multi-document summarization using an approximate oracle score. In Proceedings of the COLING/ACL 2006 main conference poster sessions. 152–159.
[26] Mariana Damova and Ivan Koychev. 2010. Query-based summarization: A survey. (2010).
[27] 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).
[28] Jade Goldstein, Vibhu O Mittal, Jaime G Carbonell, and Mark Kantrowitz. 2000. Multi-document summarization by sentence extraction. In NAACL-ANLP 2000 Workshop: Automatic Summarization.
[29] Vishal Gupta and Gurpreet Singh Lehal. 2010. A survey of text summarization extractive techniques. Journal of emerging technologies in web intelligence 2, 3 (2010), 258–268.
[30] Jon M Kleinberg. 1999. Authoritative sources in a hyperlinked environment. Journal of the ACM (JACM) 46, 5 (1999), 604–632.
[31] Piji Li, Wai Lam, Lidong Bing, Weiwei Guo, and Hang Li. 2017. Cascaded attention based unsupervised information distillation for compressive summarization. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2081–2090.
[32] Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In Text summarization branches out. 74–81.
[33] DzmitryBahdanau,KyunghyunCho,andYoshuaBengio.2014.Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473 (2014). [
[34] Yang Liu. 2019. Fine-tune BERT for extractive summarization. arXiv preprint arXiv:1903.10318 (2019).
[35] DarakshaParveen,Hans-MartinRamsl,andMichaelStrube.2015.Top- ical coherence for graph-based extractive summarization. In Proceed- ings of the 2015 Conference on Empirical Methods in Natural Language Processing. 1949–1954.