JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 06 Issue 01
2023
*Research Article*
https://doi.org/10.56979/601/2023

# Comparative Evaluation of Text Similarity Matrices for Enhanced Abstractive Summarization on CNN/Dailymail Corpus

**Muhammad Hammad Asif[1*] and, Aman Ullah Yaseen[1]**

[1]Depatmetof Creative Technologies, Air University, Islamabad, Pakistan.
*Corresponding Author: Muhammad Hammad Asif. Email: hammadasif1437@gmail.com

**Received:** October 12, 2023 **Accepted:** December 01, 2023 **Published:** December 05, 2023

**Abstract:** The importance of extracting essential information from extensive textual data is becoming increasingly crucial in our contemporary data-abundant environment. Using the CNN/Dailymail dataset, this study rigorously investigates the effect of text similarity matrices—including Cosine Similarity, Jaccard Index, and the Universal Sentence Encoder (USE)—on abstractive text summarization. It assesses how these matrices influence the quality of summaries. The research utilises comprehensive assessment techniques, such as BLEU scores, to examine the performance of each matrix. This delineates the effectiveness and limitations of employing quantitative analysis and graphical representations to succinctly and coherently summarize extensive textual content. While the Cosine Similarity metric exhibits notable efficacy in ensuring both coherence and informativeness, it marginally trails the Jaccard Index in these aspects. Despite its competitive standing, the Universal Sentence Encoder displays intricate attributes that significantly impact the quality of generated summaries. Moreover, the research delves into the potential implications for the realm of natural language processing (NLP) and offers pragmatic recommendations for practitioners, technology developers, and academics. The empirical findings of this study offer substantial insights into the pivotal role text similarity matrices play in condensing abstractive texts, thereby enriching comprehension.

**Keywords:** Text summarization; Abstractive Summary, Similarity Measure; Universal Sentence Encoder.

## 1. Introduction

The significance of extracting crucial insights from vast textual data is increasingly pivotal in today's data-rich landscape. Natural language processing (NLP), amalgamating computer science, linguistics, and machine learning principles [1], encompasses a wide spectrum of functionalities, including language translation, automated dialogue systems, and speech-to-text conversion. Within NLP, abstractive text summarization emerges as a compelling solution to this challenge. Focusing on text similarity matrices in abstractive summarization, this study scrutinizes news articles from the Dailymail dataset. It provides an extensive analysis of rapid advancements in deep learning, particularly emphasizing their practical applications in question answering, text summarization, and sentiment analysis. The responsibility of maintaining critical data integrity during the compression of vast textual datasets underscores the growing significance of these tasks.

Two primary classifications—extractive and abstractive summarization—define summarization techniques [2]. Abstractive summarization crafts concise synopses mirroring human-like coherence by interpreting and paraphrasing rather than directly quoting the source material. Conversely, extractive summarization preserves the original text by synthesizing succinct summaries using existing sentences or phrases. Abstractive methods prioritize novelty but demand a deeper understanding of context, while extractive methods emphasize precision yet may lack coherence. Widely employed across disciplines, these methodologies streamline information retrieval, condense articles, and summarize documents. Hybrid approaches integrate these methods to maximize their individual strengths, striving for accuracy, cohesiveness, and informative summaries.

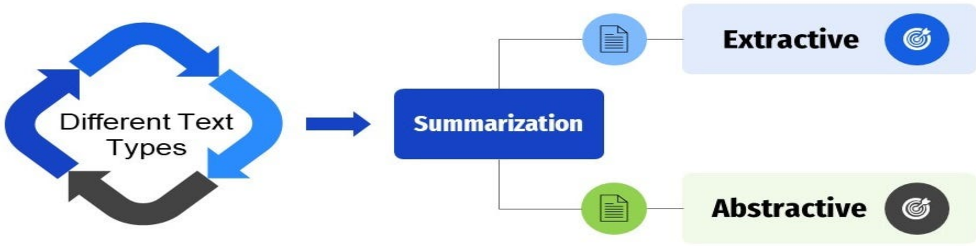


**Figure 1.** Type of Text Summarization

In order to produce abstractive summaries, a variety of deep learning methodologies are applied, one of which is Seq2Seq models [2], which utilizes recurrent neural networks (RNN) including LSTM and GRU. In recent times, transformer models have proven to be effective by incorporating attention mechanisms. Algorithms for dynamically modifying text summarization utilizing the universal sentence encoder have been developed in recent research to address practical applications. By utilizing the cosine similarity and Jaccard coefficient to measure sentence similarity and dissimilarity, respectively, this algorithm overcomes its limitations. Evaluation metrics like BLEU, METEOR [4], and F1 are considered to measure structural and contextual relevance in summaries. Automatic summarization in NLP aims to condense input text into compact versions preserving essential details, akin to abstracts or news releases.

In the domain of abstractive text summarization techniques, two prominent approaches emerge: methods predicated on structure and methods predicated on semantics. The former category comprises the subsequent: rule-based methodologies, which extract text extraction rules from patterns that have been extracted; template-based approaches, which are tailored to summaries individual or multiple documents; ontology-based techniques, which employ knowledge graphs that are specific to the domain; and tree-based methodologies, which utilize dependency trees to identify critical sentences for summarization. Conversely, approaches grounded in semantics utilize semantic graphs to discern between nouns and verbs within written materials. Rich semantic graphs are utilized in conjunction with SVM classification and other techniques in these methodologies to extract fundamental concepts. In addition, deep learning-based models, with the seq2seq model being particularly noteworthy, have gained recognition for their capacity to produce abstractive summaries and address diverse natural language processing (NLP) challenges, even in the presence of foreign words. Diverse approaches to summarization have surfaced in an attempt to enhance the process. Unified extractor models, guided summarization, contextualized rewriting, bottom-up approaches, and pointer generator networks are some of these methodologies. The goals of these initiatives are to address difficulties that arise during the process of summarization, including context-dependent refinement and the selection of repetitive words.

Abstractive text summarization and advanced natural language processing have garnered increased attention in recent years, largely as a result of advancements in deep neural networks (DNNs). GPT, ULMFit, BERT, ELMO, GPT-2 [5], and GPT-3 are language models that have significantly advanced numerous NLP tasks, including abstractive summarization. Transfer learning, which enables unsupervised pre-training of language models on tasks involving rich data and subsequent fine-tuning for specific purposes such as summarization, has been instrumental in reducing training time and costs. Although recent advancements in benchmark performance are being made with more recent models based on BERT and GPT-2, there is a lack of research into warm-starting sequence-to-sequence (seq2seq) models, which are commonly employed for abstractive summarization and utilize pre-trained models. Selected seq2seq techniques that are founded on Transformers and employ pre-trained model checkpoints that are publicly available, such as BERT, exhibit comparable performance to larger models (e.g., T5, BART[6], PEGASUS, ProphetNet[7]) while requiring considerably less training capital. This investigation into the combination of models demonstrates cutting-edge performance, specifically in the domain of abstractive text summarization, with a specific emphasis on generating headlines—succinct overviews of the given text.

Extractive methods rely on rudimentary heuristics to identify and concatenate the most relevant sentences, without taking into account grammatical or syntactical principles. As a consequence, the output produced is both unintelligible and incoherent. The aforementioned situation required the creation of abstract techniques [8]. To ensure a cohesive and uninterrupted summary, it is imperative to incorporate supplementary contextual information pertaining to the tokens present in the input text. As a result, a compilation of models that produce unique phrases in a manner similar to that of a human reader paraphrasing is necessary. A plethora of abstractive summarization models have been previously documented in the academic literature. Recent surveys have provided evidence that these techniques consist of rule-based [9], graph-based [8], and semantic modelling [8].

However, numerous NLP tasks have been enhanced by recent developments in deep learning that were not accounted for in the aforementioned models. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are modern approaches to abstractive summarization that are built upon the deep learning model architecture. The LSTM and GRU architectures, which augment the initial RNNs, are discussed in [9]. GANs, or generative adversarial networks, are an additional neural architecture category distinct from CNNs and RNNs.

Transformer [11], a deep learning model composed of numerous encoder and decoder layers, simulates the global dependencies of sequential input by employing the attention mechanism. To provide greater specificity, the self-attention mechanism assigns different weights to different input components in accordance with their contextual relevance. During the generation of the output sequence, these components are encoded in hidden state layers. In addition, transformer models utilize multi-head attention, a technique that concurrently applies attention to detect and distinguish between diverse patterns and correlations within the incoming data. Utilizing the encoder-decoder paradigm, which generates output by encoding and decoding data into hidden layers, Transformer generates output. These models are classified as semi-supervised due to the fact that they are refined explicitly following unsupervised pretraining on massive datasets. Approaches grounded in this framework achieve state-of-the-art outcomes across various text generation endeavors, including abstractive summarization.

In recent times, surveys have been conducted to assess and examine the distinctions between previous abstractive methodologies, including those that utilized deep learning models, before the Transformer architecture was introduced [12]. It is imperative to recognize that the primary emphasis of the research detailed in this article pertains to well-established techniques for pretrained language models. These techniques, which are dependent on the transformer, are described in greater detail in the following section.

The T5 [13] text-to-text transfer transformer is an architectural design that is strikingly similar to the Transformer architecture. The framework offers an all-encompassing structure for text-to-text conversion of a series of natural language processing (NLP) tasks. A prefix that is unique to the given task is appended to the input sequence in order to identify it. In the pretraining phase, both supervised and unsupervised learning are implemented. An element of the unsupervised objective involves concealing random gaps in tokens through the use of unique sentinel tokens. The decoder acquires the capability to forecast the tokens that are omitted from the output layer, whereas the encoder is provided with the "corrupted" phrase.

Bidirectional auto-regressive transformers, or BART [14], are an abstractive summarization-incorporating multitask deep learning approach. By employing an assortment of textual alterations, the "denoising" autoencoder from BART discerns the connections between a given document and its "corrupted" counterpart. Document rotation, text infilling, sentence permutation, and random token deletion or masking are some examples. This autoencoder is utilized by the sequence-to-sequence model as both a bidirectional encoder and a left-to-right auto-regressive decoder. To optimize a reconstruction loss function (cross-entropy) prior to training, the decoder generates tokens which possess an increased likelihood of being present in the source document.

PEGASUS, representing pretraining with extracted gap-sentences for abstractive summarization, is a deep learning methodology specifically engineered to undergo pre-training with the aim of achieving abstractive summarization as the final outcome [15]. Gap sentence generation (GSG) is introduced as an innovative pretraining target for models based on Transformers. This goal is specifically designed for the task of abstractive text summarization, as it obscures entire sentences as opposed to the narrower text spans utilized in alternative approaches. This results in a "gap" in the input document, which is subsequently filled by instructing the model to consider the remaining sentences.

As predicated in [16], In order to address the escalating size and computational complexity of the models, the researchers were motivated to explore techniques for compressing large pretrained models into smaller ones without sacrificing accuracy or execution speedier inference. Illustrative of this is the research conducted by [7], which proposes a number of comprehension strategies, including: I direct knowledge distillation (KD), which facilitates the transfer of knowledge from a larger model to a smaller one called the "student" model; (ii) pseudo-labels, which substitute the ground truth target documents of the teacher model with those of the student model; and (iii) shrink and finetune (SFT), which reduces the teacher model to the dimensions of the "student" model.

As an essential component of information retrieval and natural language processing, extractive summarization condenses lengthy texts into succinct summaries. Traditional methods encompass topic word extraction, sentence clustering based on user-queried topics, and frequency-driven strategies that assign weights to sentences containing frequently occurring words using word probability (Term Frequency) and TF-IDF (Term Frequency-Inverse Document Frequency). Enhancing the efficiency of TF-IDF in sentence selection, centroid-based summary generation is implemented to group TF-IDF topic-associated words. Utilizing machine learning models such as Support Vector Machines (SVM) and Decision Trees, supervised learning approaches have surfaced to rank sentences automatically according to their coherence within the summary and their relevance to the document's content. Latent Semantic Analysis (LSA) [17] generates a topic-sentence matrix through the utilization of singular value decomposition; sentences with the highest scores are chosen to be included in the summary. In contrast, decision trees divide sentences into segments according to attributes including length, frequency of terms, and document position. These segments are then classified as either pertinent or extraneous for the purpose of summarization. These methodologies symbolize crucial advancements in the automation of sentence selection for extractive summarization, making substantial contributions to the accessibility and administration of extensive textual data.

The primary aim of extractive techniques is to discern the most crucial sentences from a given document in order to generate a succinct summary that encapsulates the fundamental concepts presented in the original material. For extractive summarization, numerous algorithms have been proposed, each employing a unique strategy for sentence ranking and extraction. The aforementioned methods include statistical algorithms that employ metrics such as the frequency of words or sentences, graph-based algorithms that utilize concepts from graph theory (e.g., centrality and measures for detecting communities), and semantic algorithms that model sentences and terms into a co-occurrence matrix that is examined using distributional semantics [18].

## 2. Materials and Methods

This methodology stands as a pivotal endeavor, driven by the pressing need to comprehensively understand how text similarity matrices influence abstractive summarization. Our research hinges on meticulous experimental design, model selection, and data preprocessing, forming the bedrock for a systematic exploration of these matrices' strengths, limitations, and practical implications.

Our methodology is predicated on the implementation of the BART model, which is a context-aware transformer architecture and a major advancement in NLP. The system's ability to produce intricate abstractive summaries is consistent with our research goals, offering the potential for refined outcomes that are grounded in the subtleties of the source material. By utilizing the CNN/Dailymail dataset, which is widely acknowledged as a benchmark, we can enhance the dependability of our findings by guaranteeing consistent and significant evaluations. The CNN/Daily Mail dataset comprises an extensive collection of news articles that cover a wide range of topics, including entertainment, politics, sports, and more. Its extensive collection of 280,000 pairs and concise summaries render it a favored option among researchers for the purpose of training and assessing text summarization models. The importance of it in abstractive summarization is substantial for a multitude of reasons. To begin with, the content is derived from reputable sources such as CNN and the Daily Mail, which guarantees inclusivity in terms of subjects and aesthetics that are vital for the evaluation and training of models. Furthermore, the inclusion of summaries generated by humans provides benchmarks of superior quality against which to assess the efficacy of models and fosters the development of novel ideas. Complexity, richness of language, and detailed content of the dataset necessitate advanced language comprehension for the creation of coherent abstract summaries; thus, it serves as a standard against which model coherence, informativeness, and fluency are evaluated. The extensive utilization of this dataset has spurred progress in the field of abstractive summarization methods, rendering it an essential asset in the improvement of natural language comprehension and generation.

Our methodological framework involves an investigation into a range of text similarity metrics, including both conventional and modern techniques such as sentence embedding-based similarities and cosine similarity. These tools afford us the capability to scrutinize the intricate textual interconnections that are inherent in the process of abstractive summarization. As we advance through this chapter, we extend a warm invitation to the reader to join us on a scholarly investigation firmly rooted in rigorous methodology. Our collective aim is to reveal empirical discoveries that possess the capacity to fundamentally transform the domain of abstract text summarization. By utilizing this approach, we are strategically positioned to exploit the potential of text similarity matrices, thereby paving the way for informed methodologies and groundbreaking developments in the field.

The cosine similarity is an essential metric that finds application in text analysis and natural language processing. It enables the evaluation of textual relationships through the comparison of similar vectors in a quantifiable manner. The algorithm assesses the degree of similarity among sentences or documents and generates results ranging from -1 (that is, ideal similarity) to 0 (total dissimilarity) (zero similarity). It is widely employed in the process of summarizing texts (lack of similarity). Although it can be useful for a range of text analysis tasks including clustering, document retrieval, and sentence similarity determination, its straightforwardness in extractive summarization hinders semantic comprehension and term frequency sensitivity. As a result, it becomes necessary to combine this approach with supplementary methods in order to achieve more sophisticated abstractive summarization techniques. Another crucial metric, the Jaccard Index, assesses the degree of word or concept overlap that exists between textual components such as sentences and documents. While proficient in determining similarity, it encounters difficulties in comprehending semantics and contextual nuances; therefore, it is necessary to incorporate more advanced summarization methods. The utilization of similarity matrices based on sentence embedding has made substantial strides in the field of text summarization through the incorporation of semantic and contextual subtleties. Despite demanding computational resources and labeled datasets, their ability to convey semantic comprehension and contextual relevance establishes them as essential tools in abstract summarization, prompting careful consideration of method choice and computational demands in practice.

Evaluation metrics play a pivotal role in assessing the performance of diverse deep-learning methods, and several prominent ones stand out in this regard. BLEU [19], founded on n-gram comparisons between original and summarized texts, focuses on precision and brevity, though it overlooks sentence structure. ROUGE encompasses various matrices, including ROUGE-N, Recall, Precision, and F1 Score, providing insights into n-gram matching and context, ensuring the inclusion of essential information in generated summaries. As opposed to BLEU, METEOR evaluates n-grams while taking synonyms and word stems into account. By employing metrics such as precision, recall, harmonic means, and penalty terms, the degree of similarity between the candidate and reference texts is assessed. In contrast to the algorithm for text summarization proposed in [10] that makes use of transformer models, dynamic text changes, and graph reduction techniques, the chosen evaluation metric for its assessments is BLEU. The aforementioned evaluation matrices provide extensive metrics for assessing the quality of summarization. Each matrix concentrates on evaluating a distinct facet of the fidelity and caliber of the generated summaries.

## 3. Results

This section provides a critical analysis of the functions and efficacy of three crucial text similarity matrices: Cosine Similarity, Jaccard Index, and Universal Sentence Encoder (USE). The aim of this analysis is to determine whether an innovative selection mechanism that utilizes sentence-level similarity could be implemented to facilitate the summarization of abstract texts. The BLEU metric continues to be the principal quantitative evaluation criterion, utilized to assess the overall efficacy and precision of the produced abstractive summaries pertaining to reference summaries. BLEU scores function as a quantitative metric that can be utilized to assess the efficacy of summaries in capturing essential information and lexical similarity.

By incorporating similarity matrices at the sentence level, this approach distinguishes itself from traditional summarization methods in an effort to enhance the abstractive summarization process. The utilization of these matrices shifts the focus from assessing the similarity of entire documents to examining a more specific subset of sentences. This facilitates the generation of abstractive synopses that are more coherent and suitable for the provided context. The objective of this research endeavor is to analyze the implications of cosine in various domains, such as summarization contexts, the identification of similarities based on vector orientation, and the determination of the cosine angle between vectors.

The present study examines the utilization of the Jaccard utility to assess similarity at the sentence level, its approach to determining overlap, and its application in evaluating similarities based on sets.

The objective of this research endeavor is to examine the capacity of USE [21] to transform sentences into significant embeddings, assess its versatility across different linguistic contexts, and investigate its potential in the domain of abstractive summarization. The objective of this concentrated examination is to offer a comprehensive and nuanced comprehension of the unique functions and relative efficacy of Universal Sentence Encoder (USE), Jaccard Index, and Cosine Similarity as they pertain to the enhancement of abstractive summaries via the novel sentence-level similarity-based selection mechanism.

The results of this study showcase the varying impact of different similarity metrics on the performance of the BART model for summary generation. When the BART model operated in isolation, it attained a moderate BLEU score of 0.033, indicating its baseline performance. However, when integrated with different similarity metrics, the impact on the summary quality differed significantly. The integration of cosine similarity and Jaccard similarity led to a reduction in BLEU scores, demonstrating a decrease in the summary quality compared to BART operating alone. This reduction suggests that these similarity metrics might not be as conducive to enhancing the BART model's summary generation capabilities.

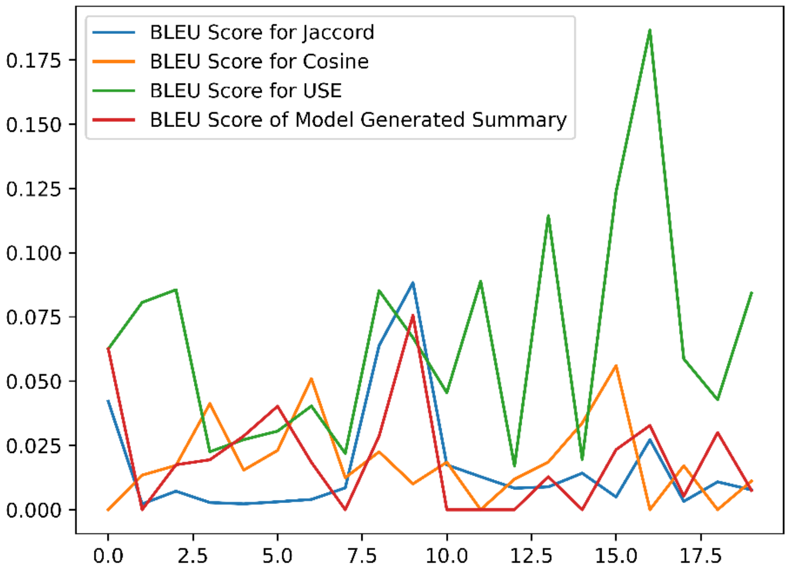


**Figure 2.** Comparative Graph of BLEU for different Similarity Matrices

**Table 1.** BLEU Score Comparison

| Model | BLEU Score |
|---|---|
| BART | 0.033 |
| BART + Cosine | 0.019 |
| BART + Jaccard | 0.017 |
| BART + USE | 0.086 |

## 4. Discussion

On the contrary, a substantial enhancement was noted following the implementation of the Universal Sentence Encoder (USE), which exhibited results that were considerably more favorable than those achieved through the utilization of alternative similarity metrics. Through the achievement of a noteworthy BLEU score of 0.086, the BART model, facilitated by USE, exhibited its effectiveness in augmenting the caliber of the produced summaries. The substantial performance improvement highlights the feasibility of incorporating advanced embedding techniques, such as the USE, to significantly enhance the BART model's summarization capabilities.

A notable discrepancy exists in the performance of various similarity metrics; of these, the USE demonstrates superiority over the cosine and Jaccard similarities. This highlights the crucial significance of embedding techniques in enhancing the quality of summaries. This implies that the integration of sophisticated embedding techniques, such as the USE, has the potential to significantly enhance the efficacy of summarization models by producing summaries of superior quality.

Overall, the results of this research emphasize the importance of judiciously selecting and incorporating appropriate similarity metrics or embedding techniques when attempting to enhance the effectiveness of summarization models like BART. The significant improvement identified with the USE highlights its potential to significantly improve the quality of generated summaries, thus offering promising opportunities for future research and advancements in the field of text summarization methodologies.

**5. Conclusions**

The concluding section provides a concise overview of the main findings and emphasizes the broader ramifications of the research for the process of summarizing abstract texts. Significant findings emerged from our investigation into text similarity matrices in relation to abstractive summarization. The evaluation of the efficacy of various matrices—Cosine Similarity, Jaccard Index, and Universal Sentence Encoder—revealed nuanced impacts on the summaries' coherence and quality. The summary outcomes were distinct in their impact due to the individual contributions and constraints of each matrix. It is imperative to acquire a thorough comprehension of the operation of text similarity matrices in order to advance the development of abstractive summarization techniques. Our research substantiates the criticality of understanding the influence that these matrices exert on the production of summaries.

Our research contributes significantly to the advancement of knowledge regarding text similarity matrices. By uncovering novel insights and conducting empirical evaluations, this study provides a significant addition to the ever-evolving domain of abstractive summarization. It contributes nuanced perspectives that enrich the discourse and guide further inquiries in this domain. The implications of our findings extend beyond concise overviews, potentially fostering advancements in the field of natural language processing and related academic disciplines. Further research could explore alternative matrices, improve methodologies, and integrate diverse datasets, thus contributing to a more comprehensive understanding of the performance of matrices across multiple domains.

In summary, this research sheds light on the significant influence that text similarity matrices have on the results of abstractive summarization. This highlights the significance of selecting and integrating an informed matrix, which acts as a catalyst for the progression of research and practical implementations in the wider domain of natural language processing.

**References**


1. Cambria, E., & White, B. (2014). Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]. IEEE Computational Intelligence Magazine, 9(2), 48–57. https://doi.org/10.1109/MCI.2014.2307227
2. Munot, N., & Govilkar, S. S. (2014). Comparative study of text summarization methods. International Journal of Computer Applications, 102(12).
3. Gong, Y., & Liu, X. (2001). Generic text summarization using relevance measure and latent semantic analysis. Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 19–25. https://doi.org/10.1145/383952.383955
4. Banerjee, S., & Lavie, A. (2005, June). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization (pp. 65-72).
5. Bhat, I. K., Mohd, M., & Hashmy, R. (2018). Sumitup: A hybrid single-document text summarizer. In Soft Computing: Theories and Applications: Proceedings of SoCTA 2016, Volume 1 (pp. 619-634). Springer Singapore.
6. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., ... & Zettlemoyer, L. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461.
7. Qi, W., Yan, Y., Gong, Y., Liu, D., Duan, N., Chen, J., ... & Zhou, M. (2020). Prophetnet: Predicting future n-gram for sequence-to-sequence pre-training. arXiv preprint arXiv:2001.04063.
8. Radev, D. R., Jing, H., Styś, M., & Tam, D. (2004). Centroid-based summarization of multiple documents. Information Processing & Management, 40(6), 919-938.
9. Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). Text summarization techniques: a brief survey. arXiv preprint arXiv:1707.02268.
10. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.
11. Liu, X., & Xv, L. (2019, December). Abstract summarization based on the combination of transformer and LSTM. In 2019 International Conference on Intelligent Computing, Automation and Systems (ICICAS) (pp. 923-927). IEEE.
12. Saggion, H., & Poibeau, T. (2013). Automatic text summarization: Past, present and future. Multi-source, multilingual information extraction and summarization, 3-21.
13. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. The Journal of Machine Learning Research, 21(1), 5485-5551.
14. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., ... & Zettlemoyer, L. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461.
15. Zhang, J., Zhao, Y., Saleh, M., & Liu, P. (2020, November). Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In International Conference on Machine Learning (pp. 11328-11339). PMLR.
16. Nallapati, R., Zhou, B., Gulcehre, C., & Xiang, B. (2016). Abstractive text summarization using sequence-to-sequence rnns and beyond. arXiv preprint arXiv:1602.06023.
17. Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. Journal of the American society for information science, 41(6), 391-407.
18. Rieger, B. B. (1991). On distributed representation in word semantics. Berkeley, CA: International Computer Science Institute.
19. Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002, July). Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics (pp. 311-318).
20. Rawat, R., Rawat, P., Elahi, V., & Elahi, A. (2021, April). Abstractive Summarization on Dynamically Changing Text. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1158-1163). IEEE.
21. Cer, D., Yang, Y., Kong, S. Y., Hua, N., Limtiaco, N., John, R. S., ... & Kurzweil, R. (2018). Universal sentence encoder. arXiv preprint arXiv:1803.11175.